Journal of Basrah Researches (Sciences) **50**(1), 74 (2024)
DOI: https://doi.org/10.56714/bjrs.50.1.7

Face Anti-Spoofing Detection with Multi-Modal CNN Enhanced by ResNet

Hala S. Mahmood[1*], **Salah Al-Darraji**[2]

[1]Department of Computer Science, College of Education for Pure Sciences, University of Basrah, Basrah, Iraq.
[2]Department of Computer Science, College of Computer Science & Information Technology, University of Basrah, Basrah, Iraq.

ARTICLE INFO

Received 09 January 2024
Accepted 19 February 2024
Published 20 June 2024

Keywords:

Face anti-spoofing, presentation attack detection, PAD, face recognition, ResNet-50, deep learning.

Citation: H.S. Mahmood, S. Al-Darraji, J. Basrah Res. (Sci.) **50**(1), 74 (2024). DOI:https://doi.org/10.56714/bjrs.50.1.7

ABSTRACT

The growing prevalence of face recognition technology in various applications, including mobile devices, access control, and financial transactions, highlights its importance. However, the vulnerability of face recognition systems to attacks has been demonstrated, underscoring the necessity of addressing potential weaknesses that attackers may exploit. The paper delves into face presentation attack detection (PAD) within biometric systems, which is crucial for ensuring the reliability and security of face recognition algorithms. To address this issue, the paper proposes a method for face presentation attack detection using ResNet-50 in conjunction with multi-modal data, incorporating RGB, depth, infrared (IR), and thermal channels. The method explores diverse strategies to combine results from each modality, investigating various fusion techniques such as majority voting, weighted voting, average pooling, and a stacking classifier. The system has been tested on the WMCA dataset. It exhibits strong performance compared to existing methods, notably achieving an impressive ACER ratio of 0.087% with the stacking classifier. This approach proves effective by consolidating multiple modalities without requiring individual scenario-specific models, indicating promise for real-world applications.

1. Introduction

Automated authentication systems need to be protected against spoofing attacks to prevent illegal entry [1], [2]. Face anti-spoofing (FAS) is crucial for enhancing the security of face recognition systems by protecting them from presentation attacks. Face recognition has made remarkable progress, with state-of-the-art systems exceeding the performance of humans [3]. A substantial portion of this achievement may be ascribed to the accessibility of extensive annotated face datasets often gathered via the internet. In contrast, datasets for face anti-spoofing assaults need a laborious procedure of manual data acquisition, resulting in a restricted number of distinct individuals and samples.

[*] **Corresponding author email:** hala.shaker@uobasrah.edu.iq

ISSN: 1817-2695 (Print); 2411-524X (Online)
Online at: https://jou.jobrs.edu.iq

The prevailing FAS datasets mostly consist of RGB images alone. Additionally, prior multi-modal datasets have only included a restricted number of subjects, therefore posing a risk of overfitting the training data for the present methods. To enhance liveness detection, anti-spoofing algorithms may use other image modalities, such as infrared, thermal, and depth channels, which are obtained via specialized cameras [4]. The combination of these modalities is anticipated to enhance liveness detection since they provide complementary information. The deployed systems must possess the capability to ascertain the vitality of the individual in the camera's field of view by means of identifying and rejecting any forms of FAS, such as printed images, replay attacks, 3D masks, and other similar methods.

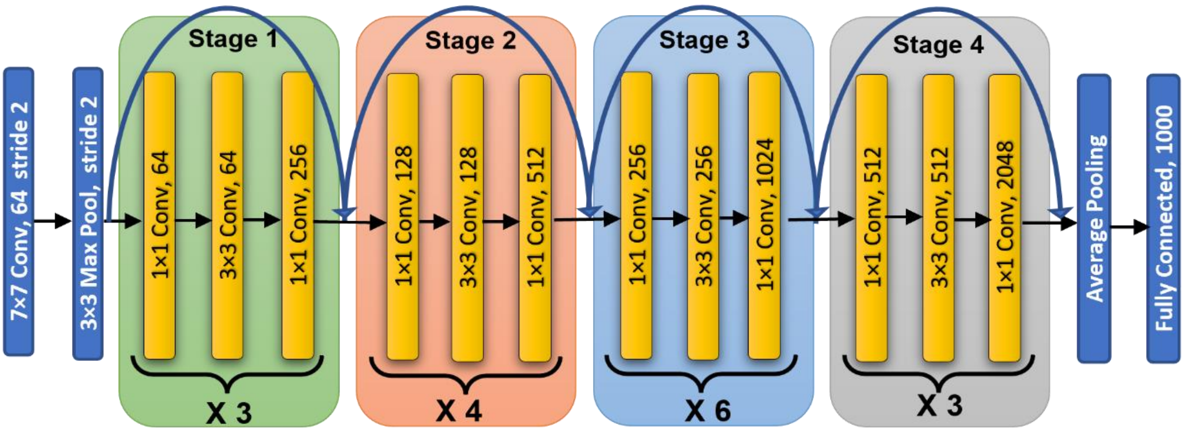

Fig. 1. The architecture of ResNet-50.

Anti-spoofing algorithms may gain advantages by using other image modalities, such as infrared (IR) and depth channels, which are obtained via specialized cameras. The combination of these modalities is anticipated to enhance liveness detection since they provide complementary information. IR cameras are impervious to electronic displays and may thwart assaults from phones and tablets. Additionally, the depth channel facilitates the differentiation between flat-printed surfaces and facial contours. The approach aims to extract comprehensive and varied information from each modality in order to improve the system's capacity to differentiate between *bonafide* and fake faces.The proposed approach is evaluated on the wide multi-channel presentation attack (WMCA) [5] dataset and explores different techniques for combining the results from each modality.The paper evaluates the performance of convolutional neural networks (CNN) for face anti-spoofing, specifically using the ResNet-50 architecture [6]. ResNet-50 is a 50-layer deep learning model that has been shown to perform effectively in various applications. Its architecture consists of multiple residual blocks, which help the network learn residual functions instead of explicitly learning the desired underlying mapping, allowing for more efficient and effective training, see Fig. 1. This research paper aims to explore the importance of implementing anti-spoofing measures in automated authentication systems, focusing on the identification and prevention of face presentation assaults. The research utilizes CNN to address face anti-spoofing by integrating RGB, IR, depth, and thermal modalities. Each modality undergoes a separate analysis, resulting in unique outcomes. These CNN architectures will provide positive outcomes for FAS in many scenarios. These results will be compared with the outcomes of papers [5], [7], revealing that the current findings exhibit superior performance. Subsequently, the results of all modalities will be combined using different techniques, such as majority voting [8], weighted voting [9], average/pooling [10], and stacking classifiers [11]. These techniques aim to capitalize on the diversity of available information, enhancing the performance of the resulting model in various applications.

The main contributions can be encapsulated as follows:

- This study introduces a multi-modal strategy for face anti-spoofing, leveraging four distinct modalities: RGB, depth, IR, and thermal. This integration of diverse modalities enhances the capability to differentiate between authentic and fraudulent faces.
- Various methodologies for amalgamating outcomes from individual modalities are explored. A novel model is trained to utilize these modality-specific results as features, facilitating the prediction of the final anti-spoofing verdict. Employing a stacking approach enables the model to learn from combined information, enabling more informed decision-making.
- To evaluate the model's efficacy, cross-validation techniques are employed, addressing dataset partitioning concerns and ensuring a more robust assessment of the model's performance. The paper's structure is organized as follows: Section 2 covers pertinent literature, while section 3 introduces the proposed method and network architecture. Section 4 provides an account of the experimental study conducted on the proposed PAD algorithm, which was trained using the WMCA database. This section includes comprehensive information regarding the dataset, methodology, working specifics, and assessment methodologies. Section 5 focuses on results and discussions, and the paper concludes in Section 6.

2. Related Work

Face recognition has gained widespread usage in applications such as access control and E-payment due to its practicality [12]. However, this technology faces security concerns, particularly regarding presentation attacks as obtaining facial images has become easier, given the prevalence of social networks. Despite this broad scope, a majority of face Presentation Attack Detection research focuses primarily on two types of attacks: photo and replay attacks. This emphasis could be attributed to the dominance of publicly available databases that primarily contain these attack types, such as NUAA [12], CASIA-FASD [13], OULUNPU [14], and RECOD-Mtablet [15]. Notably, these datasets primarily contain RGB data, limiting the diversity of information available for related algorithms. Even newer datasets like HiFiMask [16] and Synth A Spoof [17] are based solely on RGB data. Recent studies [18] have indicated that face PAD systems based solely on RGB data exhibit relatively poor performance, even in detecting the aforementioned two attack types. Moreover, their performance degrades further when tested against unseen attack scenarios. These experiments utilize a database encompassing a broader spectrum of presentation attacks, including both 2D and 3D attacks, along with partial attacks. This experimentation highlights that RGB-based PAD systems exhibit subpar performance even when evaluated against known attack scenarios. Evolved presentation attacks, such as 3D attacks and silicone masks [19], highlight the limitations of visible cameras in detecting sophisticated fraudulent attempts. Advancements in attack methods have prompted the development of new sensor technologies like depth cameras, multi-spectral cameras, and infrared light cameras, expanding the options for face PAD approaches. Previous studies have explored fusion strategies for different modalities, with some approaches combining binary scores from RGB, depth, and IR models, showcasing superior performance on datasets like CeFA [20]. Others propose decision-level fusion strategies that aggregate scores from multiple models using depth or IR modalities for enhanced live/spoof classification. Specific datasets like Msspoof [21], and CSMAD [22] provide multi-modal information, combining RGB, IR, Depth, and thermal images. However, limitations in the number of subjects and samples in these datasets might pose risks of overfitting when evaluating face PAD algorithms. Recently released datasets like WMCA aim to overcome limitations by increasing dataset sizes and including multiple modalities (RGB, depth, IR, and thermal), presenting new challenges for liveness identification [5]. Various studies have leveraged modern CNN in face anti-spoofing research, using them as effective feature extractors to discern live and spoof faces. Some approaches focus on utilizing reflectance properties to differentiate between facial skin and mask materials, while others employ pretrained models for feature extraction.

In summary, while face recognition technology has seen extensive use, challenges persist, particularly concerning sophisticated presentation attacks. The evolution of new sensor technologies and larger, multi-modal datasets present opportunities for improved liveness detection and robust anti-spoofing measures.

3. The Proposed Method

This section delineates the different phases of the suggested presentation attack detection framework.

3.1. Pre-processing

Preprocessing plays a crucial role in PAD by improving accuracy through the standardization and alignment of input images. This procedure effectively reduces variances instance, illumination, and obstructions. The MTCNN algorithm is used for detecting faces in the color channel. After obtaining the face bounding box, the supervised descent method (SDM) is used to recognize the facial landmarks inside this bounding box. Alignment is accomplished by converting the image in order to align the centers of the eyes and mouths with predetermined coordinates. The aligned facial photos undergo a conversion to grayscale and are then enlarged to a resolution of 128×128 pixels. Regarding non-RGB channels, preprocessing involves the spatial and temporal alignment of images from different channels with the color channel. A similar alignment procedure is carried out for these channels, using face features identified in the RGB channel. An additional normalization process utilizing mean absolute deviation (MAD) guarantees that the range of non-RGB facial images is transformed into an 8-bit format [23].

3.2. Network Architecture

The paper introduces a CNN design using a multi-modal ResNet-50 architecture for facial impersonation detection. The ResNet-50 employed in this study involves inputs comprising RGB, depth, infrared, and thermal images, each having dimensions of 128×128 pixels, sourced from the WMCA dataset [12]. The ResNet-50 architecture employs a bottleneck design in its building blocks, which involves the use of 1×1 convolutions referred to as "bottlenecks". This design helps to decrease the number of parameters and speed up the training process for each layer. In contrast to the conventional ResNet-50, its bottleneck design incorporates a structure of three layers as opposed to two, see Fig. 1. The architecture of the 50-layer ResNet consists of many components: a convolutional layer with a 7×7 kernel and a stride of 2, followed by 64 kernels with a stride of 2, a max pooling layer with a stride of 2, and other layers with varying kernel sizes arranged in certain patterns. The architecture encompasses 12 layers iterating four times, consisting of 1×1,128 kernels, 3×3,128 kernels, and 1×1,512 kernels. Subsequently, there are 18 layers, iterating six times, featuring 1×1,256 kernels, 3×3,256 kernels, and 1×1,1024 kernels. Finally, an additional 9 layers, repeating three times, utilize 1×1,512 kernels, 3×3,512 kernels, and 1×1,2048 kernels, culminating in a 50-layer network design. The model's training involves employing the "cross-binary" loss function using the "adam" optimizer [34]. The evaluation metric for this model is set to "accuracy".

In this context, individual modalities undergo separate processing, each subjected to the anti-spoofing algorithm, resulting in distinct outputs, see Fig. 2. These outputs are then amalgamated through various fusion techniques:

- Majority Voting [8]: This fusion method involves each classifier within the ensemble making predictions, and the ultimate result is chosen by picking the class that receives the majority of votes.
- Weighted Voting [9]: Unlike equal-weighted predictions, this approach assigns varied weights to classifier predictions based on their performance or confidence. These weights are usually determined through cross-validation or model performance on a validation set.
- Averaging/Pooling [10]: Here, the predicted probabilities or scores from multiple classifiers are averaged to yield the final prediction. In the case of multi-class classification, probabilities for each class from every classifier are averaged.
- Stacking/Stacked Generalization [11]: This technique involves training a meta-classifier that takes individual classifier predictions as additional features to make the final prediction.

The efficacy of these fusion techniques is evaluated to determine the most optimal method for the system's ultimate output.

The proposed architecture, exhibits promising outcomes in face anti-spoofing. Leveraging multi-modal data fusion and diverse fusion methods augments the system's security, resilience, and accuracy in anti-spoofing tasks. These findings contribute significantly to advancing face anti-spoofing methodologies, highlighting the importance of leveraging multiple modalities and fusion strategies for the development of reliable facial recognition systems.

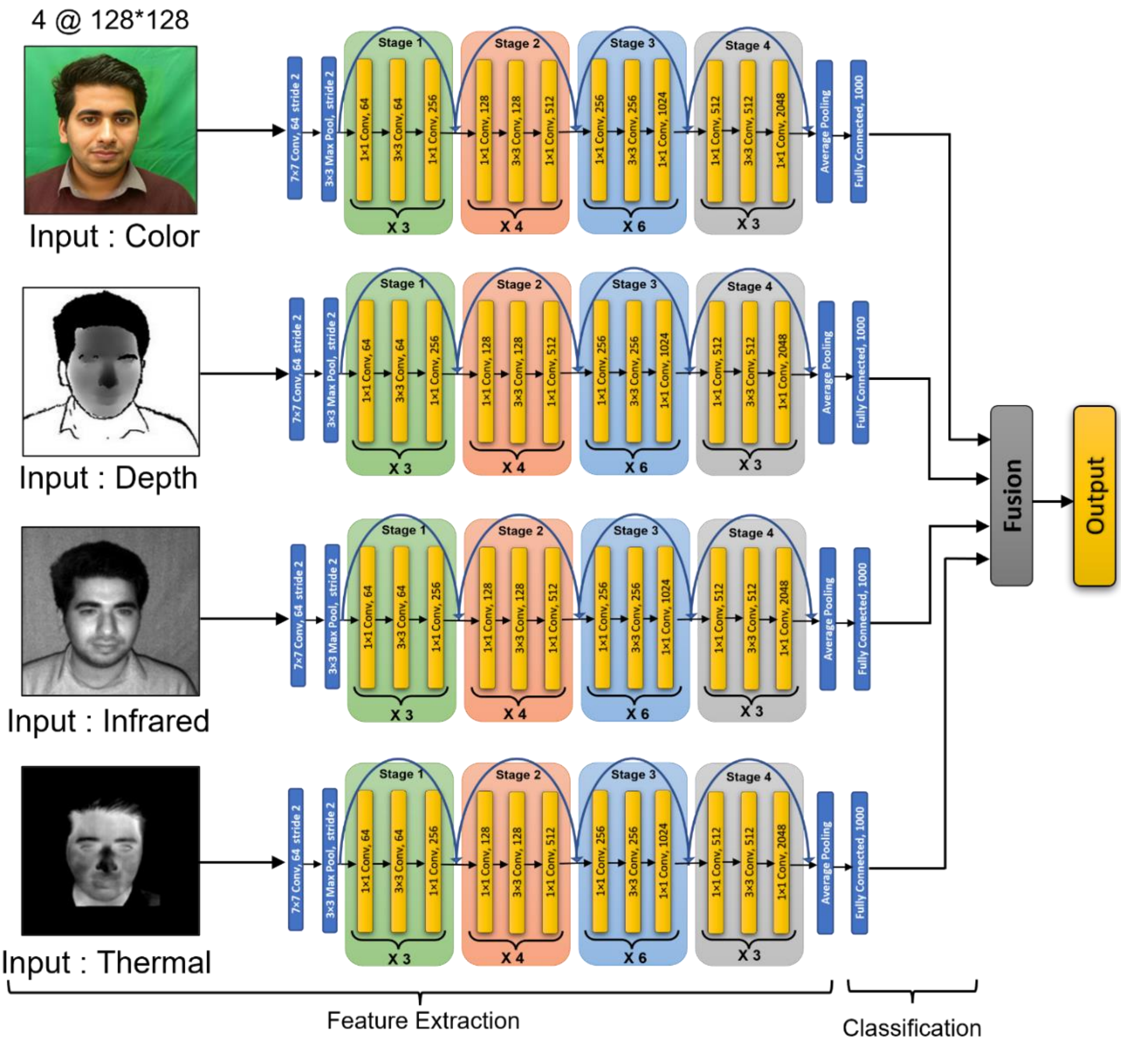

Fig. 2. The proposed architecture.

4. Experimental Results

This section offers in-depth explanations regarding the database used and the experimental outcomes attained using the proposed approach. The subsequent subsections delineate the various phases of the suggested framework.

4.1. Dataset

The methodology utilized the Wide Multi-Channel Presentation Attack database (WMCA) [5] to evaluate its approach. This database encompasses 1679 videos, comprising 347 *bonafide* presentations and 1332 attack instances. Data expansion involved extracting approximately 50 frames from each video, resulting in a total of 83,950 images utilized for augmentation purposes (refer to Table 1) [5]. The attacks encompass seven types: print, glasses, replay, rigid mask, fake head, flexible mask, and paper mask, as shown in Fig. 3. The WMCA database contains recordings in four distinct modalities: color data, depth maps, infrared images, and thermal data. Capture devices for RGB, depth, and infrared data included the Intel RealSense SR300 camera, while thermal data was sourced from the Seek Thermal Compact PRO camera. All images maintain a standardized size of 128×128 pixels and have been geometrically aligned to ensure uniformity. Samples illustrating these four image modalities are showcased in Fig. 4, with each image displayed across the four modalities: color intensity (C), depth (D), infrared imaging (I), and thermal data (T). See Fig. 4.

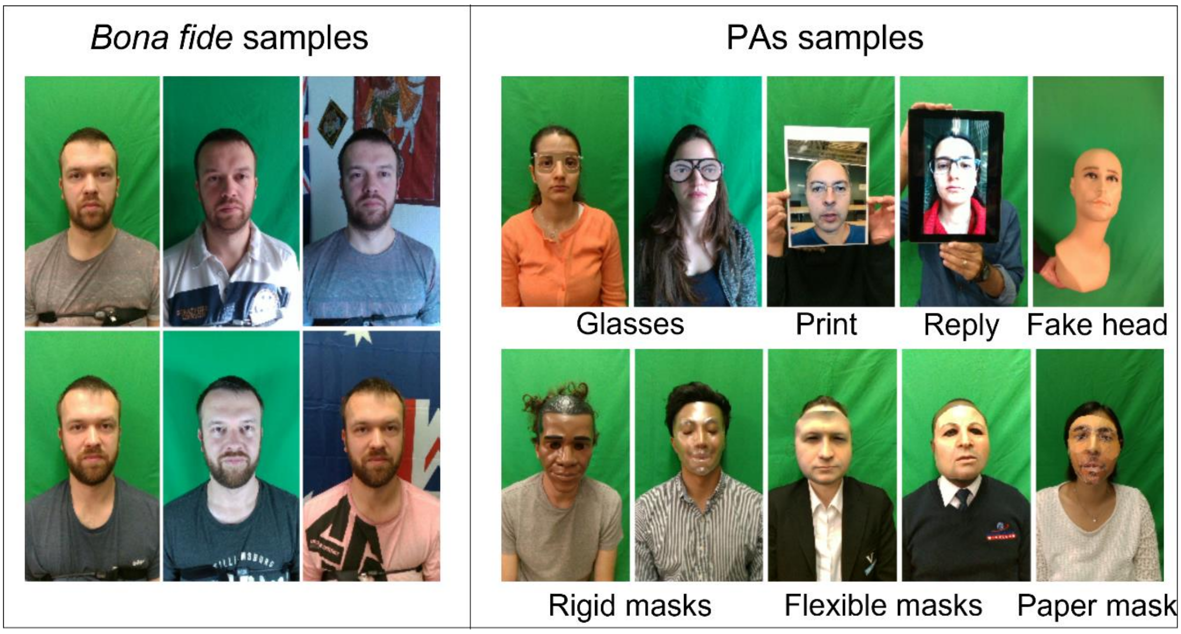

Fig. 3. Examples of bonafide data across six sessions and various PAs. [5].

Table 1. The WMCA dataset's main statistics [5].

Type	#Video
Bonafide	347
Glasses	75
Print	200
Replay	348
Fake head	122
Rigid mask	137
Flexible mask	379
Paper mask	71

4.2. Protocol

The model was trained on 1679 photos from the WMCA database, with an additional 83950 images added by including about 50 frames per video [5]. This expanded dataset was partitioned into training and testing subsets, maintaining a 70:30 ratio. This division ensured sufficient data for the face discrimination model's performance evaluation. In precise proportions, 70% of the data was designated as the training set, and the remaining 30% was incorporated into the test set. Only 50 uniformly sampled frames from each video were chosen, treating each frame as a separate sample. A biometric sample consists of geographically and temporally aligned frames from each of the four modalities [5], [24]. The proposed network underwent 50 epochs of training using an Adam optimizer and a batch size of 100.

The proposed multi-modal fusion method's performance was assessed using cross-validation, specifically the k-fold cross-validation technique. Estimating a model's performance on unseen data and assessing its generalizability are frequent uses of this method in machine learning. Each cycle of k-fold cross-validation uses one of the k subsets of the dataset to train the model and then evaluates its performance on the remaining fold. The validation set is utilized precisely once for each fold, and the operation is repeated k times.

4.3. Channel Selection

In this work, four modalities are use: color, depth, infrared, and thermal. The color modalities effectiveness against replay and print attacks is limited, so the depth modality is employed as a reliable alternative.

Furthermore, the incorporation of infrared and thermal modalities serves to further enhance the effectiveness of the approach. Various combinations of these modalities were tested and evaluated, with an emphasis on their ability to discriminate between real and fake images.

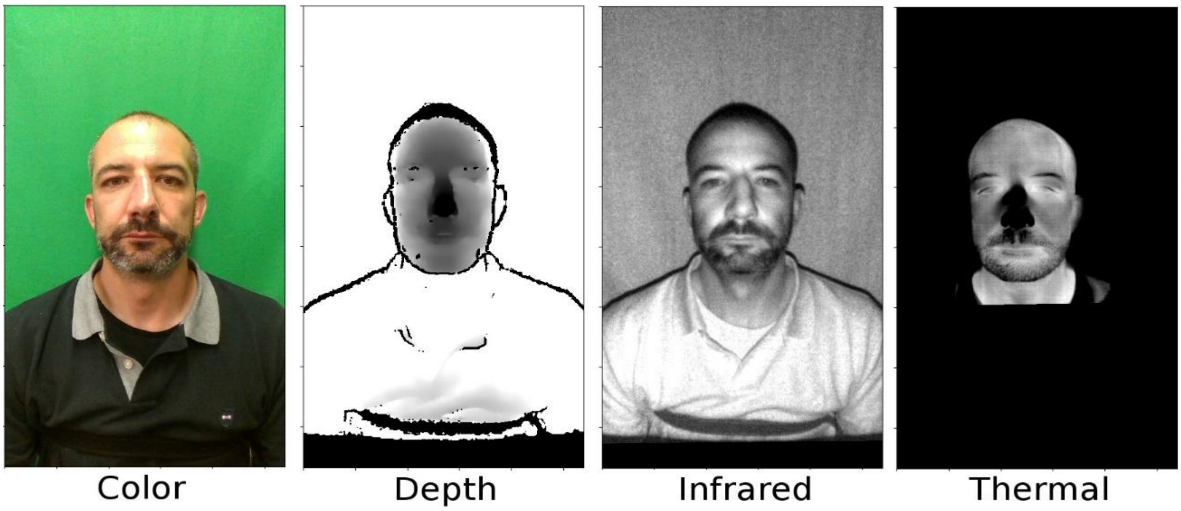

Fig. 4. The WMCA dataset sample of four different modalities [5].

Notably, the integration of all four modalities yields substantial performance improvements. Fig. 4 shown the four different images.

- **Color Images:** RGB images, standard in digital imaging, consist of red, green, and blue color channels per pixel, portraying a broad spectrum of colors for visual perception and display purposes.
- **Depth images:** Depth images convey object distances from the camera rather than color data, commonly applied in 3D reconstruction, augmented reality, and gesture recognition.
- **Infrared (IR) Images:** Infrared imaging captures the infrared radiation emitted by objects. IR images detect infrared wavelengths that are beyond the visible spectrum.
- **Thermal Images:** Thermal images capture the heat emitted by objects and represent it as a visual map of temperature distribution. These images visualize temperature variations across objects or scenes.

4.4. Evaluation Metrices

In the field of biometrics, especially in the evaluation and performance assessment of biometric systems, particularly for assessing facial recognition and anti-spoofing systems, the metrics APCER, BPCER, and ACER are often used by the ISO/IEC 30107-3 standards [25]. They are closely associated with the components of TN (true negatives), FP (false positives), TP (true positives), and FN (false negatives) present in a confusion matrix [26]. These values serve as the foundation for the computation of the aforementioned metrics. Their application and significance can be described as follows [25]:

- APCER (Attack Presentation Classification Error Rate): APCER measures the error rate associated with incorrectly accepting presentation attacks as real-face images. In other words, it quantifies the system's susceptibility to spoofing attempts. The APCER is computed as:

$$APCER\ (\%) = \frac{FP}{FP + TN} \times 100 \quad (1)$$

To compute APCER, the system is tested with a set of known presentation attacks (fake images or videos) and determines the percentage of these attacks that are incorrectly classified as real.

- BPCER (Bona-fide Presentation Classification Error Rate): BPCER measures the error rate related to unjustified rejection of real-face images. It shows that the system is unable to distinguish between authentic users. The BPCER equation is:

$$BPCER\ (\%) = \frac{FN}{FN + TP} \times 100 \quad (2)$$

To compute BPCER, the system is tested with a set of real face images and determines the percentage of these real samples that are incorrectly classified as impostors or rejected.

- ACER (Average Classification Error Rate): ACER provides an overall assessment of the face recognition system's performance by combining both APCER and BPCER. It represents the average of the two error rates. The equation for ACER is:

$$ACER\ (\%) = \frac{APCER + BPCER}{2} \quad (3)$$

5. Results and Discussion

In order to combine different types of data effectively, this study processed each type separately using ResNet-50 and then merged the results using methods such as majority voting, weighted voting, average pooling, and stacking classifiers. The performance assessment of the fusion scenario was based on ACER. Table 2 indicates that the stacking classifier achieved the most favorable ACER ratio of 0.087%, showcasing its superior performance among the fusion techniques evaluated. As a result, the post-fusion CDIT technique is the best way to go. It combines the results of different processing methods to use different kinds of information and make face anti-spoofing work better. The combination of different fusion methods further enhances the overall success of the post-fusion CDIT stage. Comparatively, this study evaluated two algorithms: RDWT-Haralick-SVM, based on an SVM classifier [27], and the MC-CNN algorithm using an artificial neural network for feature extraction [28]. Reference [5] showcased the RDWT-Haralick-SVM algorithm achieving an ACER of 3.44%, while the MC-CNN method obtained 0.3% ACER using the CDIT metric. Meanwhile, reference [7] reported an ACER of 2.91% with a dataset of 1679 images and 1.18% with 83,950 images. Nevertheless, the proposed method demonstrated superior performance compared to these results, as presented in Table 3.

Table 2. Results of proposed method

Fusion Type	APCER %	BPCER %	ACER %
Majority voting	93.710	99.989	96.849
Weighted voting	99.968	100.0	99.984
Average / pooling	100.0	100.0	100.0
Stacking classifier	0.175	0	**0.087**

Table 3. The test set for the CDIT data and the entire dataset comprising 83,950 images.

Method	APCER (%)	BPCER (%)	ACER (%)
Basic feature set + SVM [7]	0.11	2.24	1.18
Basic feature set + RF [7]	3.23	6.50	4.87
RDWT-Haralick + SVM [5]	6.39	0.49	3.44
MC-CNN [5]	0.60	0	0.30
Proposed Method	0.17	0	**0.08**

To test how well the suggested multi-modal fusion method worked, a k-fold cross-validation with k = 5 was used to split the dataset into five groups. During each iteration, four subsets were utilized for training, reserving one subset for validation. The method was iterated five times, guaranteeing that each subset served as the validation set once only. The model achieved 99% accuracy, indicating its robustness. The high accuracy achieved on unseen data indicates the fusion method's practical utility in real-world face anti-spoofing applications. Its ability to make accurate predictions on new and unseen samples demonstrates its effectiveness in preventing face-spoofing attacks, establishing its reliability. The area under the curve (AUC) and receiver operating characteristic (ROC) curves were also used to test the fusion method. The AUC value was 99%, as shown in Fig. 5. This signifies outstanding discriminative power, distinguishing between genuine and spoofed faces with exceptional accuracy. A high AUC value of 99% signifies near-perfect performance in differentiating real from fake facial images. Generally, a proficient model's ROC curve shows a steep rise towards the upper-left corner, showing a high rate of correctly identified positives and a low rate of falsely identified positives, so confirming the strength and precision of the fusion approach.

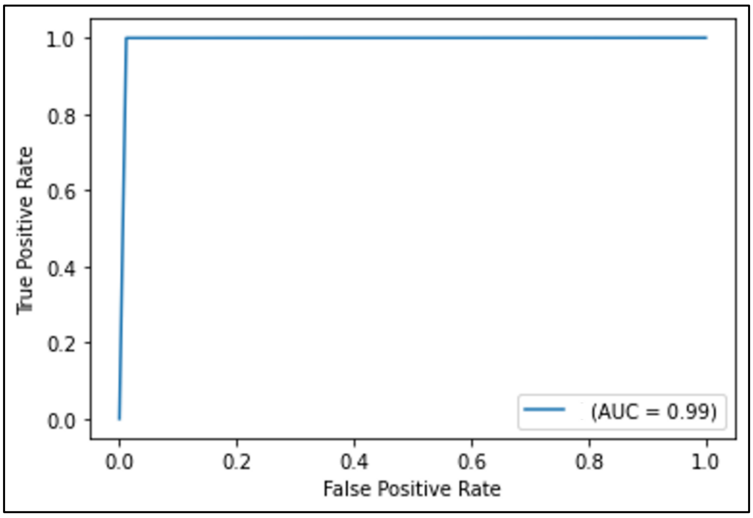

Fig. 5. ROC curve of proposed method.

6. Conclusion

The paper presents a technique designed to identify facial presentation attacks by employing multi-modal images and utilizing a RESNET-50 CNN-based anomaly detection model. The objective of this approach is to discern various attack modes while ensuring a high level of generalization performance. To demonstrate its practical usability, the method is applied and evaluated using the WMCA dataset. Additionally, it introduces a novel face anti-spoofing method that utilizes RESNET-50 and integrates RGB, IR, depth, and thermal modalities. This method incorporates diverse fusion techniques, such as majority voting and stacking classifiers, to enhance the system's overall performance. In future endeavors, there exists a chance to enhance this study through the incorporation of sophisticated deep learning methods. Furthermore, expanding the scope of research beyond the current dataset and performing cross-dataset analyses shows potential. By applying the proposed algorithm to a variety of datasets, this approach seeks to assess the efficacy of multi-modal FAS techniques within a broader context.

7. References

[1] A. Kadhim, S. Al-Darraji, “Face Recognition System Against Adversarial Attack Using Convolutional Neural Network.,” Iraqi Journal for Electrical & Electronic Engineering, vol. 18, no. 1, 2022. Doi:https://doi.org/10.37917/ijeee.18.1.1.

[2] M. A. Mohammed, M. A. Hussain, Z. A. Oraibi, Z. A. Abduljabbar, V. O. Nyangaresi, “Secure Content Based Image Retrieval System Using Deep Learning,” J. Basrah Res.(Sci.), vol. 49, no. 2, pp. 94–111, 2023. Doi:https://doi.org/10.56714/bjrs.49.2.9.

[3] P. J. Phillips et al., “Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms,” Proceedings of the National Academy of Sciences, vol. 115, no. 24, pp. 6171–6176, 2018. Doi:https://doi.org/10.1073/pnas.1721355115.

[4] W. Liu, X. Wei, T. Lei, X. Wang, H. Meng, A. K. Nandi, “Data-fusion-based two-stage cascade framework for multimodality face anti-spoofing,” IEEE Trans Cogn Dev Syst, vol. 14, no. 2, pp. 672–683, 2021. Doi:https://doi.org/10.1109/TCDS.2021.3064679.

[5] A. George, Z. Mostaani, D. Geissenbuhler, O. Nikisins, A. Anjos, S. Marcel, “Biometric face presentation attack detection with multi-channel convolutional neural network,” IEEE Transactions on Information Forensics and Security, vol. 15, pp. 42–55, 2019. Doi:https://doi.org/10.1109/TIFS.2019.2916652.

[6] I. Z. Mukti, D. Biswas, “Transfer learning based plant diseases detection using ResNet50,” in 2019 4th International conference on electrical information and communication technology (EICT), IEEE, 2019, pp. 1–6. Doi: https://doi.org/10.1109/EICT48899.2019.9068805.

[7] A. Denisova, “An improved simple feature set for face presentation attack detection,” in Proc. 2022 WSCG Computer Science Research Notes conf.,pp. 16-23. Doi:http://hdl.handle.net/11025/49574.

[8] Y. Zheng, E. A. Essock, “A local-coloring method for night-vision colorization utilizing image analysis and fusion,” Information Fusion, vol. 9, no. 2, pp. 186–199, 2008.Doi:https://doi.org/10.1016/j.inffus.2007.02.002.

[9] J. Kittler, M. Hatef, R. P. W. Duin, J. Matas, “On consolidating classifiers,” IEEE Trans Pattern Anal Mach Intell, vol. 20, no. 3, pp. 226–239, 1998. Doi: https://doi.org/10.1109/34.667881.

[10] L. Kuncheva, “Combining pattern classifiers methods and algorithms. john wiley&sons,” Inc. Publication, Hoboken, 2004. Doi: https://doi.org/10.1198/tech.2005.s320.

[11] D. H. Wolpert, “Stacked generalization,” Neural networks, vol. 5, no. 2, pp. 241–259, 1992.Doi:https://doi.org/10.1016/S0893-6080(05)80023-1.

[12] X. Tan, Y. Li, J. Liu, L. Jiang, “Face liveness detection from a single image with sparse low rank bilinear discriminative model,” in Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part VI 11, Springer, 2010, pp. 504–517. Doi: https://doi.org/10.1007/978-3-642-15567-3_37.

[13] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, S. Z. Li, “A face antispoofing database with diverse attacks,” in 2012 5th IAPR international conference on Biometrics (ICB), IEEE, 2012, pp. 26–31. Doi: https://doi.org/10.1109/ICB.2012.6199754.

[14] Z. Boulkenafet, J. Komulainen, L. Li, X. Feng, A. Hadid, “OULU-NPU: A mobile face presentation attack database with real-world variations,” in 2017 12th IEEE international conference on automatic face & gesture recognition (FG 2017), 2017, pp. 612–618. Doi:https://doi.org/10.1109/FG.2017.77.

[15] W. R. Almeida et al., “Detecting face presentation attacks in mobile devices with a patch-based CNN and a sensor-aware loss function,” PLoS One, vol. 15, no. 9, p. e0238058, 2020.Doi:https://doi.org/10.1371/journal.pone.0238058.

[16] A. Liu et al., “Contrastive context-aware learning for 3d high-fidelity mask face presentation attack detection,” IEEE Transactions on Information Forensics and Security, vol. 17, pp. 2497–2507, 2022. Doi: https://doi.org/10.1109/TIFS.2022.3188149.

[17] M. Fang, M. Huber, N. Damer, “Synthaspoof: Developing face presentation attack detection based on privacy-friendly synthetic data,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 1061–1070.

[18] O. Nikisins, A. Mohammadi, A. Anjos, S. Marcel, “On effectiveness of anomaly detection approaches against unseen presentation attacks in face anti-spoofing,” in 2018 International Conference on Biometrics (ICB), IEEE, 2018, pp. 75–81.Doi:https://doi.org/10.1109/ICB2018.2018.00022.

[19] E. Nesli, S. Marcel, “Spoofing in 2d face recognition with 3d masks and anti-spoofing with kinect,” in IEEE 6th International Conference on Biometrics: Theory, Applications and Systems (BTAS’13), 2013, pp. 1–8. Doi: https://doi.org/10.1109/BTAS.2013.6712688.

[20] A. Liu, Z. Tan, J. Wan, S. Escalera, G. Guo, S. Z. Li, “Casia-surf cefa: A benchmark for multi-modal cross-ethnicity face anti-spoofing,” in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2021, pp. 1179–1187.

[21] I. Chingovska, N. Erdogmus, A. Anjos, S. Marcel, “Face recognition systems under spoofing attacks,” Face Recognition Across the Imaging Spectrum, pp. 165–194, 2016.Doi:https://doi.org/10.1007/978-3-319-28501-6_8.

[22] S. Bhattacharjee, A. Mohammadi, S. Marcel, “Spoofing deep face recognition with custom silicone masks,” in 2018 IEEE 9th international conference on biometrics theory, applications and systems (BTAS), IEEE, 2018, pp. 1–7. Doi: https://doi.org/10.1109/BTAS.2018.8698550.

[23] K. Zhang, Z. Zhang, Z. Li, Y. Qiao, “Joint face detection and alignment using multitask cascaded convolutional networks,” IEEE Signal Process Lett, vol. 23, no. 10, pp. 1499–1503, 2016.Doi:https://doi.org/10.1109/LSP.2016.2603342.

[24] H. S. Mahmood, S. Al-Darraji, “A Multi-Modal Convolutional Neural Network for Face Anti-Spoofing Detection,” Iraqi Journal for Electrical and Electronic Engineering, 2024.

[25] I. Standard, “Information technology–biometric presentation attack detection–part 3: testing and reporting,” International Organization for Standardization: Geneva, Switzerland, 2017.Doi:https://doi.org/10.1145/3038924.

[26] H. Wu, F. J. Meng, “Review on evaluation criteria of machine learning based on big data,” in Journal of Physics: Conference Series, IOP Publishing, 2020, p. 052026.Doi:https://doi.org/10.1088/1742-6596/1486/5/052026.

[27] K. Erubaar Ewald, L. Zeng, Z. Yao, C. B. Mawuli, H. Sani Abubakar, A. Victor, “Applying CNN With Extracted Facial Patches Using 3 Modalities To Detect 3d Face Spoof,” 2020.Doi:https://doi.org/10.1109/ICCWAMTIP51612.2020.9317329/20/$31.00.

[28] X. Wu, R. He, Z. Sun, T. Tan, “A light CNN for deep face representation with noisy labels,” IEEE Transactions on Information Forensics and Security, vol. 13, no. 11, pp. 2884–2896, Nov 2018. Doi: https://doi.org/10.1109/TIFS.2018.2833032.

Journal of Basrah Researches (Sciences) **50**(1), 74 (2024)
DOI: https://doi.org/10.56714/bjrs.50.1.7

اكتشاف مكافحة انتحال الوجه باستخدام شبكة CNN متعددة الوسائط المعززة بواسطة ResNet

هالة شاكر محمود*[1]، صلاح فليح فالح [2]

[1] قسم علوم الحاسوب، كلية التربية للعلوم الصرفة، جامعة البصرة، البصرة، العراق.
[2] قسم علوم الحاسوب، كلية علوم الحاسوب وتقنية المعلومات، جامعة البصرة، البصرة، العراق.

معلومات البحث

الاستلام	9 كانون الأول 2023
القبول	19 شباط 2024
النشر	30 حزيران 2024

الكلمات المفتاحية

مكافحة انتحال الوجه، كشف هجوم العرض التقديمي، PAD، التعرف على الوجه، ResNet-50، التعلم العميق.

Citation: H.S. Mahmood, S. Al-Darraji, J. Basrah Res. (Sci.) **50**(1), 74 (2024). DOI:https://doi.org/10.56714/bjrs.50.1.7

الملخص

إن الانتشار المتزايد لتقنية التعرف على الوجوه في مختلف التطبيقات، بما في ذلك الأجهزة المحمولة والتحكم في الوصول والمعاملات المالية، يسلط الضوء على أهميتها. ومع ذلك، فقد تم إثبات ضعف أنظمة التعرف على الوجوه أمام الهجمات، مما يؤكد ضرورة معالجة نقاط الضعف المحتملة التي قد يستغلها المهاجمون. تتعمق الورقة في الكشف عن هجوم عرض الوجه (PAD) ضمن أنظمة القياسات الحيوية، وهو أمر بالغ الأهمية لضمان موثوقية وأمن خوارزميات التعرف على الوجوه. ولمعالجة هذه المشكلة، تقترح الورقة طريقة للكشف عن هجوم عرض الوجه باستخدام ResNet-50 بالتزامن مع البيانات متعددة الوسائط، بما في ذلك RGB والعمق والأشعة تحت الحمراء (IR) والقنوات الحرارية. تستكشف الطريقة استراتيجيات متنوعة للجمع بين النتائج من كل طريقة، والتحقيق في تقنيات الدمج المختلفة مثل تصويت الأغلبية، والتصويت المرجح، والتجميع المتوسط، ومصنف التراص. تم اختبار النظام على مجموعة بيانات WMCA. إنه يعرض أداءً قويًا مقارنة بالطرق الحالية، ولا سيما تحقيق نسبة ACER مثيرة للإعجاب تبلغ 0.087% مع مصنف التراص. وقد أثبت هذا النهج فعاليته من خلال توحيد طرائق متعددة دون الحاجة إلى نماذج فردية خاصة بسيناريوهات محددة، مما يشير إلى تطبيقات واعدة في العالم الحقيقي.

***Corresponding author email:** hala.shaker@uobasrah.edu.iq

ISSN: 1817-2695 (Print); 2411-524X (Online)
Online at: https://jou.jobrs.edu.iq