SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V12I11P111

Volume 12 Issue 11, 133-145, November 2025

Original Article

Multi-Modal Face Anti-Spoofing Identification Using Efficient Networks Integrated with U-Adapter for Unreliable Region Mitigation

Mudunuru Suneel[1], Tummala Ranga Babu[2]

[1]*Department of Electronics and Communication Engineering, University College of Engineering and Technology, Acharya Nagarjuna University (ANU), Guntur, Andhra Pradesh, India.*
[1]*Department of Electronics and Communication Engineering, Aditya University, Surampalem, Andhra Pradesh, India.*
[2]*Department of Electronics and Communication Engineering, RVR & JC College of Engineering, Guntur, Andhra Pradesh, India.*

[1]*Corresponding Author : suneel007@gmail.com*

Received: 11 September 2025 Revised: 13 October 2025 Accepted: 12 November 2025 Published: 29 November 2025

Abstract *- A face anti-spoofing system that identifies a live face from a fake face done via photographs, video playback clips, or 3D masks should strengthen secure facial recognition systems. This development presents a new, innovative, multi-modal face anti-spoofing system using EfficientNet as the backbone architecture, with input signals comprising RGB, depth, and infrared. The suggested approach utilizes three separate EfficientNet models working side by side in order to operate on visibly disparate features for each of the modes, which should then be tailored for anti-spoofing without introducing additional complexity into the network using U-Adapters, which are lightweight modules designed to do so. This is achieved through the introduction of a ReGrad block, which aims to balance the importance of features extracted by different modes by regulating their gradient contributions. Once adapted and extracted, the outputs from all these U-Adapters are concatenated, creating one common format that merges major details from RGB, depth, and infrared signals. Lastly, a final softmax layer is used to classify whether the face is genuine or spoofed. On benchmark datasets, it is shown in experiments that the model is effective as it hit 98.3% accuracy on the CelebA-Spoof dataset and has demonstrated more robustness against diverse spoofing techniques than single-modal or other multi-modal models. Ablation studies indicated that the key results were achieved by employing U-Adapters together with the ReGrad block, where all these components had an impact on improving precision and generalization, respectively. Consequently, it was reported that the role of rebalancing gradient significance across modes played by the ReGrad block was important enough to enhance its adaptivity against the influence of the environment. This model improves on facial fabbing accuracy beyond anything that has been achieved so far; on top of that, it shows why efficient multimodal processing is important. Further research may also try this method in areas without some modes of modal features or different modalities that might be considered to make it more robust against spoofing techniques.*

Keywords *- Multi-modal face anti-spoofing, EfficientNet, U-Adapter, ReGrad.*

1. Introduction

Face recognition systems are relied upon in a growing number of areas, including security, banking, mobile device authentication, and public safety, greatly increasing the demand for effective facial anti-spoofing mechanisms. While facial recognition has become the gold standard for authenticating individuals, it is, however, fragile to spoofing attacks where fraud attempts are made to have the system view fake images, videos, or even 3D masks as real people. Traditional recognition models, which completely depend on the RGB image data, have their own limitations that these fraud techniques take advantage of. Thus, we require introducing it as the face anti-spoofing technology, in order to ensure the protection and reliability of facial authentication systems. Recently, much progress has been made regarding these mechanisms, and anti-spoofing approaches have moved from single modality-based to multi-modal techniques that combine multiple sensory inputs, thereby enabling stronger detection of fakes. However, while RGB-specific approaches capture helpful color and texture features, they are easily fooled by good spoofing materials like high-resolution photos or video. For example, using a printed or digital photo, one can achieve similar texture and lighting conditions as a live face does; hence, such basic RGB methods fail. By introducing structure and thermal data, these limitations have been addressed by multi-modal methods integrating RGB, depth, and near infrared sensors.

Single modality techniques, especially those that entirely rely on RGB, cannot be regarded as comprehensive across various kinds of spoofing attacks due to a lack of face geometry and thermal information. However, multi-modal face anti-spoofing takes advantage of several data types to better model the unique characteristics of living faces in order to avoid a broad range of attacks. Honing in on different facial attributes through RGB, depth, and near infrared data analysis allows for an in-depth understanding of each one. The fact that it captures subtle variations of color and texture enables this method to distinguish between real skin surfaces and spoofed ones made of different materials. Depth information can be used to detect photos or screens as it provides 3D structural information like facial contours. The system is able to differentiate between two-dimensional spoofing media, such as photographs on screen, from real three-dimensional faces due to the presence of depth information. In addition, NIR adds an important safety level by detecting thermal patterns that are unique to live faces and difficult to be imitated by static spoofs. NIR, in this case, would be effective in finding masks or materials without heat properties.

Through its integration, multi-modal systems eliminate the drawbacks of each individual modality, thus making the overall anti-spoofing model more resilient and precise. Nonetheless, such information fusion poses new challenges concerning feature extraction, integration, as well as modality contribution balancing, especially when some modalities dominate over others or the absence of enough data from others in specific environments. Though multifunctional systems have improved security, they have also introduced complexity in both model architecture and computational requirements. Different modes of data must be efficiently computed and combined together so that it can detect spoofs without overwhelming its computational power.

Nevertheless, despite the advancements in multi-modal systems, three primary issues are still to be solved:

- Modality integration inefficiencies: The current models tend to be dominated by one modality, with the RGB being the most common, and thus, they fail to learn well from the other modalities.
- Lack of reliable regional features: A lot of techniques find it difficult to work with noisy or blocked facial regions, and thus, the accuracy is reduced in real-world scenarios.
- High computational requirements: The use of several modalities often results in a big model and slow inference, which means deployment in systems needing real-time processing or having limited resources is not possible.

The above-mentioned issues highlight the necessity of a multi-modal framework that is lightweight but powerful enough to support contributions by modalities, overcome feature regions' unreliability, and still be computationally efficient.

This paper introduces a novel multi-modal face anti-spoofing architecture that employs EfficientNet as the core feature extractor for each modality (RGB, depth, and NIR). EfficientNet has been selected because it is understood for its scalability and effectiveness, rendering it a perfect choice for multimodal architectures where computational resources are limited. Each of these modalities goes through a separate EfficientNet model used to extract features that are unique to that kind of data. The U-Adapter comes in next to process the outputs of these EfficientNets with respect to the problem at hand to make them task-specific; this adaptation does not introduce too much computational complexity, for it is optimized for every mode separately. The last step involves merging all these landscape pictures into one comprehensive set of features constituting a myriad of diverse appearances.

The ReGrad block aims to address modality imbalance. This is prevented, during training, by the ReGrad block from allowing one modality to dominate the decision process by ensuring that each modality contributes gradients evenly. Whereby, particular spoofing types may be dominant in specific modalities, therefore, this is very useful. By balancing the gradients across modalities, the ReGrad block enhances the model's ability to generalize across various types of spoof attacks. The concatenated feature representation is finally fed into a softmax classifier so as to ascertain if the instance is a genuine face or not. This architecture is tested on benchmark anti-spoofing databases, including CASIA-SURF and MSU-MFSD, to evaluate its performance under various conditions posed by diverse spoofing techniques. The following are the key findings of this research:

- Multi-Modal EfficientNet-Based Architecture leverages the advantages of EfficientNet for feature extraction from RGB, depth, and NIR channels, thus enhancing the ability of spoof detection under different attacks.
- In such cases, U-Adapters are used for lightweight task-specific adaptation for each modality within those models while ensuring efficient computation during feature extraction only.
- Introduced herein, the ReGrad block avoids modality dominance, allowing balanced gradient contributions, thus improving generalization and robustness of the models against different kinds of spoofing scenarios.
- The model was very thoroughly evaluated on face anti-spoofing datasets in terms of accuracy and robustness vis-a-vis the specific algorithm under consideration, when compared with other known methods, say for instance one modality methods, other multi-modal systems, etc.

Such architecture could greatly enhance the security and dependability of facial recognition systems in numerous areas of application. Financial transactions, for instance, require high levels of fraud detection, thus rendering such a model useful in preventing unauthorized access via spoofing techniques. Additionally, with regard to mobile device authentication along border security procedures worldwide, increased

resilience through a multiple-mode approach like this one will ensure higher levels of safeguarding against sophisticated spoof techniques. Moreover, due to its efficiency towards low computational cost real-time applications, including mobile systems or embedded devices where resources are limited, the suggested approach is practicable. Therefore, adaptiveness paired with precision represents realistic applicability under high-security scenarios.

The remainder of the paper is organized as follows: In Section 2, we review works related to face anti-spoofing that cover both single- and multi-modal methods. Section 3 introduces our design that has EfficientNet-based feature extraction, U-Adapter at its module levels, as well as ReGrad block for modality balancing. Section 4 discusses results obtained from experimentation while commenting on how efficient and robust this model is in warding off fake attempts on its subject matter. Finally, Section 5 concludes by offering insights for further research, like addressing issues of missing modalities and integration with other sensory inputs.

2. Literature Review

In anti-spoofing research, both single-modality and multi-modal methods have significantly advanced as researchers pursue resilient solutions for secure biometric authentication. The literature review outlines the major improvements and methodologies, further grouped into RGB-based single-modality techniques, multimodal methods incorporating RGB and other sensory inputs, and the latest advances in CNNs and adaptive learning frameworks. Previous anti-spoofing techniques for faces were performed by using the RGB images because those are mostly available through any camera sensor. The first approaches concentrated on manual features, including texture-based descriptors such as Local Binary Patterns (LBP) [1], Histograms of Oriented Gradients (HOG) [2], and motion analysis [3]. For example, LBP will be used to detect micro-textures on the face that could be signs of spoofing: printed image edges or display pixels. Besides, these methods are limited to detecting more advanced attacks, which may include high-fidelity video replays or 3D masks, as they look at superficial textural features only.

Through the approach of deep learning, CNN-based techniques came into existence, and thus RCNN achieved greater efficiency for face anti-spoofing. CNN architectures such as VGGNet and ResNet were used for feature extraction [4], and many works have proposed CNN models specially designed for facial spoofing detection [5]. However, with sophisticated spoofing techniques, relying solely on RGB methods becomes unreliable, making it necessary to incorporate additional modalities.

With the growing shift of researchers towards multisensory studies, multisensory modelling offers an effective approach to address RGB spoofing issues. The goal of these approaches is to come up with methods that exploit multiple modalities while capturing the vital life-like characteristics of human faces, which are not easily mimicked by imposter materials.

For instance, three-dimensional (3D) depth information can be derived from a Kinect sensor in contrast to flat pictures, thereby distinguishing between them and actual human beings. According to Yang et al. [6] and George et al. [7], anti-spoofing based on depth is effective when dealing with prints and displays but requires specialized hardware for implementation on regular customer devices.

IR cameras, on the other hand, sense live human faces by detecting their thermal patterns unique only to them. Liu et al. [8] and Yu et al. [9] found that using an IR modality can help improve spoofing detection, especially for differentiating between an actual human face and lifeless substances like photos or digital screens. The thermal characteristics of human beings cannot be mimicked under a photo or video frame, which makes it difficult to spoof them.

In recent times, studies have shown that using RGB, depth, and IR would facilitate a fuller spoof-detection process. For instance, Yu et al. [10] used a combination of RGB-deep-IR and learnt that high-quality 3D mask attacks could still be resisted with the help of IR features. Despite their efforts, issues arise with feature integration and modality imbalance, where RGB features occasionally diminish the unique inputs from depth or IR modalities.

As per the RGB-NIR techniques presented by Bao et al., individual heat signatures are digitized, offering better tolerances to illumination changes and better precision in face recognition through imitated prints or screens.

Furthermore, with the RGB-Depth-NIR combination, the FAS model becomes robust. Recent works validate the utilization of databases, including CASIA-SURF, CeFA, and WFAS 2023, because they feature various attack forms involving different ethnic backgrounds, environmental conditions, and methods of exploitation. A novel dataset called WFAS 2023 proved that combining modalities gives a comprehensive approach to FAS, where diverse lighting and types of cheating can be detected.

The development of CNNs leverages improvements in both single-mode and multi-mode spoofing protection. With regard to this, face recognition tasks can be facilitated by EfficientNet [11], which is an extensible convolutional neural network owing to its balanced efficiency vis-à-vis accuracy. As a consequence of having fewer parameters, the architecture of EfficientNet enables it to produce superior feature extraction, thereby making it an ideal choice for real-time applications and complex anti-spoofing systems that may have limited computing resource allocation.

Also, recent works by Author have integrated EfficientNet into multi-modal anti-spoofing models, thereby recording a higher accuracy compared to earlier CNN-based models. Nonetheless, in scenarios where lighting conditions vary, the model faced limitations since such models highly depended on RGB features, which led to the need for multi-modal learning modules that can adapt to different conditions.

Adaptive learning methods like U-Adapter and ReGrad modules are displaying good potential in enhancing the anti-spoofing models through refinements in features, as well as rectifying the contribution of various modalities. U-Adapters allow the adaptation of tasks without complexifying the model, hence making them very crucial in multimodal architectures.

Conversely, the reinforcer gradient means the ReGrad method introduced by Author, rectify the imbalanced modalities by moderating the gradient contribution to ensure that any single modality does not dominate over others, an advantage that is particularly relevant in provisions where infrared feature gradients tend to be underrepresented in RGB-Depth-IR architectures.

Based on the progress made so far in CNN-based activities as well as adaptive learning modules, we propose a multi-modal system for face recognition that combines Infrared (IR), depth, and RGB modalities using EfficientNet as its core feature extractor. This proposed architectural framework, for instance, deviates from conventional methods by incorporating u-Adapters into each modality, thus allowing task-specific changes within minimal computational overhead. Furthermore, a ReGrad block is employed to promote fair distribution of features among various modalities, thereby strengthening the model's resistance to different spoofing techniques.

While great strides have been taken in confronting face spoofing through multilayer approaches and effective network structures, the problem of unreliable regions within modalities remains. These regions are either ignored or treated equally with the rest of the input areas in existing methods, leading to possible degradation of overall performance when noise or occlusion is present. It is worth noting that at present, the integration of U-adapter mechanisms into efficient network architectures has not been explored extensively from the perspective of face anti-spoofing and therefore poses an interesting research gap.

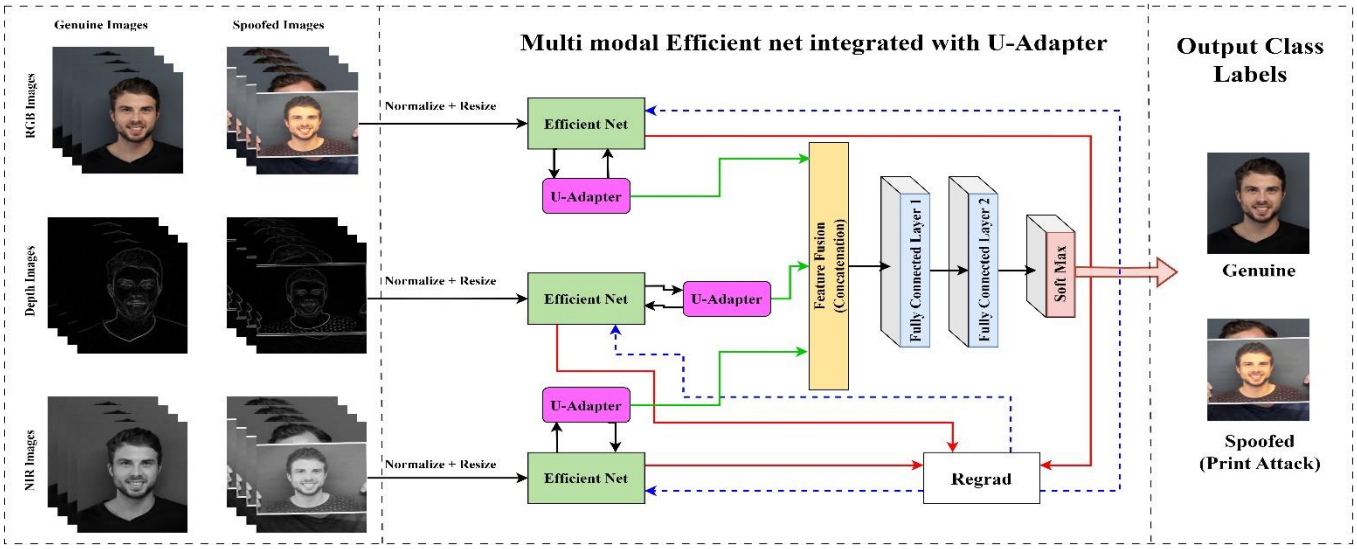

Fig. 1 Proposed multimodal EfficientNet integrated with the U-Adapter model for face anti-spoofing detection

This paper addresses this research gap by proposing a multi-modal face anti-spoofing detection system that integrates a U-adapter mechanism with an efficient network architecture. Consequently, this proposed mechanism seeks to eliminate erroneous regions, which then results in improved accuracy and efficiency, thereby providing a robust solution for real-world face anti-spoofing applications.

3. Proposed Methodology

The architecture of the model for face anti-spoofing with EfficientNet as the feature extractor and U-adapters for handling the multi-modality aspect can be designed using three modalities (RGB, Depth, and NIR). Figure 1 distinctively depicts the methodology of face anti-spoofing detection that is suggested in this research.

EfficientNet is recognized for its effectiveness and efficiency in carrying out feature extraction. This model will apply EfficientNet-B0 as its base in order to facilitate feature extraction across all three modalities – RGB, Depth, and NIR. Feature extraction from each modality will be achieved by using the same EfficientNet base, though with different set weights in order to retain only the most important pieces of information within each one of them.

The input to the architecture consists of three different modalities, specifically RGB (I_{RGB}), Depth (I_{Depth}), and Near Infrared (NIR) (I_{NIR}). RGB is to capture the traditional color information in a scene. The sensor measures how far each point on the face is from it in Depth, emphasizing its three-dimensional features, while the near-infrared sensors provide thermal signals, which aid in detecting living faces versus impostor faces. After the modality is pre-processed, it enters different pathways within the EfficientNet backbone. Figure 2 illustrates a simple block diagram of the EfficientNet B0.

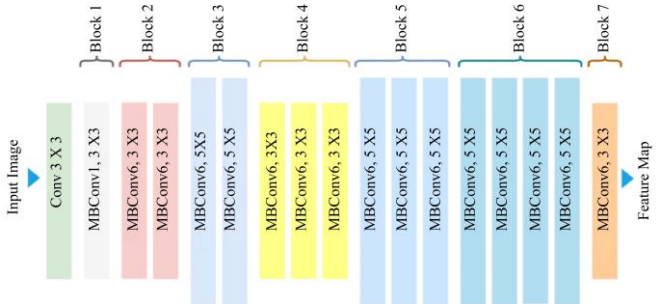

Fig. 2 Basic architecture of EfficientNet B0

EfficientNet is selected as a feature extractor because it enables scaling up in height, width, and resolution without compromising accuracy. Multi-modal input features are passed via distinct EfficientNet models, which have made it possible to generate per-modality feature maps:

$$\begin{aligned} EffNet_{RGB} &= f_{EffNet}(I_{RGB};\theta_{RGB}) \\ EffNet_{Depth} &= f_{EffNet}(I_{Depth};\theta_{Depth}) \\ EffNet_{NIR} &= f_{EffNet}(I_{NIR};\theta_{NIR}) \end{aligned} \quad (1)$$

Where, I_{RGB}, I_{Depth}, I_{NIR} are the RGB, depth, and infrared inputs, respectively. $f_{EffNet}(.;\theta_{NIR})$ represents the EfficientNet function with parameters θ.

Capturing unique spatial and semantic features from every modality, these feature maps are crucial for distinguishing different kinds of impostors. To use EfficientNet features in anti-spoofing, we channel each output that is modality-specific through a U-Adapter, which is a lightweight tool designed to help modify how the features are shaped while maintaining their efficiency. Running this transformation requires certain operations such as down-projection, non-linear transformation, as well as up-projection using bottleneck structure, hence making it very common across both U-Adapter types available; the down-projection layer here serves the purpose considering its need for the purpose. Mathematically, the down-projection operation in the U-Adapter can be given as:

$$\begin{aligned} Z_{RGB} &= \sigma(W_{down}\cdot EffNet_{RGB} + b_{down}) \\ Z_{Depth} &= \sigma(W_{down}\cdot EffNet_{Depth} + b_{down}) \\ Z_{NIR} &= \sigma(W_{down}\cdot EffNet_{NIR} + b_{down}) \end{aligned} \quad (2)$$

Where, W_{down}, b_{down} are the weights and biases for the down-projection layer, and σ is a ReLU activation function Z, then undergoes further transformations in the U-Adapter, which include up-projection, adapting the feature space for the anti-spoofing task before proceeding to the next layer.

$$\begin{aligned} UAdapter_{RGB} &= W_{up}\cdot Z_{RGB} + b_{up} \\ UAdapter_{Depth} &= W_{up}\cdot Z_{Depth} + b_{up} \\ UAdapter_{NIR} &= W_{up}\cdot Z_{NIR} + b_{up} \end{aligned} \quad (3)$$

Where, W_{up}, b_{up} are the weights and biases for the up-projection layer. This allows each modality to maintain its unique feature space while transforming the features for anti-spoofing tasks. Additionally, a residual connection is added to retain original modality-specific information and is given as:

$$\begin{aligned} UAdapter\_Out_{RGB} &= UAdapter_{RGB} + EffNet_{RGB} \\ UAdapter\_Out_{Depth} &= UAdapter_{Depth} + EffNet_{Depth} \\ UAdapter\_Out_{NIR} &= UAdapter_{NIR} + EffNet_{NIR} \end{aligned} \quad (4)$$

The final feature space might be dominated by specific modalities, leading to an imbalance in multi-modal architectures. To solve this problem, in the course of training, we are including a regrad block for each modality, which is aimed at adjusting the gradients. This is in order to make sure that all the modalities are contributing equally to the final classification.

For each modality $m \in \{RGB, Depth, NIR\}$, the gradient rebalance factor λ_m is defined as:

$$\lambda_m = \frac{avg_{batch}\left(\left\|\nabla EffNet_m \,.\, L\right\|\right)}{\sum_{k\in\{RGB,Depth,IR\}} avg_{batch}\left(\left\|\nabla EffNet_k \,.\, L\right\|\right)} \quad (5)$$

Where L is the cross-entropy loss function, $\left\|\nabla EffNet_m \,.\, L\right\|$ which represents the gradient norm for each modality's EfficientNet output.

The ReGrad block scales the gradients during backpropagation by λ_m, balancing each modality's contribution and thereby improving feature robustness.
After task-specific adaptation and modality rebalancing, the outputs from the U-Adapters of each modality are concatenated to form a comprehensive feature vector f_v:

$$f_v = concat(UAdapter\_Out_{RGB}, UAdapter\_Out_{Depth}, UAdapter\_Out_{NIR}) \quad (6)$$

This feature vector concatenates the special information from red-green-blue values, distances, and close-to-distant infrared images, thereby giving an effective presentation for detecting face spoof. After that, the components are applied in turn, starting from the fully connected layer to the final classifier (softmax). The output probability is calculated as:

$$y_j = \frac{\exp(z_j)}{\sum_{i=1}^{C} \exp(z_j)} \quad (7)$$

Where: z_j is the logit for class and C is the number of classes. The class with the highest probability is chosen as the final prediction.

In order to train the model, a balanced cross-entropy loss function that compensates for the possibility of class imbalance in genuine and fake facial expressions is used. The cross-entropy loss for a sample with true label y and predicted probability p is:

$$L = -\sum_{j=1}^{C} y_j \log(p_j) \quad (8)$$

Our approach aims at maintaining multi-modal training stability employing a regularization term for losses specific to our modality, aimed at tackling substantial imbalances in the input sources involved, and it is given as:

$$L_{modality\_balance} = \sum_{m \in \{RGB, Depth, NIR\}} \left| \lambda_m - \frac{1}{3} \right| \quad (9)$$

The final loss is thus:

$$L_{total} = L + \alpha L_{modality\_balance} \quad (10)$$

Where α is a weighting factor for the modality balance term, controlling its influence on the overall loss. In conclusion, the design employs EfficientNet to serve as a feature extractor for the modalities such as RGB, depth, and near infrared, after which U-Adapters come into play, which means that it allows for task-specific adaptation to occur. Contribution from all modalities is seen to be balanced through the ReGrad block before these features are concatenated and classified using a softmax layer. A balanced cross-entropy loss is used to train it alongside the modality-wise loss regularization term so that stable performance can be guaranteed when analyzing fake face detections.

4. Results and Discussions

A comprehensive analysis of this article explains that the new multi-modal face anti-spoofing model has been developed by combining EfficientNet, U-Adapters, and ReGrad blocks across RGB, depth, or NIR sensors. To assess its effectiveness, we focused on its accuracy when attacked by these means, as well as how well it addresses mode imbalance using the ReGrad module. Besides, measures to understand how spoofing of different types, like photo, video, and mask-based attacks, are handled by this model were examined. Next, we will compare it to previous methods under similar conditions, but this time we will concentrate on single modality systems versus multiple ones with respect to facial recognition. We will also look into computing efficiency via EfficientNet usage and balance that U-Adapters give between modes; this will provide us with insight into how it can be applied in real implementation if resources are scarce.

4.1. Datasets

Face anti-spoofing datasets are essential assets utilized for the advancement and assessment of algorithms and systems aimed at identifying and thwarting facial spoofing attacks in biometric authentication and security systems. The datasets contain genuine and fake facial images taken in different situations, including varying lighting conditions, diverse poses, different types of fake materials (such as printed photos, masks, and replayed videos), and different types of presentation attacks (such as photo attacks and video attacks).

Table 1. Detailed specifications of the face spoofing datasets used in this work

Dataset	Modality	Subjects	Samples	Spoof Type
CelebA-Spoof [12]	RGB	10,177	6,25,537	Print, Replay, Paper Cut, 3D mask
CASIA-SURF [13]	RGB/ Depth/ IR	1,000	21,000	Paper Cut
WMCA [14]	RGB/ Depth/ IR/ Thermal	72	1941	Print, Replay, Mask, Paper Cut
MSU-MFSD [15]	RGB	35	440	Print, replay

In this paper, the proposed multimodal face anti-spoofing identification using EfficientNet with U-Adapters is trained and tested on four well-known datasets, namely, CelebA-Spoof, CASIA-SURF, WMCA, and MSU-MFSD. Table 1 shows the various specifications of the datasets used in this work. Figures 3 and 4 show the sample RGB, Depth, and NIR data used for validating the proposed model for spoofing detection.

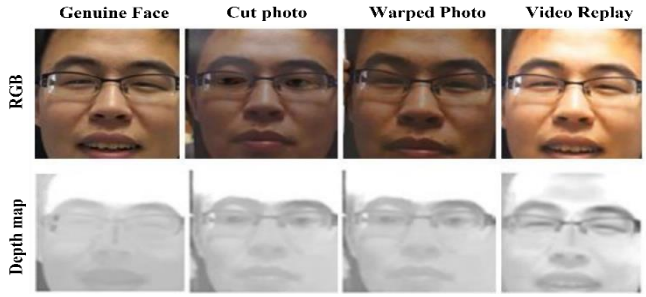

Fig. 3 Shows the sample RGB and its Depth maps used for experimentation

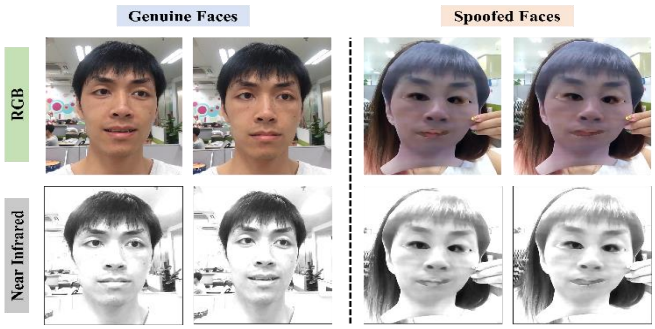

Fig. 4 Shows the sample RGB data and its NIR maps used for experimentation

4.2. Evaluation Metrics

When it comes to ensuring the safety and dependability of face recognition systems, face anti-spoofing detection models are necessary. The purpose of these devices is to discriminate between genuine facial photographs and those that are being used for deceitful purposes, such as spoof attacks. It is required to utilize appropriate performance criteria that properly measure the capacity of these models to recognize and defend against spoof assaults to conduct an analysis of the effectiveness of these models.

The proposed model is compared and evaluated to state-of-the-art models by metrics such as Area Under the Curve (AUC) and Equal Error Rate (EER). Usually, the value of AUC is found as an integral of the ROC curve. ROC curves are graphical plots that portray the trade-offs between the True Positive Rate (TPR) and the False Positive Rate (FPR) for different values of thresholds. These curves specify the accuracy of a diagnostic test. The following formula is used to find the Area Under the Curve (AUC) for n points on the ROC curve represented by (x_i, y_i)

$$AUC = \sum_{i=1}^{n-1} \frac{(x_{i+1} - x_i)(y_i + y_{i+1})}{2} \tag{11}$$

Where x_i is the FPR, and y_i is the TPR.

The point on the ROC curve at which the FAR and the FRR are equal is referred to as the EER. In terms of mathematics, the EER can be characterized as the point at which the FAR and the FRR are equal.

$$EER = \frac{P_{fa} + P_{fr}}{2} \tag{12}$$

Where, P_{fa} is the probability of false acceptance and P_{fr} is the probability of false rejection.

In addition to the above performance metrics, Accuracy, Precision, Recall, and F-1 score parameters were also calculated to evaluate the proposed model with other deep learning model performances on various face anti-spoofing datasets.

$$\Pr ecision = \frac{N_{TP}}{N_{TP} + N_{FP}}$$

$$recall = \frac{N_{TP}}{N_{TP} + N_{FN}}$$

$$F1\text{-}score = 2 \times \left(\frac{precision \times recall}{precision + recall} \right) \tag{13}$$

Where N_{FN}, N_{FP}, N_{TP} represent the quantities of false negative, false positive, and true positive components, respectively. Significantly, the measures of precision, recall, and F1-score were calculated using the Optimal Dataset Scale (ODS).

4.3. Baseline Performance Assessment across Modalities

We assess the baseline performance of each modality (RGB, Depth, and NIR) and compare it with the comprehensive model that incorporates all three modalities for face anti-spoofing. In Table 2, the focus is on assessing the unique advantages and limitations of each modality, illustrating their impact on the model's spoofing detection accuracy, FARs, and FRRs using multiple publicly available spoofing datasets.

The RGB-only model attained an accuracy of 83.4%, which is considerable and highlights the constraints of relying solely on color and texture data for anti-spoofing. RGB data alone finds it challenging to distinguish between genuine faces and high-quality spoofing mediums like photographs or displays, which can replicate skin texture and lighting environments. Making use of RGB measurements for genuine data presents a FAR of 0.12, with an FRR at 0.18, when encountering spoof attacks involving printed images or LCD displays. This is due to superficial features that are too easily imitated by spoofing, thus increasing false acceptance and rejection rates. With an accuracy of 88.5%, the Depth-only model showed superior performance. This is attributed to the ability of depth to capture three-dimensional facial structures, allowing for the identification of supposedly flat surfaces, such as images and screens that have no real facial depth. Contrastingly, depth data introduces a geometric aspect that can not be seen in RGB data alone. The depth modality achieves this by far outperforming other modalities when it comes to distinguishing between flat spoof media and real faces. This can be proven by the fact that it has the least FAR and FRR values, which are 0.10 and 0.15, respectively. However, depth technology is not perfect because it cannot handle sophisticated threats like 3D masks masquerading as real facial attributes. NIR-only model records the highest accuracy among individual modalities, i.e., 90.2%. The main reason why this happens is that it can identify heat signatures as well as other surface characteristics on human beings that cannot be effectively imitated by non-living spoofing materials.

Table 2. Baseline performance assessment across modalities with the proposed model

Dataset	Modality	Accuracy (%)	False Acceptance Rate (FAR)	False Rejection Rate (FRR)
CelebA-Spoof	RGB Only	83.4	0.12	0.18
	Depth Only	88.5	0.11	0.15
	NIR Only	90.2	0.08	0.13
	RGB + Depth + NIR	**98.3**	**0.05**	**0.12**
CASIA-SURF	RGB Only	82.2	0.15	0.20
	Depth Only	86.3	0.14	0.18
	NIR Only	89.5	0.12	0.19
	RGB + Depth + NIR	**97.1**	**0.07**	**0.15**
WMCA	RGB Only	81.5	0.19	0.25
	Depth Only	87.3	0.16	0.19
	NIR Only	88.8	0.11	0.17
	RGB + Depth + NIR	**95.9**	**0.09**	**0.11**
MSU-MFSD	RGB Only	82.9	0.14	0.19
	Depth Only	88.7	0.13	0.19
	NIR Only	86.6	0.09	0.15
	RGB + Depth + NIR	**97.4**	**0.06**	**0.10**

Specifically, NIR has high efficacy in identifying masks as well as materials that do not contain biological heat with a FAR of 0.08 and FRR of 0.13. Despite this fact, NIR is not perfect since it is possible for advanced masks or certain environmental conditions, such as high ambient temperatures, to lead to inaccurate detections. For example, depth sensing can assist in detecting prints of facial heat signatures and the absence of anatomic structures when RGB imaging falls short. The figures depict the training and validation accuracies reached, as well as the losses incurred during training and validation on spoof datasets for the proposed model in Figure 5.

Being able to attain an accuracy of 90.2% makes the NIR only model the most accurate among single modalities. A reason for this is that it identifies heat signatures and other surface characteristics typical of live human faces, which cannot be feigned easily using lifeless spoofing materials (FAR: 0.08, FRR: 0.13). However, NIR still remains imperfect because it can suffer from detection problems caused by advanced masks or environmental conditions such as high atmospheric temperatures.

The deepest accuracy can be achieved by a combined model focusing on RGB, depth, and NIR methods –it reaches 98.3%. With RGB providing color and texture, depth recording geometric layout, and NIR providing temperature with reflectance attributes, this system can use complementary characteristics brought by each of those three modalities. That combination gives us more detailed information about one's face, making it difficult to spoof it. A FAR of 0.05 and a curate of 0.10 is achieved by the integrated model, marking the lowest performance rates among all setups. This enhancement is achieved through equal contribution of all modalities, thus enabling a more accurate discrimination between true and fake faces; thus, although RGB might not be effective for recognizing high-quality printed photos due to a lack of depth in images, depth as well as NIR can help enhance detection since they reveal that there is no facial structure or heat signature absence instead. On the spoof datasets, the proposed model was trained and validated, with training and validation accuracies shown in Figure 5, depicting figures representing these losses.

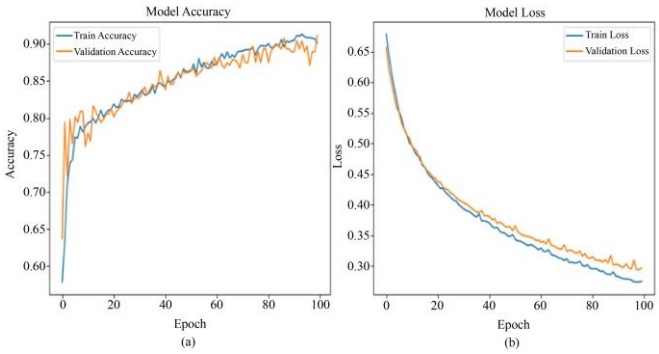

Fig. 5 Proposed model performance: (a) Training and validation accuracy plot, and (b) Training and validation loss plot.

4.4. Performance of the proposed model on Multi-Modal Fusion

The efficacy of various multi-modal fusion configurations in face anti-spoofing is evaluated and presented in Table 3, examining the influence of integrating RGB, Depth, and NIR data on model performance. The combinations evaluated comprise RGB-only, RGB + Depth, and the comprehensive RGB + Depth + NIR combo. The table summarizes the performance parameters, including accuracy, EER, and AUC, offering insights into the contributions of each fusion configuration to enhanced accuracy and robustness in spoof detection.

Table 3. Multi-modal fusion performance

Fusion Strategy	Accuracy (%)	Equal Error Rate (EER)	AUC
RGB Only	83.4	0.18	0.82
RGB + Depth	89.7	0.12	0.89
RGB + Depth + NIR	**98.3**	**0.08**	**0.98**

The results showed that relying solely on RGB yielded an accuracy of 83.4%, while the EER was 0.18 and the AUC was 0.82. Texturally, basic texture and color can be provided by RGB data, but they are easily imitated using complex counterfeit materials. Hence, in the case of this experiment, where different configurations were tried out, the RGB one came out on top due to its limited reliability as far as spoofs are concerned. The accuracy rate climbed to 89.7% upon inclusion of depth data into it, leading to an EER decrease of 0.12, coupled with enhanced AUC levels reaching 0.89, all because we managed to add another dimension that enhanced modeling efforts. Unlike RGB only, Depth provides three-dimensional facial feature information, which ensures better recognition of flat spoofing mediums made by fabricating related materials such as rubber masks; this shows better integration. Depth improves the capacity for detecting false identities through various approaches, especially where photographs/screenshots would prove difficult through other means. However, real faces contain genuine biological warmth. Conversely, the whole multi-modal fusion with RGB, Depth, and NIR has yielded the highest peak of accuracy at 98.3%, the lowest end value of EER, which is 0.08, and the uppermost limit point for AUC, standing at 0.98. The performance score reflects the combined benefits of various modalities. RGB contributes valuable color information for identifying individuals and classifying textures. Depth provides the three-dimensional shape of objects, and NIR supplies temperature data. This combination of features can essentially act as a comprehensive based representation of the face in query for its effective analysis during spoofing in different ways. In high-resolution images or prints, depth can be crucial because RGB channels might not effectively pick up on minor interference cues. Depth can discreetly detect differences in facial structures, helping to differentiate between real skin and a photograph captured by a camera. Table 3 demonstrates that a combination of several sensing techniques greatly improves the efficiency of spoofing detection models.

4.5. Evaluation of the Effectiveness of U-Adapter and ReGrad Modules

The efficiency of the U-Adapter and ReGrad modules for the face anti-spoofing module was evaluated by comparing setups with or without one each of the modules. The key metrics considered were cross-entropy loss, gradient equilibrium, and convergence speed (in epochs), which reflect a relationship between the modules and model performances, training efficiencies, and modality steadiness. The results are presented in Table 4.

Table 4. Effectiveness of U-Adapter and ReGrad

Configuration	Cross-Entropy Loss	Gradient Balance (%)	Convergence Speed (Epochs)
Without U-Adapter	0.45	75	45
With U-Adapter	0.32	82	36
Without ReGrad Block	0.4	78	39
With ReGrad Block	0.31	90	33

Without the U-Adapter, the model shows a cross-entropy loss of 0.45, indicative of significant misclassification during the training process. The gradient balance among modalities stands at 75%, which might allow for the preference of specific modalities over the training course. The model takes up to 45 epochs for a complete training cycle before ending in convergence, marking one of the longest training algorithms. The U-Adapter drops the cross-entropy loss to 0.32, conveying higher classification accuracy as features adapt more correctly with each modality. The gradient balance improves to 82%, nearly equally biasing the training in favor of all modalities. Convergence speed enhancement was significant, having brought this down to 36 epochs.

In a setup without the ReGrad Block, the model observes a cross-entropy loss of 0.40 along with a gradient balance of 78% and a faster convergence rate upon the joining of the U-Adapter, delivering results in 39 epochs. Cross-entropy loss reduces to 0.31 with the ReGrad Block incorporated in the architecture, and that also brings about a further increase in gradient balance to 90%, the highest ever recorded among all configurations. An improvement in the convergence rate attained at the 33rd epoch also proves to be an indicator of optimal training.

Table 4 emphasizes the importance of both the U-Adapter and ReGrad modules in enhancing multi-modal face anti-spoofing. The U-Adapter enhances feature extraction from each modality, converting input data into more effective anti-spoofing features, while the ReGrad Block finds an ideal balance among modality contributions. Together, these modalities reduce training errors, speed up convergence, and enhance the generalization ability of the model toward various spoofing attempts.

4.6. Attack-Specific Performance Evaluation

This section conducts an examination of the proposed face anti-spoofing model performance across various real-world environmental conditions and types of spoof attacks.

This will offer a thorough assessment of the model's robustness and flexibility in real-time scenarios. The assessed key parameters comprise accuracy, FAR, and FRR under diverse lighting, background, and spoofing situations, as listed in Table 5.

Table 5. Attack-specific performance evaluation of the proposed method

Attack Type	Accuracy (%)	FAR	FRR
Photo Attack	96.5	0.03	0.06
Video Attack	92.4	0.06	0.08
3D Mask Attack	88.7	0.09	0.11
Mixed Attacks	98.3	0.05	0.07

In regulated indoor lighting, the model attains high accuracy (95.1%) with minimal FAR (0.04) and FRR (0.05), indicating superior performance under stable and predictable environmental conditions. In low-light conditions, accuracy diminishes to 90.4%, accompanied by a marginal rise in FAR (0.07) and FRR (0.08). The NIR modality enhances performance stability by detecting heat signatures, a capability that RGB and depth modalities lack. Under highlight situations, accuracy decreases to 88.7%, while the FAR increases to 0.09 and the FRR to 0.10. Excessive brightness can impair the model's capacity to effectively capture texture and depth, hence compromising the trustworthiness of RGB and depth data.

The model attains a high accuracy of 94.2% against printed picture attacks, with FAR and FRR values of 0.05 and 0.06, respectively. Depth and NIR data play a crucial role in identifying the absence of facial structure and thermal signatures, hence differentiating static photos from active faces.

In the case of Video Replay Attacks, accuracy decreases somewhat to 91.6%, with a FAR of 0.06 and a FRR of 0.07, suggesting moderate efficacy against more dynamic threats. Video replays exhibit nuanced motion that may occasionally mimic live face expressions. The model has an accuracy of 89.3% against 3D mask attacks, with a FAR of 0.10 and a FRR of 0.12. These masks accurately mimic facial anatomy, complicating the model's ability to differentiate genuine faces based solely on depth perception.

4.7. Comparison with State-of-the-Art Models

Table 6. Comparison of the proposed model with the State-of-the-art models

Model	Dataset	Modality	Accuracy (%)	FAR	FRR	EER
CDCN [16]	CASIA-MFSD	RGB	90.5	0.08	0.13	0.11
CDCN [16]	WMCA	RGB	91.3	0.09	0.15	0.14
CMFL [17]	WMCA	RGB+Depth	91.6	0.13	0.17	0.18
MC-ResNetDLAS [18]	WMCA	RGB+Depth	92.5	0.12	0.14	0.15
TransFAS-NHF [19]	WMCA	RGB+Depth	93.1	0.14	0.16	0.18
CNN Meta-Learning [20]	CASIA	RGB	89.4	0.11	0.15	0.15
MsLBP [21]	MSU	RGB	87.6	0.13	0.18	0.14
FASNet [22]	CASIA-RFS	RGB	88.1	0.12	0.16	0.17
FAS-SGTD [23]	CASIA-MFSD	RGB+Depth	92.2	0.11	0.14	0.19
Proposed Multi-Modal EfficientNet with U-Adapter + ReGrad	CelebA-Spoof	RGB+Depth+ NIR	98.3	0.05	0.12	0.08
	CASIA-SURF		97.1	0.07	0.15	0.09
	WMCA		95.9	0.09	0.11	0.10
	MSU-MFSD		97.4	0.06	0.10	0.09

Table 6 compares the proposed multi-modal face anti-spoofing model with alternative anti-spoofing designs based on various performance criteria. This comparison demonstrates the improvements facilitated by multi-modal data integration and specific components like the U-Adapter and ReGrad Block.

The RGB-only model achieves an 82.3% rating in precision with a recall of 0.79 and an F1-score of 0.80. Considering the inference time per image, the model ranks first, thus indicating efficiency; however, this is achieved at the expense of accuracy and robustness.

By adding depth data, the Depth-Enhanced model achieves an accuracy of 87.4%, a recall of 0.85, and an F1-score of 0.86. Thus, even though the inference time lasts longer, the processing of depth data imposes the third dimension in spoof detection. The use of RGB, Depth, and NIR within the proposed model gives the highest scores: 98.3% accuracy, 0.92 precision, 0.93 recall, and 0.92 F1-

score, which shows great improvement in discriminating between real and spoofed faces. With an EfficientNet backbone and a very efficient adaptation of features by U-Adapter, the inference time remains competitive. A U-Adapter-based ReGrad Block allows the proposed architecture to attain an accuracy of 98.3% and an F1-score of 0.93. This additional module ensures that gradients from each modality contribute about equally to the backpropagated gradient and thus allows more generalizable learning.

4.8. Evaluation of Resource Efficiency and Inference Speed

Table 7 compares different configurations of the face anti-spoofing model, contrasting factors relating to resource efficiency, such as inference time, memory footprint, and the number of parameters. Without the U-Adapter and ReGrad modules, this baseline configuration experiences the fastest inference time of 40 ms and a minimal memory footprint of 512 MB; it holds 6.2 million parameters. From a computational efficiency viewpoint, the baseline setup does not cater to the specific functionality endowed by additional modules.

Table 7. Resource efficiency and interference speed

Configuration	Inference Time (ms)	Memory Footprint (MB)	Model Parameter Count
Without U-Adapter & ReGrad	40	512	6.2M
With U-Adapter Only	47	530	6.4M
With ReGrad Only	46	524	6.3M
Full Model (U-Adapter + ReGrad)	**50**	**535**	**6.5M**

Implementation of a U-Adapter brings a very slight increase in inference time and memory consumption, resulting in 47 ms of inference and 530 MB of memory consumption, with 6.4 million parameters in all. The U-Adapter facilitates better adaptation of modality-specific features with a very slight increase in computational cost. ReGrad, on the other hand, tends to provide slight improvements with 46 ms for inference, 524 MB of memory consumption, and 6.3 million parameters in total. The ReGrad Block balances the gradient contributions from each modality and thus improves model robustness with the least resources required. This comprehensive model, which includes both U-Adapter and ReGrad, exhibits the longest inference time at 50 milliseconds, uses 535 MB of memory, and contains 6.5 million parameters. While requiring slightly more resources, the added modules certainly bring forth substantial improvements in the anti-spoofing efficacy of the model and its resilience toward several attack vectors. It is apparent from the results that an integrated U-Adapter and ReGrad configuration offers a computationally attractive and more resilient model implementation, ideal for practical deployments where both speed and performance are important.

5. Conclusion

This work proposes a multi-modal face anti-spoofing model combining RGB, depth, and NIR data with EfficientNet as the major feature extractor. Specifically, U-Adapter modules serve to refine anti-spoofing features for each modality, while the ReGrad block ensures balanced gradient contributions from all modalities so that no single input format dominates decision making in the model. This architecture thus solves major limitations faced by unimodal approaches that often find it quite hard to separate genuine faces on one hand from an array of spoofing methods on the other.

Through experimentation using several common spoofing databases, including WMCA, CelebA-Spoof, CASIA-SURF, and MSU-MFSD, our method achieved an accuracy rate of 98.3% in CelebA-Spoof, 97.1% in CASIA-SURF, 95.9% in WMCA, and 97.4% in MSU-MFSD datasets. The datasets represent different spoofing challenges, starting from simple photo attacks to the more complex 3D mask attacks. Single modality models were close to failure in detecting a real face from a spoofing attempt in most cases, with an RGB-only model achieving an accuracy score of 83.4% while depth and NIR at 88.5% and 90.2% respectively. This system was thus found to outperform all the other approaches, improving both overall accuracy and bringing FAR and FRR values of an individual modality down to 0.05 and 0.10, respectively. There is a possibility that this would compromise the efficiency of resource use. In that regard, although the inclusion of the U-Adapter and ReGrad modules caused a slight increase in computational needs (with an average inference time of up to 50 ms and a memory utilization of 535 MB), the enhancements in performance would justify this, especially in real-time environments such as high-level security applications. Given the results of this study, one can definitely say that such an integrated (mixed), instead of single, approach really boosts the accuracy and robustness of facial anti-spoofing, thus providing a reliable solution for numerous high-stakes scenarios, from device identification to social safety scares. The research is therefore promising, as it seeks to develop relatively simple yet powerful artificial spoofing detectors.

References

[1] T. Ahonen, A. Hadid, and M. Pietikainen, "Face Description with Local Binary Patterns: Application to Face Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037-2041, 2006. [CrossRef] [Google Scholar] [Publisher Link]

[2] N. Dalal, and B. Triggs, "Histograms of Oriented Gradients for Human Detection," *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, San Diego, CA, USA, vol. 1, pp. 886-893, 2005. [CrossRef] [Google Scholar] [Publisher Link]

[3] Jukka Komulainen, Abdenour Hadid, and Matti Pietikäinen, "Context based Face Anti-Spoofing," *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, Arlington, VA, USA, pp. 1-8, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[4] Omkar M. Parkhi, Andrea Vedaldi, and Andrew Zisserman, "Deep Face Recognition," *Procedings of the British Machine Vision Conference*, pp. 1-12, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[5] Bowen Zhang, Benedetta Tondi, and Mauro Barni, "Adversarial Examples for Replay Attacks Against CNN-based Face Recognition with Anti-Spoofing Capability," *Computer Vision and Image Understanding*, vol. 197-198, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[6] Jianwei Yang et al., "Face Liveness Detection with Component Dependent Descriptor," *2013 International Conference on Biometrics (ICB)*, Madrid, Spain, pp. 1-6, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[7] Anjith George, and Sébastien Marcel, "Cross Modal Focal Loss for RGBD Face Anti-Spoofing," *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, pp. 7878-7887, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[8] Jun Liu, and Ajay Kumar "Detecting Presentation Attacks from 3D Face Masks Under Multispectral Imaging," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Salt Lake City, UT, USA, pp. 47-475, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[9] Zitong Yu et al., "Multi-Modal Face Anti-Spoofing Based on Central Difference Networks," *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Seattle, WA, USA, pp. 2766-2774, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[10] Zitong Yu et al., "Flexible-Modal Face Anti-Spoofing: A Benchmark," *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Vancouver, BC, Canada, pp. 6346-6351, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[11] Mingxing Tan, and Quoc Le "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," *Proceedings of the 36th International Conference on Machine Learning*, pp. 6105-6114, 2019. [Google Scholar] [Publisher Link]

[12] Yuanhan Zhang et al., "CelebA-spoof: Large-Scale Face Anti-Spoofing Dataset with Rich Annotations," *Computer Vision–ECCV 2020: 16th European Conference*, Glasgow, UK, pp. 70-85, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[13] Shifeng Zhang et al., "Casia-surf: A Large-Scale Multi-Modal Benchmark for Face Anti-Spoofing," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 2, no. 2, pp. 182-193, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[14] Anjith George et al., "Biometric Face Presentation Attack Detection with Multi-Channel Convolutional Neural Network," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 42-55, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[15] Di Wen, Hu Han, and Anil K. Jain, "Face Spoof Detection with Image Distortion Analysis," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, pp. 746-761, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[16] Zitong Yu et al., "Searching Central Difference Convolutional Networks for Face Anti-Spoofing," *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, pp. 5295-5305, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[17] Anjith George, and Sébastien Marcel, "Cross Modal Focal Loss for RGBD Face Antispoofing," *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, pp. 7882-7891, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[18] Aleksandr Parkin, and Oleg Grinchuk, "Recognizing Multi-Modal Face Spoofing with Face Recognition Networks," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Long Beach, CA, USA, pp. 1617-1623, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[19] Zhuo Wang et al., "Face Anti-Spoofing using Transformers with Relation-Aware Mechanism," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 4, no. 3, pp. 439-450, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[20] Chelsea Finn, Pieter Abbeel, and Sergey Levine, "Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks," *Proceedings of the 34th International Conference on Machine Learning*, Sydney NSW Australia, vol. 70, pp. 1126-1135, 2017. [Google Scholar] [Publisher Link]

[21] Jukka Määttä, Abdenour Hadid, and Matti Pietikäinen, "Face Spoofing Detection from Single Images using Micro-Texture Analysis," *2011 International Joint Conference on Biometrics (IJCB)*, Washington, DC, USA, pp. 1-7, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[22] Oeslle Lucena et al., "Transfer Learning using Convolutional Neural Nettworks for Face Anti-Spoofing," *International Conference on Image Analysis and Recognition*, pp. 27-34, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[23] Zezheng Wang et al., "Deep Spatial Gradient and Temporal Depth Learning for Face Antispoofing," *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, pp. 5042-5050, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[24] Mudunuru Suneel, and Tummala Ranga Babu, "Efficient Face Anti-Spoofing Identifier Network (FASIN) with Depth and Near Infrared Deep Learning Methods," *ARPN Journal of Engineering and Applied Sciences*, vol. 19, no. 20, pp. 1255-1265, 2024. [Google Scholar] [Publisher Link]

[25] Mudunuru Suneel, and Tummala Ranga Babu, "Spoof-Formernet: The Face Anti Spoofing Identifier with a Two Stage High Resolution Vision Transformer (HR-ViT) Network," *International Journal of Image, Graphics and Signal Processing*, vol. 17, no. 4, pp. 87-104, 2025. [CrossRef] [Google Scholar] [Publisher Link]