Detecting Presentation Attacks on ID cards Using Feature Refinement

Raghavendra Mudgalgundurao
NTNU Gjøvik
raghavem@stud.ntnu.no

Patrick Schuch
NTNU Gjøvik
patrick.schuch2@ntnu.no

Raghavendra Ramachandra
NTNU Gjøvik
raghavendra.ramachandra@ntnu.no

Kiran Raja
NTNU Gjøvik
kiran.raja@ntnu.no

***Abstract*—Identity documents (IDs) verify a person's identity in various applications such as banking, travel, and border control. Systems used for verifying ID cards can be attacked using different attack mediums (e.g., printed ID, screen displays) and thus need to be equipped with presentation attack detection (PAD) systems. Developing better PAD schemes is often limited by the availability of datasets. To address the shortage of relevant datasets for training PAD systems, we introduce DASAC (Document Authentication and Synthetic Attack Collection)[1]. The dataset includes synthetic ID card data generated using Unconditional Latent Diffusion (ULD), referred to as DASAC-ULD generated. Additionally, it features a subset created through carefully designed template transfer techniques, known as DASAC-Crafted Template. We further present a dual-stage architecture based PAD scheme developed using EfficientNet-Transformer network to detect diverse presentation attacks. Further, leveraging mixup data augmentation to enhance model robustness, the proposed approach achieves an Equal Error Rate (EER) of 3.14% with a Bona Fide Presentation Classification Error Rate (BPCER) of 2.42% and 1.33% at Attack Presentation Classification Error Rates (APCER) of 5% and 10%, respectively. Additional ablation studies, has been uploaded to the GitHub link [2].**

***Index Terms*—ID Cards, Presentation Attack Detection, Biometrics.**

I. Introduction

Biometric access control systems are widely deployed in critical infrastructure (e.g., border security, financial institutions) and rely on International Civil Aviation Organization (ICAO)-compliant documents like passports and ID cards. The COVID-19 pandemic accelerated demand for remote authentication, driving advancements in digital document verification [1, 2]. However, ID card verification systems remain vulnerable to presentation attacks (PAs) conducted with physical counterfeits (e.g., variable-quality printed forgeries) and digital attacks (e.g., high-resolution screen replays) [3, 4].

ICAO-compliant NFC-enabled documents offer intrinsic Presentation Attack Detection (PAD) capabilities, however, their wide-spread adoption is constrained by infrastructure costs [5]. Thus, non-NFC ID card verification systems need integrated PAD, and PAD methods have been proposed leveraging texture analysis, pattern recognition, and machine learning—are essential for detecting material anomalies in forged credentials [4, 6, 7]. Techniques are further devised to detect both print-based and digital attacks, emphasizing robustness against evolving spoofing technologies [8]. In the following section, we review recent research on PAD methods through a taxonomy that classifies attacks into physical counterfeits (e.g., printed forgeries) and digital forgeries (e.g., screen replays, synthetic manipulations), contextualizing their detection methodologies.

Recent advances in ID card PAD techniques consider diverse methodologies and datasets. Gonzalez et al. proposed two key approaches: (1) a face quality assessment framework using MagFace [9] on Chilean ID cards (74,939 images), where low-quality face regions indicate attacks [4], and (2) a dual-stage MobileNetV2 architecture distinguishing bona fide documents from composite/synthetic forgeries (Stage 1) and fine-grained attack types (Stage 2), achieving a $BPCER_{100}$ of 92% on a 190,000-image dataset [8]. Kiselev et al. [10] introduced a compact dataset of 500 video clips featuring 50 different types of identity documents, marking one of the earliest publicly available resources for identity document analysis and recognition in video streams. Their work established baselines for face detection accuracy, Optical Character Recognition precision across four primary document fields, and data extraction from video streams.

Mudgalgundurao et al. [7] proposed a pixel-wise supervision for PAD model where German ID cards were used to detect the attacks. Although achieving an EER of 2.2%, the approach was aimed at localizing printer artifacts and moiré patterns from screens to detect attacks. However, the approach was tested on an in-house dataset, making it difficult for other researchers to develop new PAD models. In a similar direction, Kunina et al. [11] focused only on screen attack detection in documents, noting that moiré patterns predominantly appear on document edges. Their approach, utilizing Hough transforms for pattern detection, achieved 95.4% precision on the DLC-2021 dataset of simulated documents and screen captures.

Despite those advancements, existing research is constrained by limited data diversity, often focusing on single-type documents and synthetic IDs. Thus, we introduce a synthetic ID card dataset generated via an Unconditional Latent Diffusion (ULD) model, which eliminates external conditioning inputs, reducing computational complexity and noise [12] while capturing essential security features. Making use of this dataset, we propose a two-stage PAD architecture where EfficientNet-B4 [13] in the first stage extracts features, capturing fine-grained variations across different ID card instances. A transformer-based second stage [14] leverages self-attention mechanisms to model long-range dependencies and contextual relationships within ID card embeddings. To enhance robustness, we incorporate mixup augmentation [15] for diverse training data, addressing common attack types such as replay, print, crafted template transfer, and synthetic ID card attacks. Specifically, our contributions to this work can be listed below:

- Synthetic ID Card Dataset: We generate 15,000 synthetic attack images via ULD and an additional 1,340 Crafted Template Transfer images by modifying existing ID cards and passports while retaining security features like holograms. Our dataset covers ULD-generated IDs, Crafted Template Attacks (CTA), Texture Transfer Attacks (TTA), and Screen & Print Attacks.
- Novel Two-Stage PAD Architecture: Unlike prior works limited to a single ID type, our approach supports diverse IDs (e.g., driving licenses, passports, and entry cards), leveraging mixup augmentation for enhanced attack detection across different scenarios.

[1]Dataset is available upon request for research purposes only. Access to dataset is subject to agreement on terms of use that ensure the data is used exclusively for non-commercial research activities.

[2]https://github.com/Raghavendra-MG/EUSIPCO_Supplementary_Paper

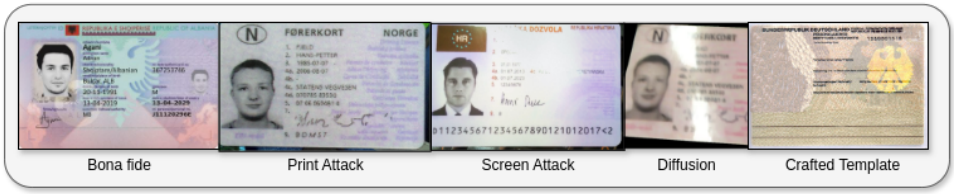

Fig. 1: Example images from our DASAC dataset include bona fide, print, replay attack, diffusion, and crafted texture transfer types.

Database	Bona fide	Screen	Print	Synthetic	Size (images)	Attack Types	Synthetic Data	Augmentation	Availability
Chilean ID Cards[3]	6,588	24,778	6,972	-	38,338	Screen, Print	No	Limited	Restricted
MIDV-500*[10]	15,000	-	-	-	15,000 + 500 videos	None	No	None	Public
KID34K[16]	13,746	13,729	7,187	-	34,662	Screen, Print	No	Limited	Public
DASAC-ULD Generated (Ours)	-	-	-	15,000	15,000	Synthetic	Yes (15K ULD)	Extensive	Research use (Non-commercial)
DASAC-Crafted Template (Ours)	-	-	-	1,340	1,340	Synthetic	Yes (1.3K Template)	Extensive	Research use (Non-commercial)
DASAC-Texture Transfer (Ours)	-	2,221	2,226	-	4,447	Screen, Print	No	Extensive	Research use (Non-commercial)
DASAC-Screen Attack (Ours)	-	65,023	-	-	65,023	Screen	No	Extensive	Research use (Non-commercial)
DASAC-Print Attack (Ours)	-	-	65,023	-	65,023	Print	No	Extensive	Research use (Non-commercial)

TABLE I: Summary of major ID document datasets including class distributions, attack types, and key dataset characteristics. *MIDV also includes 500 videos of ID cards.

Database	Bona fide			Replay			Screen			Synthetic		
	Train	Validation	Test	Train	Validation	Test	Train	Validation	Test	Train	Validation	Test
MIDV-500[10]	55667	7783	16323	26392	6298	13365	26392	6298	13365	-	-	-
KID34K[16]	11342	862	1539	4208	382	692	2543	864	1463	-	-	-
DASAC- ULD Generated (Ours)	-	-	-	-	-	-	-	-	-	10500	1500	3000
DASAC- Crafted Template (Ours)	-	-	-	-	-	-	-	-	-	938	134	268

TABLE II: Breakdown of the dataset samples to training, validation, and testing sets.

- Comprehensive Evaluation: We benchmark our models against State-of-the-Art (SOTA) approaches on our new dataset and conduct targeted ablation studies to analyze strengths and limitations, offering insights for advancing PAD research.

In the rest of the paper, we first present the newly created ID card attack database in Section II along with a detailed rationale for our proposed approach in Section III. We present different baseline evaluations along with proposed approach in Section IV, and demonstrate the applicability of the proposed approach in detecting the attacks effectively. We provide a discussion on the conducted ablation studies in Section V. We further present the limitations of our work in Section VI, and Section VII presents the conclusion and future work.

II. Document Authentication and Synthetic Attack Collection (DASAC) Database

Due to the sensitive nature of ID card data, open-source datasets for research remain scarce. Among the limited resources, MIDV-500 [10] offers 500 video clips from 50 unique identity documents, including 17 ID cards, 14 passports, 13 driving licenses, and 6 miscellaneous documents, with high-resolution (4032×3024) image settings were captured, featuring various individuals under different lighting and background conditions. KID34K [16] addresses forgery detection with 34,662 images of 82 replica ID cards (46 falsified identities), subdivided into 13,746 bona fide samples, 13,729 screen-displayed attacks, and 7,187 printed copies, acquired under realistic conditions using 12 smartphones and diverse digital devices (tablets, monitors). Similarly, the Chilean ID Cards dataset [3] provides 6,588 bona fide images captured at 1280×720 resolution using Samsung S6/S8 and iPhone 11 Pro devices, alongside 24,778 screen-based attacks (displayed on Acer Aspire Nitro 5 and HP 22w monitors) and 6,972 print-based attacks generated via an Epson L4160 printer on regular and glossy paper. Table I comprehensively summarizes existing and contributed datasets.

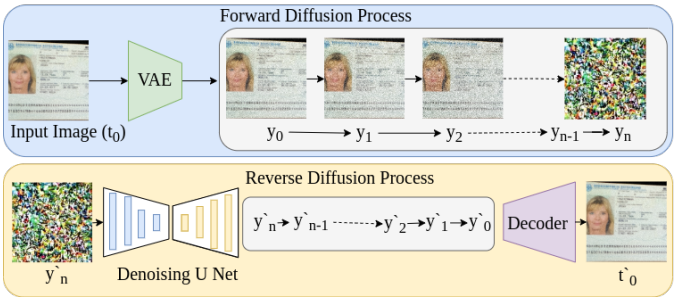

Fig. 2: A graphical representation of our trained ULD network, which inputs a cropped ID and generates synthesized ID cards.

A. Our dataset - DASAC

To address the scarcity of public datasets, we developed a new ID card dataset using ULD, where we make use of MIDV-500 [10] and KID34K [16] datasets. Our pipeline utilizes the open-source ID card detection by Steidle [17] to extract ID cards from MIDV-500 video frames. These high-resolution ID card frames were then manually curated to ensure data quality by discarding images that included background noise. The curated dataset is used to create different attacks that include crafted template attacks (Sec II-A2), texture transfer attacks (Sec II-A3), and screen/print attacks (Sec II-A4). In the below section, we outline each step involved in generating synthetic ID cards. Example images from our DASAC dataset are shown in Figure 1.

1) Generating Images from Unconditional Latent Diffusion Model: The ULD performs a diffusion process in a compressed latent space to synthesize high-fidelity ID card images. The model architecture

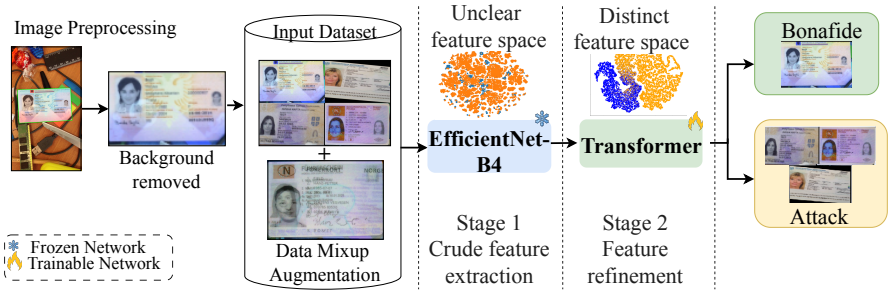

Fig. 3: The input is an ID card image that passes through an EfficientNet-B4 model to extract embeddings. These embeddings are then fed into a transformer network. The transformer uses multiple layers of attention mechanisms and feed-forward networks to process the embeddings efficiently. We visualize the embeddings from the EfficientNet-B4 network and their transformed counterparts after training the transformer network using t-distributed stochastic neighbor embedding (t-SNE) representation.

integrates a 2D UNet [18] to reverse the diffusion process from Denoising Diffusion Probabilistic Models (DDPM) [19], as illustrated in Figure 2. Our implementation utilizes pre-trained weights from [20] and incorporates a frozen Variational Autoencoder (VAE) [21] to encode high-dimensional images into a lower-dimensional latent space, optimizing computational efficiency.

For training, we utilized the curated dataset described earlier, ensuring high-quality ID card images for generating synthetic samples using the Unconditional Latent Diffusion Model. The training pipeline employs a DDPMScheduler [22] for noise management during forward diffusion and inference, while an AdamW optimizer minimizes the loss between predicted and actual noise in the latent representations. We use an Exponential Moving Average (EMA) to stabilize training via temporal weight averaging, promoting consistent gradients and convergence.

2) Crafted Template Attacks: To enhance the dataset diversity with realistic examples, we create Crafted Template Attacks (CTA) making use of 15 different templates available in MIDV-500 dataset (passport and ID cards). Each of the different type of ID card is used to create a blank template by removing all the details, such as face image and demographic details using GIMP[3] in a careful manner. We generate a new subset of attacks, with random demographic details such as name, date of birth, ID number, and validity dates and added face images from CelebA dataset yielding 1,340 attack ID cards.

3) Texture Transfer Attacks: We apply screen and print attack textures from texture transfer method of Benalcázar et al [6] to bona fide samples sourced from MIDV-500 and KID34K datasets. Through this methodology, we generated a comprehensive dataset of 4,454 ID card images that simulate print and replay attacks, thereby expanding the robustness of our attack detection approach.

4) Screen and Print Attacks: Screen attacks are captured by displaying ID cards from MIDV-500 and KID34K datasets on a Samsung LU32J509 LED monitor, recorded using Samsung Galaxy S20 (3840×2160) and Note 20 Ultra (1920×1080) under varied lighting/quality settings. Print attacks involve grayscale/color ID cards printed with a Canon TS-5000, captured using the same smartphones at 1920×1080 resolution. This ensures diverse attack types (screen/print) and capture conditions for robust PAD evaluation.

III. Proposed Approach for PAD

Our proposed PAD framework, as illustrated in Figure 3 enhances attack detection through three key components: mixup data augmentation [15] to improve data diversity of the training data, EfficientNet-B4 [13] for extracting fine-grained features from ID cards, and a transformer architecture [14] that exploits the contextual relationships between the features through self-attention mechanisms to detect the attacks.

A. Mixup Augmentation: Class-Specific Implementation

Mixup augmentation [15] generates synthetic training examples by linearly interpolating data points and their corresponding labels. We create class-specific augmentation to generate massive training data. We make use of both label-based mixup and data-based mixup to make the training of PAD more robust. By providing such diverse augmented data for bona fide and attack classes, we aim to enhance the model's discriminative capability and generalization to different kinds of ID cards. Given input images x and corresponding labels $y \in 0,1$ (where 0 represents bona fide and 1 represents attack samples), the mixing process is controlled by α which is parameterized in a β distribution. The interpolation strength between samples is controlled by $\lambda \in [0,1]$ drawn from the β distribution, where a symmetric distribution ensures balanced interpolation between classes.

Sample Generation: For each bona fide sample $x_b[i]$ with label $y_b[i]$, a randomly selected attack sample $x_a[i]$ with label $y_a[i]$ is mixed using the following equation:

$$f_{samples}[i] = \lambda \cdot x_b[i] + (1-\lambda) \cdot x_a[i] \quad (1)$$

The corresponding label mixing is as follows:

$$f_{labels}[i] = \lambda \cdot y_b[i] + (1-\lambda) \cdot y_a[i] \quad (2)$$

During training, these mixed samples are incorporated using a modified loss function:

$$L_{mixup} = \lambda L(\mathcal{F}(f_{samples}), f_{labels}) \quad (3)$$

where L represents the base loss function and $\mathcal{F}(\cdot)$ denotes the model's predictions.

Our class-specific mixup implementation synthesizes samples along the bona fide and attacks using $\lambda \sim \beta(\alpha,\alpha)$. When $\alpha = 1$, λ follows a uniform distribution, ensuring unbiased interpolation between classes. For $\alpha < 1$, the β distribution becomes U-shaped, favoring extreme λ values (e.g., $\lambda \approx 0.05$ or 0.95), which generates samples strongly weighted toward one class while preserving subtle features of the other. As α decreases further, this polarization effect intensifies, enhancing the model's ability to detect subtle adversarial artifacts while maintaining robustness through controlled feature interpolation. By setting the $\lambda \sim \beta(\alpha,\alpha)$ equals the number of bona fide samples in a batch, the symmetric β distribution ensures balanced interpolation between classes in our work.

[3]https://www.gimp.org

Approach	EER%	BPCER @ APCER=5%	BPCER @ APCER=10%
Gonzalez[4]	7.61	10.41	5.95
PixelWise Model[7]	21.02	56.13	36.35
Gonzalez and Tapia[8]	23.46	70.42	47.59
With Mixup Strategy			
Efficientnet-B4 ($\alpha=0.2$)	7.09	12.97	3.51
MobileNet V2 ($\alpha=0.2$)	4.89	4.81	3.37
MobileNetV3-Large ($\alpha=0.2$)	7.74	9.84	6.64
ViT ($\alpha=0.2$)	21.68	56.11	38.98
Proposed ($\alpha=0.2$)	**3.14**	**2.42**	**1.33**
Cross-Database Evaluation on IDNet using our Proposed model [23]			
Screen and Print	14.75	17.82	16.15
Copy-Move	53.60	97.83	94.55
Crop-Replace	58.79	99.35	98.44
Face Morphing	50.53	95.02	90.57
Inpaint-Rewrite	58.59	99.31	98.09
Face Replacement	50.58	94.11	88.79

TABLE III: Comparison of the proposed approach against baseline and state-of-the-art (SOTA) models. The SOTA models were evaluated without the mixup strategy and are presented as benchmarks.

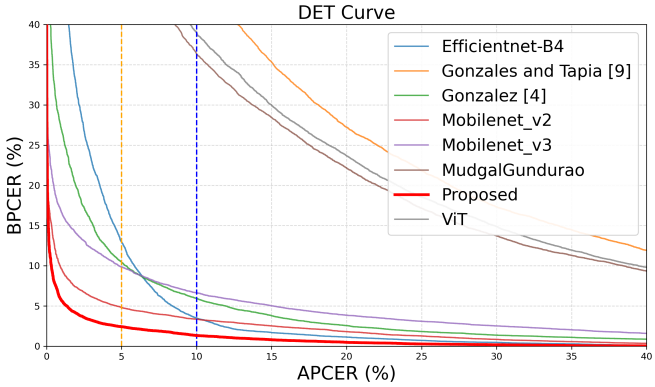

Fig. 4: DET plots for baseline and proposed methods.

B. Feature Refining for Enhanced PAD

Our PAD framework leverages a pre-trained EfficientNet-B4 [13] as a parameter-efficient backbone to extract generic features, which are then task-adapted via a transformer encoder. The transformer refines features through self-attention mechanisms, modeling discriminative bona fide vs attack feature relationships. By combining EfficientNet's discriminative local features with the transformer's global contextual modeling [14], our approach robustly captures fine-grained attack patterns critical for PAD, as shown in Figure 3.

C. Implementation Details

Input images are resized to 380×380 and processed by a pre-trained EfficientNet-B4 to extract 1792-D features, projected to 512-D via a linear layer. A three-layer transformer encoder employs: (i) 8-head multi-head self-attention, (ii) 2048-D feed forward network, and (iii) dropout (0.2). Classification uses a softmax-activated fully connected layer. Training employs Adam optimizer ($lr = 10^{-4}$) with beta parameters annealed from (0.9,0.999) to (0.1,0.1) every 30 epochs. The ReduceLROnPlateau scheduler (factor = 0.1, patience = 10) stabilizes the training.

IV. EXPERIMENTS

In this section, we present the results of selected baselines, evaluation metrics, and experimental results used to assess PAD performance. We compare our approach against three different PAD schemes for ID (Gonzalez [4], PixelWise Model [7], and Gonzalez and Tapia's [8]) and different deep learning models, such as MobileNet V2, MobileNetV3-Large, EfficientNetB4, and Vision Transformer. We evaluate all methods using standardized metrics from the ISO/IEC 30107-3:2017 [24] guidelines for PAD, including APCER, BPCER, and EER. We make the dataset disjoint for training, validation, and testing setup, with data split into 70% for training, 10% for validation, and 20% for testing as indicated in Table II.

A. Results and Discussion

The results presented in Table III compare the performance of various approaches using EER and BPCER at APCER = 5%, and APCER = 10%, and in Figure 4, the Detection Error Tradeoff (DET) curve is presented. To systematize the comparison, we focus on BPCER@APCER=5%, which reflects the ability of each experiment to correctly accept bona fide samples while controlling the false acceptance rate at a reasonable threshold.

Among the baseline methods, Gonzalez [4] method achieves the lowest BPCER of 10.41%, whereas the PixelWise Model [7] and Gonzalez and Tapia's [8] show higher BPCERs of 56.13% and 70.42% respectively, indicating poor robustness to presentation attacks. The proposed PAD model, trained with mixup α = 0.2, outperforms all others, achieving a BPCER of 2.42%, showcasing its effectiveness in distinguishing between bona fide and attack samples. MobileNet V2 also uses mixup at the same α, achieves a BPCER of 4.81%, outperforming several baseline models, but performs lower than the proposed PAD model.

The proposed PAD model, trained with mixup α = 0.2 outperforms all others, achieving a BPCER of 2.42%, demonstrating its effectiveness in distinguishing between attack and bona fide samples. MobileNet V2, also using mixup at the same α, follows with a BPCER of 4.81%, outperforming many baselines but not matching the proposed method. EfficientNet-B4 and MobileNetV3-Large also improve over SOTA methods but show elevated BPCERs compared to our approach. Interestingly, Vision Transformer (ViT) performs poorly in this setting with a BPCER of 56.11%. Furthermore, we observe that while the proposed model remains effective at higher α values (e.g., 0.5 and 1.0), $\alpha=0.2$ offers the best tradeoff between robustness and generalization. [4]

In parallel work, the PAD-IDCard 2024 competition [25] established a standardized benchmark for ID card presentation attack detection using a sequestered dataset containing genuine and attack samples from four countries. The top-performing team achieved a 74.30% average rank with BPCER values frequently exceeding 40% at APCER=10%, while the MobileVIT baseline trained on private data achieved 1.84% BPCER. Despite the different test methodologies that preclude direct comparison, these results show that DASAC is a good approach to training robust PAD models.

[4]For a complete comparison of all the alpha values, please refer to Table **STIII** provided in the supplementary material.

Approach	EER%	BPCER @ APCER=5%	BPCER @ APCER=10%
Proposed ($\alpha = 0.2$)	**3.14**	**2.42**	**1.33**
Reduced Transformer Depth	24.58	45.76	37.85
Increased Attention Heads	20.19	85.74	62.42
Reduced Feed-Forward Dimensionality	22.91	81.72	62.22

TABLE IV: Results from ablation studies to demonstrate the impact of varying model parameter configurations.

V. Ablation Studies

To analyze the impact of various components on the performance of the proposed PAD model, we conduct an extensive ablation study. We study the impact of transformer depth, role of attention mechanism, and the feed-forward capacity as illustrated in Table IV. We further study the impact of the rate of mixup by varying the α coefficient as provided in the supplementary material (Table STIII) due to page constraints.

Transformer Depth: Reducing the number of transformer layers degrades the model's performance, as shown in Table IV, suggesting the role of depth in capturing complex patterns in ID card features.

Attention Mechanism: Increasing the number of attention heads from 8 to 16 in an attempt to potentially capture more intricate relations in the image. However, this results in no improvement in PAD.

Feed-Forward Network Capacity: We also investigate the effect of reducing the dimensionality of the feed-forward network within each transformer layer from 2048 to 1024. The reduction in feed-forward capacity results in degraded PAD performance, indicating the role of larger dimensions.

Attack-Specific Analysis We evaluate the model's effectiveness across different attack types by comparing each attack type against bona fide samples. Results detailed in supplementary material in Tables ST4–ST6 show that our model performs particularly well in detecting ID cards generated using diffusion while also maintaining high detection rates for print and screen attacks.

Class Activation Map Analysis We further study the Class Activation Maps (CAM), the visualization provided in the supplementary material (Figure SF7 - SF24) indicates which regions drive the classification as either bona fide or attacks. These visualizations aid in interpreting the model's focus and validating its decision-making process.

VI. Limitations of our work

While our proposed method demonstrates promising results, it has limitations. First, privacy constraints limit bona fide ID data, causing class imbalance, favouring attack samples. Second, synthetic data generation (ULD) improves diversity but cannot fully replicate intricate security features, which may reduce the detection of advanced forgeries. Third, despite data augmentation, performance may decline under real-world conditions such as harsh lighting and motion blur. Addressing data scarcity, improving synthetic fidelity, and environmental robustness remains critical for future work.

VII. Conclusion and future directions

Verifying ID cards in unsupervised settings remains challenging due to prevalent presentation attacks involving digital or photocopied IDs. To tackle this, we introduced a synthetic ID card dataset generated via ULD, and complemented it with crafted template transfer attacks, print and replay attacks. This dataset will be made available for non-commercial research to help advance the presentation attack detection techniques.

Further, we have presented a new PAD framework that utilizes the strengths of two different architectures in a cascaded manner to detect various attacks on ID cards. Using mixup strategy, the model effectively learns subtle distinctions between bona fide and attack samples. Our proposed approach demonstrates better PAD performance with a BPCER of 2.42%, and 1.33% at APCER of 5% and 10%, respectively. Future work can be extended to simulate attacks under varies captured conditions - such as reflections, rotations, and background noise. Additionally, alternative methods for combining complementary architectures could further improve robustness.

References

[1] Y. Shi and A. K. Jain, "Docface: Matching id document photos to selfies," in *IEEE International Conference on Biometrics Theory, Applications and Systems BTAS*.

[2] Y. Shi and A. K. Jain, "DocFace+: ID document to selfie matching," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2019.

[3] S. Gonzalez, A. Valenzuela, and J. Tapia, "Hybrid two-stage architecture for tampering detection of chipless id cards," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2020.

[4] S. González and J. Tapia, "Towards refining id cards presentation attack detection systems using face quality index," in *2022 30th European Signal Processing Conference (EUSIPCO)*.

[5] W. H. Lee, C. M. Chou, and S. W. Wang, "An nfc anti-counterfeiting framework for id verification and image protection," *Mobile Networks and Applications*.

[6] D. Benalcazar, J. E. Tapia, S. Gonzalez, and C. Busch, "Synthetic id card image generation for improving presentation attack detection," *IEEE Transactions on Information Forensics and Security*.

[7] R. Mudgalgundurao, P. Schuch, K. Raja, R. Ramachandra, and N. Damer, "Pixel-wise supervision for presentation attack detection on identity document cards," *IET Biometrics*, 2022.

[8] S. Gonzalez and J. E. Tapia, "Forged presentation attack detection for id cards on remote verification systems," *Pattern Recognition*, 2025.

[9] Q. Meng, S. Zhao, Z. Huang, and F. Zhou, "Magface: A universal representation for face recognition and quality assessment," in *IEEE/CVF conference on computer vision and pattern recognition*, pp. 14225–14234, 2021.

[10] A. Kiselev, M. Burmistrov, V. Goncharov, I. Grigoriev, and A. Zelensky, "MIDV-500: A dataset for identity documents analysis and recognition on mobile devices in video stream," in *International Conference on Document Analysis and Recognition (ICDAR)*, pp. 225–230, 2020.

[11] I. Kunina, A. Sher, and D. Nikolaev, "Screen recapture detection based on color-texture analysis of document boundary regions," , 2023.

[12] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.

[13] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," *arXiv preprint arXiv:1905.11946*, 2019.

[14] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Kaiser, and I. Polosukhin, "Attention is all you need," in *31st International Conference on Neural Information Processing Systems (NeurIPS)*.

[15] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," *arXiv preprint arXiv:1710.09412*, 2017.

[16] E.-J. Park, S.-Y. Back, J. Kim, and S. S. Woo, "Kid34k: A dataset for online identity card fraud detection," in *32nd ACM International Conference on Information and Knowledge Management*, 2023.

[17] T. Steidle, "Ml_idcard_segmentation_pytorch," 2021. Accessed: 2024-07-14.

[18] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI 2015*, Springer.

[19] J. Ho, P. Jaini, A. Srinivas, P. Abbeel, and Y. Duan, "Denoising diffusion probabilistic models," *arXiv preprint arXiv:2006.11239*, 2020.

[20] Zyinghua, "Unconditional image generation with linearly decoded masks," 2024.

[21] D. P. Kingma, "Auto-encoding variational bayes," *arXiv:1312.6114*, 2013.

[22] H. F. Team, "Diffusers: State-of-the-art diffusion models for image and audio generation in python." https://huggingface.co/docs/diffusers, 2022.

[23] H. Guan, Y. Wang, L. Xie, S. Nag, R. Goel, N. E. N. Swamy, Y. Yang, C. Xiao, J. Prisby, R. Maciejewski, *et al.*, "Idnet: A novel dataset for identity document analysis and fraud detection," *arXiv preprint arXiv:2408.01690*, 2024.

[24] ISO/IEC JTC1 SC37 Biometrics, *ISO/IEC 30107-3. Information Technology - Biometric presentation attack detection - Part 3: Testing and Reporting*. International Organization for Standardization, 2017.

[25] J. E. Tapia, N. Damer, C. Busch, J. M. Espin, J. Barrachina, A. S. Rocamora, K. Ocvirk, L. Alessio, B. Batagelj, S. Patwardhan, *et al.*, "First competition on presentation attack detection on id card," in *2024 IEEE International Joint Conference on Biometrics (IJCB)*.