Academic Journal of Research and Scientific Publishing | Vol 7 | Issue 78
Publication Date: 5 October 2025 ISSN: 2706-6495

Context-Aware Face Presentation Attack Detection (A Dual-Branch Convolutional Neural Network)

Tariq Ahmed Bahatheq

Independent AI Researcher, Saudi Arabia

Email: TariqBahatheq@outlook.com

Abstract

Received:
14 August 2025
First Decision:
25 August 2025
Revised:
20 September 2025
Accepted:
27 September 2025
Published:
5 October 2025

Face recognition systems are susceptible to presentation attacks, which can severely compromise their reliability in security-sensitive applications. Existing methods, such as Deep Pixel-wise Binary Supervision (DeepPixBis), primarily focus on facial regions, often neglecting critical contextual cues in the surrounding image that could signal spoofing attempts. This paper introduces an efficient dual-branch convolutional neural network architecture that integrates facial and contextual information for robust face presentation attack detection, all while maintaining a compact model size. The proposed model processes the extracted face and the entire image independently, producing a pixel-wise map for the face and a binary output for the full image. Trained and evaluated on the OULU-NPU dataset using standard ISO/IEC 30107-3 metrics, the proposed approach achieves state-of-the-art performance among DeepPixBis-based models in protocols II and III. Additionally, it demonstrates state-of-the-art performance in protocol II across all existing models, not just those based on DeepPixBis. Remarkably, it achieves this while being the smallest model among all existing anti-spoofing deep-learning models (1.4M parameters), demonstrating its practicality in real-world scenarios.

Keywords: Anti-spoofing, Facial recognition, Presentation Attack Detection, Deep Learning, Dual-Branch Network, OULU-NPU

1. Introduction

Face recognition is now a ubiquitous biometric technology, widely adopted in applications from mobile authentication to industrial security due to its convenience. However, this success has created a critical vulnerability: presentation attacks (PAs), or spoofing. Attackers can compromise authentication integrity using simple artifacts like printed photos, video replays, or 3D masks. In high-security settings, such failures can lead to significant breaches, making robust Presentation Attack Detection (PAD) an essential safeguard.

Early PAD research relied on hand-crafted features to detect specific spoof artifacts. This included texture analysis using Local Binary Patterns (LBP) (Määttä et al., 2011), motion analysis (Anjos & Marcel, 2011), and liveness cues like eye blinking (Pan et al., 2007). However, these traditional methods often failed to generalize to new attack types and environments. The adoption of deep learning, especially Convolutional Neural Networks (CNNs), marked a paradigm shift by enabling models to automatically learn robust, discriminative features from raw pixel data, significantly improving performance.

Despite the success of deep learning, a significant limitation persists in many state-of-the-art models: an over-reliance on the facial region while neglecting the rich contextual information present in the entire scene. These models often operate solely on tightly cropped face images, discarding valuable cues that could signal a presentation attack. For example, the edges of a handheld screen, reflections from a printed photo, or unnatural lighting inconsistencies in the background can be strong indicators of a spoof, as illustrated in Figure 1. Ignoring these contextual cues can limit the effectiveness of PAD systems, particularly against sophisticated or novel attack vectors.

To address this limitation, this paper introduces an efficient dual-branch context-aware neural network, named "Dual-PADNet." Building upon the Deep Pixel-Wise Binary Supervision framework, the primary aims of this research are to: (1) propose a novel dual-branch architecture that simultaneously processes facial and contextual information to improve detection accuracy without increasing model size; (2) demonstrate that high performance can be achieved with an efficient training strategy and an exceptionally compact model; and (3) validate the model's effectiveness on the OULU-NPU benchmark dataset, aiming for state-of-the-art performance. The rest of this paper is organized as follows: Section 2 reviews related work, Section 3

addresses the research gap and our contributions, Section 4 details our proposed methodology, Section 5 presents experimental results, Section 6 discusses the findings, and Section 7 concludes the paper with suggestions for future work.

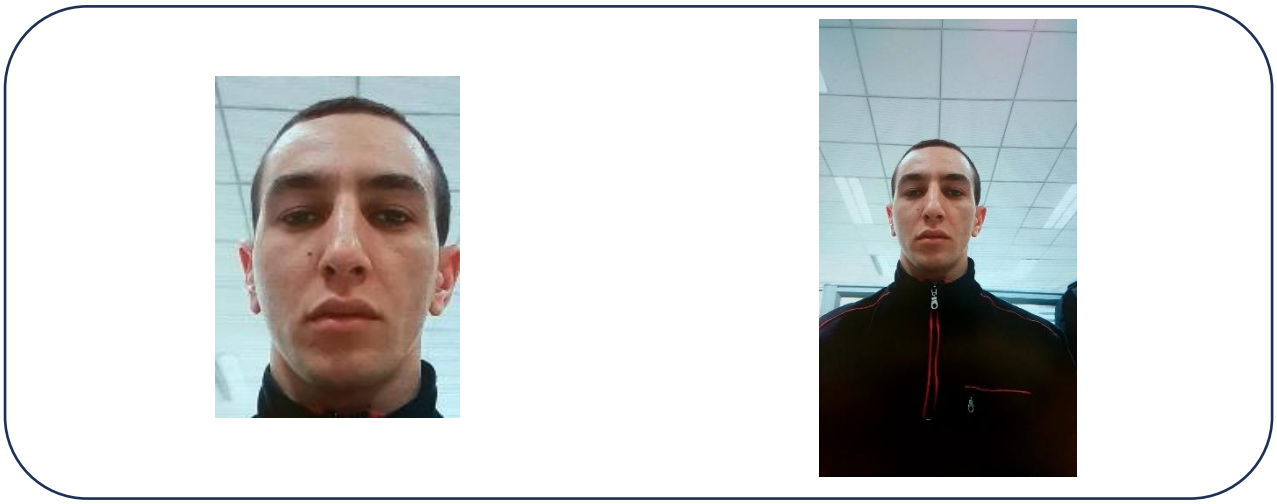

Figure 1: Looking at the face only, it is extremely difficult to determine whether the image is a spoof or not. However, when examining the full image, many artifacts show up such as the colors in the bottom-left, bottom-right, and top-right that might be caused by a reflection.

2. Related Work

The field of Presentation Attack Detection (PAD) has evolved significantly, transitioning from methods based on hand-crafted features to sophisticated deep learning architectures. This section reviews this progression and identifies the key challenges that motivate the present study.

2.1. Traditional Hand-Crafted Feature Approaches

Early efforts in PAD focused on identifying specific, pre-defined artifacts associated with spoofing attacks. These methods can be broadly categorized by the cues they analyze. Texture-based methods were among the most prominent, leveraging descriptors to capture subtle patterns that differentiate live skin from artificial materials like paper or screens. For example, Määttä et al. (2011) employed Local Binary Patterns (LBP) to analyze micro-textures, while Boulkenafet et al. (2016) extended this concept using color texture analysis. Concurrently, motion-based methods exploited temporal information, assuming that the subtle, involuntary movements of a live person differ from the static nature of a photo or the predictable motion of a video replay (Anjos & Marcel, 2011).

A third category, liveness-based methods, sought physiological signs of life, such as eye blinking (Pan et al., 2007). While foundational, these traditional approaches often struggled with generalization, as their hand-crafted features were typically sensitive to variations in lighting, environment, and attack types not seen during development.

2.2. The Shift to Deep Learning Architectures

The limitations of traditional methods paved the way for deep learning, particularly Convolutional Neural Networks (CNNs), which can automatically learn hierarchical and highly discriminative features from data. This paradigm shift led to a significant leap in performance. A foundational work in this area is Deep Pixel-wise Binary Supervision (DeepPixBis) by George and Marcel (2019), which proposed supervising the network at a pixel level. By training the model to generate a binary map distinguishing live and spoof regions, DeepPixBis encouraged the network to learn fine-grained spoofing artifacts. Other deep learning strategies have also been explored. For example, Liu et al. (2018) introduced a CNN-RNN model that incorporated auxiliary supervision using depth maps, while Atoum et al. (2017) proposed a two-stream CNN that fused information from image patches and estimated depth.

2.3. Enhancements to DeepPixBis

Building on the success of initial deep learning models, subsequent research has focused on refining architectures and loss functions. Hossain et al. (2020) proposed A-DeepPixBis, an enhancement to the DeepPixBis framework. They introduced an angular margin-based binary cross-entropy loss (A-BCE) to improve feature discriminability and incorporated an attention mechanism to guide the model toward more informative facial regions. These improvements led to more robust performance, particularly in challenging cross-dataset scenarios. Such works highlight a trend toward not just deeper or wider networks, but smarter supervision and architectural design to extract more meaningful anti-spoofing features.

2.4. Current Challenges

Despite these advancements, several critical challenges remain in the field of face PAD, defining the research gaps that current work aims to address:

- **Neglect of Contextual Information:** Many of the existing methods, including sophisticated deep learning models, operate on tightly cropped face images. This approach inherently

discards the surrounding scene, which may contain crucial evidence of an attack, such as the borders of a tablet, reflections on a printed photograph, or unnatural background elements.

- **Poor Generalization to Unseen Attacks:** Many models exhibit a significant drop in performance when evaluated on new attack types, camera sensors, or environmental conditions that were not part of their training data. Indicating a need for models that learn more fundamental and generalizable features of presentation attacks.
- **Model Size**: While most models used on this field are relatively small, their computational cost can sometimes be prohibitive for real-world deployment on resource-constrained platforms, such as mobile devices or embedded systems. There is a persistent need for lightweight models that do not sacrifice performance.

3. Research Gap and Contributions

The literature review reveals a clear research gap: a lack of PAD models that effectively leverage contextual information while remaining computationally efficient. Existing methods are predominantly face-centric, making them blind to obvious spoofing cues in the surrounding scene. Furthermore, smaller high-performing models are desired considering the deployment of facial recognition on edge devices. This paper directly addresses these limitations by proposing a model designed to be both context-aware and lightweight.

Our contributions are as follows:

- **Novel Dual-Branch Architecture**: We introduce a dual-branch network that simultaneously processes facial and contextual information, enhancing the model's ability to detect presentation attacks while achieving a state-of-the-art compact model size.
- **Efficient Training Strategy**: We only use 5 uniformly sampled frames per training video, 5 uniformly sampled frames per validation video, and train for 30 epochs. Along with one augmentation per image, we believe this reduces overfitting and computational costs without compromising performance significantly.
- **State-of-the-Art Model Size**: Previous research used DenseNet-161 (Huang et al., 2017), Bi-FBN (Roy et al., 2021), and CDCN (Yu et al., 2020). All of these are bigger than the model used here, which is the first 8 layers of DenseNet-121 amounting to only 1.4 million parameters.

- **State-of-the-Art Performance**: Our model achieves competitive results across all protocols of the OULU-NPU (Boulkenafet et al., 2017) dataset, as per ISO/IEC 30107-3 metrics (International Organization for Standardization, 2017). Particularly, it is the best model for protocols II and III among DeepPixBis-based models, and the best model for protocol II among all anti-spoofing models. See section 5.1 for protocol definitions and tables 1 & 2 for comparison with other models.

4. Methodology:

4.1. Overview:

Our proposed method addresses the limitations of existing approaches by incorporating contextual information through a dual-branch architecture. One branch processes the extracted and aligned face, while the other processes the entire image. This design allows the model to capture both facial features and contextual clues indicative of presentation attacks. The architecture is shown in figure 2.

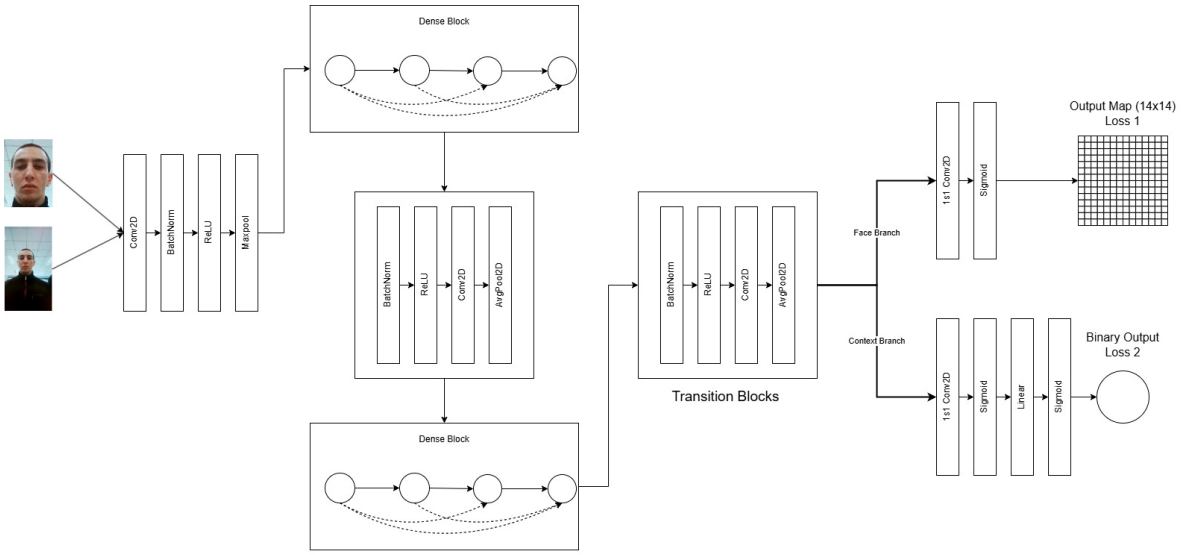

Figure 2: The extracted face and context image (full image) share the same weights for the first 8 layers of Densenet-121. Afterwards, the face branch outputs the pixel map, and the context branch outputs the binary input.

4.2. Data Preprocessing:

4.2.1. Face Extraction and Alignment

We utilize RetinaFace (Deng et al., 2020), a robust face detection and alignment method, to extract and align faces from images. RetinaFace (Deng et al., 2020) detects facial landmarks and

aligns faces to a canonical orientation, ensuring consistency across inputs. The extracted face & aligned face is padded by 25% to ensure no parts of the face are left.

4.2.2. Data Augmentation

To enhance the model's generalization capabilities, we apply various data augmentation techniques:

- Random Horizontal Flips: Flipping images horizontally with a 50% probability.
- Random Affine Transformations: Applying random rotations (±15 degrees) and scaling (90%-110%).
- Color Jitter: Randomly adjusting brightness, contrast, and saturation (±20%).

One augmentation is done per frame.

4.3. Dual-Branch Network Architecture:

Our model consists of two branches: the Face Branch and the Context Branch.

4.3.1 Face Branch:

The Face Branch processes the extracted face image and generates a 14×14 pixel-wise map indicating the likelihood of each pixel being part of a bona fide face.

- Backbone Network: We use a DenseNet-121 architecture (Huang et al., 2017) pre-trained on ImageNet for feature extraction.
- Encoder: The encoder consists of the first eight layers of DenseNet-121 (Huang et al., 2017), capturing hierarchical facial features.
- Decoder: A 256×1 convolutional layer reduces the feature map to a single-channel pixel-wise map.
- Activation: A sigmoid function is applied to produce probabilities between 0 and 1.

4.3.2. Context Branch:

The Context Branch processes the full image, capturing contextual cues that may indicate spoofing, such as background inconsistencies or artifacts.

- Shared Encoder: The Context Branch shares the same encoder architecture as the Face Branch, ensuring consistency and reducing the number of parameters.

- Decoder and Classification: Similar to the Face Branch, it uses a decoder followed by a fully connected layer to produce a binary output indicating spoof or bona fide.

The code for the model architecture can be found in our GitHub repository (Bahatheq, 2024).

4.3.3. Fusion and Decision Making

During inference, the outputs of both branches are combined to make the final decision.

- **Score Averaging**:

$$binary_{preds} = \frac{binary_{output} + scores}{2} \geq 0.5$$

Where $binary_{output}$ is the output from the Context Branch and *scores* is the mean of the pixel-wise map from the Face Branch.

4.4. Loss Function:

We employ a combined loss function that balances pixel-wise supervision and overall binary classification.

4.4.1. Pixel-Wise Binary Cross-Entropy Loss

The pixel-wise loss encourages the model to make accurate predictions at the pixel level for the Face Branch.

$$L_{\text{pixel}} = \frac{-1}{N}\sum_{i=1}^{N}[y_i log\,(p_i) + (1 - y_i) log\,(1 - p_i)]$$

Where N is the number of pixels, y_i is the ground truth label (0 for spoof, 1 for bona fide), and p_i is the predicted probability.

4.4.2. Binary Cross-Entropy Loss

The binary loss penalizes incorrect overall predictions from the Context Branch.

$$L_{\text{binary}} = -[y log\,(p) + (1 - y) log\,(1 - p)]$$

4.4.3. Combined Loss Function

We combine the two losses using a weighting factor λ=0.5:

$$L = \lambda L_{\text{pixel}} + (1 - \lambda) L_{\text{binary}}$$

4.5. Training Strategy:

4.5.1. Frame Selection:

To reduce overfitting and computational costs, we select 5 uniformly sampled frames from each video for training and validation. This approach ensures diversity in training data while maintaining efficiency. This also demonstrates the ability to use less training data and get sufficient results. Moreover, one augmentation for each sampled frame is added. However, when testing, 20 uniformly sampled frames are taken from each video. This is done to ensure that the testing is comparable to that of other research.

4.5.2 Optimization

- Optimizer: Adam
- Learning Rate: 1×10^{-4}
- Weight Decay: 1×10^{-5}
- Batch Size: 32
- Epochs: 30

4.6. Implementation Details:

- Hardware: Training and testing were conducted on NVIDIA GPUs with CUDA acceleration.
- Software: Implemented using PyTorch (Paszke et al., 2019).
- Reproducibility: All model configurations are documented for reproducibility.

5. Experiments and Results:

5.1. Dataset:

We trained and evaluated our model on the OULU-NPU dataset, a widely used and challenging benchmark for face PAD. The database was specifically designed to evaluate the generalization of PAD methods in mobile authentication scenarios and consists of 5,940 videos recorded from 55 subjects using high-resolution frontal cameras of six different smartphones in three different environments (mainly illumination and background scene). The attack types are limited to Printed Photo Attacks (created using two different high-quality printers) and Video Replay

Attacks (replayed on two different high-resolution displays). The dataset is organized into four rigorous protocols designed to test model robustness against specific variations:

- **Protocol 1:** Unseen environmental conditions (lighting and background).
- **Protocol 2:** Unseen Presentation Attack Instruments (PAIs).
- **Protocol 3:** Unseen camera sensors.
- **Protocol 4:** A combination of all three unseen conditions.

The choice of the OULU-NPU dataset over newer alternatives like CelebA-Spoof or CASIA-SURF was deliberate. Primarily, it serves as the standard evaluation benchmark for the foundational DeepPixBis and A-DeepPixBis models, against which our work is compared. Using the same dataset and protocols ensures a direct and fair comparison, accurately measuring the incremental improvements of our proposed architecture. Furthermore, the protocol design of OULU-NPU provides a robust framework for assessing a model's ability to generalize, which is a core focus of our research.

5.2. Evaluation Metrics

We use the ISO/IEC 30107-3 standard metrics (International Organization for Standardization, 2017):

- Attack Presentation Classification Error Rate (APCER): The rate at which attack presentations are incorrectly classified as bona fide.
- Bona Fide Presentation Classification Error Rate (BPCER): The rate at which bona fide presentations are incorrectly classified as attacks.
- Average Classification Error Rate (ACER): Represents the average of the APCER and BPCER, providing a consolidated performance metric.

$$ACER = \frac{APCER + BPCER}{2}$$

For all three metrics, a lower value indicates better performance with 0 being the perfect value.

5.3. Experimental Setup:

5.3.1. Training and Validation:

- Data Splits: Followed the standard splits provided in the OULU-NPU (Boulkenafet et al., 2017) dataset for fair comparison.

- Data Augmentation: Applied as described in Section 4.2.2.
- Best epoch criteria: the model with the lowest ACER is chosen as the final model.

5.3.2. Testing:

- Frame Usage: 20 uniformly sampled frames from each video are used during testing to ensure comparability with existing methods.

5.4 Results:

To assess the performance of our proposed Dual-PADNet model, we compared it against DeepPixBiS-based baselines on the OULU-NPU dataset under the four standard protocols. The evaluation metrics include APCER, BPCER, and ACER, which provide a comprehensive measure of detection accuracy. The results are summarized in Table 1 below.

Table 1. Metrics of our proposed model compared with other DeepPixBis-based models on OULU-NPU (Boulkenafet et al., 2017) for intra-dataset testing.

Protocol	Model	APCER (%)	BPCER (%)	ACER (%)
1	DeepPixBiS (George & Marcel, 2019)	**0.83**	**0.0**	**0.42**
	A-DeepPixBis (Hossain et al., 2020)	1.19	0.31	0.75
	Dual-PADNet (Ours)	2.55	**0.0**	1.27
2	DeepPixBiS (George & Marcel, 2019)	11.39	**0.56**	5.97
	A-DeepPixBis (Hossain et al., 2020)	4.35	1.29	2.82
	Dual-PADNet (Ours)	***0.52***	1.51	**1.01**
3	DeepPixBiS (George & Marcel, 2019)	11.67 ± 19.57	10.56 ±14.06	11.11 ± 9.4
	A-DeepPixBis (Hossain et al., 2020)	2.78 ± 3.47	11.16 ±16.45	6.97 ±7.57
	Dual-PADNet (Ours)	**2.60 ± 3.22**	**7.74 ± 12.33**	***5.17±5.72***
4	DeepPixBiS (George & Marcel, 2019)	36.67 ±29.67	13.33 ±16.75	25.0 ±12.67
	A-DeepPixBis (Hossain et al., 2020)	**3.86 ± 4.04**	**6.56 ±7.88**	**5.22 ±2.96**
	Dual-PADNet (Ours)	4.81 ± 7.42	17.55± 15.69	*11.18 ± 5.85*

As shown in Table 1, our Dual-PADNet architecture demonstrates superior performance among DeepPixBis-based models in Protocol II (ACER of 1.01%) and Protocol III (ACER of 5.17%). The outstanding result in Protocol II, which tests generalization to unseen spoofing devices,

suggests that the contextual branch is highly effective at identifying artifacts from different printers and screens. While the model underperforms in Protocol I, it achieves perfect BPCER of 0.0%, indicating it never misclassifies a genuine user, a critical feature for usability. The performance dip in Protocol IV highlights the extreme challenge of generalizing across all variables simultaneously, a known issue for many PAD models.

5.5. Comparison with State-of-the-Art:

To evaluate the performance of the proposed Dual-PADNet, we conducted a comparative evaluation against several state-of-the-art presentation attack detection (PAD) models on the OULU-NPU dataset. The baselines include CDCN++, Bi-FAS-S, FAS-BAS, and DeepPixBiS, tested under the four standard protocols. Performance was assessed using the established error rates APCER, BPCER, and ACER, ensuring consistency with prior PAD studies. The results of this comparison are summarized in Table 2 below.

Table 2. Metrics of our proposed model compared with best models on OULU-NPU (Boulkenafet et al., 2017) for intra-dataset testing.

Protocol	Model	APCER (%)	BPCER (%)	ACER (%)
1	CDCN++ (Yu et al., 2020)	**0.4**	**0.0**	**0.2**
	Bi-FAS-S (Roy et al., 2021)	3.13	0.83	1.97
	FAS-BAS (Liu et al., 2018)	1.6	1.6	1.6
	DeepPixBiS (George & Marcel, 2019)	0.83	**0.0**	0.42
	A-DeepPixBis (Hossain et al., 2020)	1.19	0.31	0.75
	Dual-PADNet (Ours)	2.55	**0.0**	1.27
2	CDCN++ (Yu et al., 2020)	1.8	0.8	1.3
	Bi-FAS-S (Roy et al., 2021)	1.67	1.11	1.39
	FAS-BAS (Liu et al., 2018)	2.7	2.7	2.7
	DeepPixBiS (George & Marcel, 2019)	11.39	**0.56**	5.97
	A-DeepPixBis (Hossain et al., 2020)	4.35	1.29	2.82
	Dual-PADNet (Ours)	***0.52***	1.51	**1.01**
3	CDCN++ (Yu et al., 2020)	1.7 ± 1.5	2.0 ± 1.2	1.8 ± 0.7
	Bi-FAS-S (Roy et al., 2021)	**0.69 ± 0.68**	**0.28 ± 0.68**	**0.49 ± 0.63**
	FAS-BAS (Liu et al., 2018)	2.7 ± 1.3	3.1 ± 1.7	2.9 ± 1.5

	DeepPixBiS (George & Marcel, 2019)	11.67 ± 19.57	10.56 ±14.06	11.11 ± 9.4
	A-DeepPixBis (Hossain et al., 2020)	2.78 ± 3.47	11.16 ±16.45	6.97 ±7.57
	Dual-PADNet (Ours)	2.60 ± 3.22	7.74 ± 12.33	*5.17±5.72*
4	CDCN++ (Yu et al., 2020)	4.2 ± 3.4	5.8 ± 4.9	5.0 ± 2.9
	Bi-FAS-S (Roy et al., 2021)	**2.50 ± 3.16**	**3.33 ± 4.08**	**2.92 ± 3.41**
	FAS-BAS (Liu et al., 2018)	9.3 ± 5.6	10.4 ± 6.0	9.5 ± 6.0
	DeepPixBiS (George & Marcel, 2019)	36.67 ±29.67	13.33 ±16.75	25.0 ±12.
	A-DeepPixBis (Hossain et al., 2020)	3.86 ± 4.04	6.56 ±7.88	5.22 ±2.96
	Dual-PADNet (Ours)	4.81 ± 7.42	17.55± 15.69	*11.18 ± 5.85*

Table 2 compares Dual-PADNet with a wider range of state-of-the-art models. The key finding is that our model achieves the best overall performance in Protocol II with an ACER of 1.01%, outperforming even larger and more complex models like CDCN++ and Bi-FAS-S. This is a significant result, confirming the value of contextual information for generalizing across different attack instruments. Furthermore, our model achieves a state-of-the-art BPCER of 0.0% in Protocol I and a state-of-the-art APCER of 0.52% in Protocol II. While other models show stronger performance in Protocols III and IV, our model remains competitive in these protocols, especially considering it is by far the smallest and most efficient, as will be discussed in Section 6.

6. Discussion

6.1. Impact of Contextual Information:

By incorporating the full image context, our model captures additional cues that are often overlooked in face-only approaches. Background inconsistencies, edges of spoofing devices, and lighting discrepancies can provide valuable information for PAD.

6.2. Computational Efficiency:

Our dual-branch architecture, while more complex than single-branch models, remains efficient due to shared weights, a small backbone, and less frame usage during training. This makes the model suitable for deployment in real-time applications and on devices with limited computational resources such as edge devices. Table 3 illustrates different anti-spoofing models with their sizes.

To quantify the model's computational efficiency, performance was benchmarked on an NVIDIA RTX 3070 Ti GPU. Using a batch size of 1, the model achieved an average inference latency of 12.039 ms, enabling a throughput of 83.07 frames per second (FPS), while consuming a peak VRAM of only 6.46 MB.

Table 3. Model parameters comparison

Model	Parameters
CDCN (Yu et al., 2020)	2.25 M
CDCN++ (Yu et al., 2020)	> 2.25 M
Bi-FAS-S (Roy et al., 2021)	> 4 M
FAS-BAS (Liu et al., 2018)	> 10 M
DeepPixBiS (George & Marcel, 2019)	> 3 M
A-DeepPixBis (Hossain et al., 2020)	> 3 M
Dual-PADNet (Ours)	**1.4 M**

6.3. Limitations

It is important to acknowledge the scope and limitations defined by this dataset. The OULU-NPU database exclusively contains print and replay attacks and does not include other critical attack types prevalent in modern face anti-spoofing research, such as 3D Mask Attacks, Silicone Masks, cosmetic Makeup Attacks, or AI-generated Deepfake Attacks. Consequently, the model was trained and evaluated only on the specific artifacts associated with print and replay spoofs.

While our dual-branch approach is designed to capture a broader range of anomalies—and the context branch could theoretically detect cues like the visible edges of a 3D mask—its performance against these other attack types is unverified. We have not tested the model on other types of attacks, and its effectiveness against them cannot be guaranteed. This represents a clear limitation of the current study. Therefore, the results presented in this paper are specific to the print and replay attacks found in the OULU-NPU dataset. Future work should focus on evaluating and adapting the Dual-PADNet architecture on more diverse datasets that incorporate these modern attack vectors to fully validate its generalizability.

7. Conclusion and Future Work

We have introduced an efficient dual-branch convolutional neural network architecture for face presentation attack detection. Utilizing a dual-branch architecture that processes both facial and contextual information, our approach effectively addresses the limitations of existing methods. On the OULU-NPU (Boulkenafet et al., 2017) dataset, our model achieves state-of-the-art performance in protocols II and III among DeepPixBis-based models. Notably, it also outperforms all existing models to date in protocol II. Moreover, it delivers state-of-the-art results in the BPCER metric for protocol I and the APCER metric for protocol II. Remarkably, our model stands as the smallest among all anti-spoofing models, offering superior efficiency without compromising performance.

By offering an efficient and compact solution for face presentation attack detection, our model is particularly well-suited for deployment in industrial settings. This suitability is further reinforced by its state-of-the-art performance in Protocol II and competitive performance in Protocol I, which are the most relevant protocols for industrial access control systems, as they typically involve consistent camera types.

In future work, we plan to:

- Explore Different Weighting Factors: Investigate the impact of varying lambda in the loss function.
- Extend to Other Datasets: Evaluate the model's performance on other PAD datasets to assess generalizability.
- Use A-DeepPixBis (Hossain et al., 2020) Loss Function: The Angular binary cross-entropy Loss used in their paper achieved better results than vanilla binary cross-entropy loss. Using their loss function for the face pixel map might achieve better results.
- Add a Branch for Fourier Spectra: As seen by the Bi-FPN for Face Anti-Spoofing (Roy et al., 2021) paper, Fourier Spectra improves performance
- Better Augmentation: We believe that augmentations are the key to ensuring generalizability. In fact, our model outperformed on the second protocol where lighting and camera are fixed. This signals that more & different augmentations might help it outperform on protocols III and IV

8. References:

George, A., & Marcel, S. (2019, June). Deep pixel-wise binary supervision for face presentation attack detection. In *2019 international conference on biometrics (ICB)* (pp. 1-8). IEEE. https://doi.org/10.48550/arXiv.1907.04047

Boulkenafet, Z., Komulainen, J., Li, L., Feng, X., & Hadid, A. (2017, May). 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)* (pp. 612-618). IEEE. http://dx.doi.org/10.1109/FG.2017.77

International Organization for Standardization. (2017). *Information technology — Biometric presentation attack detection — Part 3: Testing and reporting* (Standard No. ISO/IEC 30107-3:2017). https://www.iso.org/standard/67381.html

Boulkenafet, Z., Komulainen, J., & Hadid, A. (2016). Face spoofing detection using colour texture analysis. *IEEE Transactions on Information Forensics and Security*, *11*(8), 1818-1830. https://doi.org/10.1109/TIFS.2016.2555286

Määttä, J., Hadid, A., & Pietikäinen, M. (2011, October). Face spoofing detection from single images using micro-texture analysis. In *2011 international joint conference on Biometrics (IJCB)* (pp. 1-7). IEEE. https://doi.org/10.1109/IJCB.2011.6117510

Anjos, A., & Marcel, S. (2011, October). Counter-measures to photo attacks in face recognition: a public database and a baseline. In *2011 international joint conference on Biometrics (IJCB)* (pp. 1-7). IEEE. https://doi.org/10.1109/IJCB.2011.6117503

Liu, Y., Jourabloo, A., & Liu, X. (2018). Learning deep models for face anti-spoofing: Binary or auxiliary supervision. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 389-398). https://doi.org/10.1109/CVPR.2018.00048

Atoum, Y., Liu, Y., Jourabloo, A., & Liu, X. (2017, October). Face anti-spoofing using patch and depth-based CNNs. In *2017 IEEE international joint conference on biometrics (IJCB)* (pp. 319-328). IEEE. https://doi.org/10.1109/BTAS.2017.8272713

Hossain, M. S., Rupty, L., Roy, K., Hasan, M., Sengupta, S., & Mohammed, N. (2020, November). A-DeepPixBis: Attentional angular margin for face anti-spoofing. In *2020 Digital Image Computing: Techniques and Applications (DICTA)* (pp. 1-8). IEEE. https://doi.org/10.1109/DICTA51227.2020.9363382

Deng, J., Guo, J., Ververas, E., Kotsia, I., & Zafeiriou, S. (2020). Retinaface: Single-shot multi-level face localisation in the wild. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 5203-5212). http://dx.doi.org/10.48550/arXiv.1905.00641

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700-4708). https://doi.org/10.1109/CVPR.2017.243

Roy, K., Hasan, M., Rupty, L., Hossain, M. S., Sengupta, S., Taus, S. N., & Mohammed, N. (2021). Bi-FPNFAS: Bi-Directional Feature Pyramid Network for Pixel-Wise Face Anti-Spoofing by Leveraging Fourier Spectra. *Sensors*, *21*(8), 2799. https://doi.org/10.3390/s21082799

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... & Chintala, S. (2019). Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, *32*. https://doi.org/10.48550/arXiv.1912.01703

Yu, Z., Zhao, C., Wang, Z., Qin, Y., Su, Z., Li, X., ... & Zhao, G. (2020). Searching central difference convolutional networks for face anti-spoofing. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 5295-5305). https://doi.org/10.1109/CVPR42600.2020.00534

Pan, G., Sun, L., Wu, Z., & Lao, S. (2007, October). Eyeblink-based anti-spoofing in face recognition from a generic webcamera. In *2007 IEEE 11th international conference on computer vision* (pp. 1-8). IEEE. https://doi.org/10.1109/ICCV.2007.4409068

Bahatheq, T. (2024). *Dual-PADNet* [Computer software]. GitHub. https://github.com/Tariq-droid/Dual-PADNet

Doi: https://doi.org/10.52132/Ajrsp.e.2025.78.1