A Comprehensive Framework for Palm Vein Anti-Spoofing With Preprocessing Pipeline, Dataset, and Benchmark

Caiping Yan, *Member, IEEE*, Zhi Lan, Hong Li, Yuqi Li, and Zonglin Meng

***Abstract*—Non-contact palm-vein recognition has been widely adopted in security-critical applications owing to its contact-free acquisition paradigm and exceptional discriminative power. However, these systems remain susceptible to a spectrum of presentation attacks (PAs), creating significant security risks that require urgent mitigation. Progress on palm-vein anti-spoofing is currently impeded by three fundamental gaps: 1) the absence of an open-source, end-to-end pipeline for preprocessing liveness-detection data; 2) a lack of publicly available datasets specifically tailored to anti-spoofing evaluation; and 3) the unavailability of benchmark studies employing standardized protocols and reference implementations. In light of these key issues, we make the following four contributions. Firstly, we propose a new open-source preprocessing pipeline that can significantly improve model performance, reducing errors by up to 53.3%. Secondly, we introduce PVASD, a new largest known dataset comprised of 1,187,519 images belonging to 5,515 subjects, which consists of 880,241 live palm vein images along with 307,278 spoof attack images including 16 different attack types–both 2d (printed stuff) and 3d (gloves and prosthesis models) attacks captured under a variety of environments using off-the-shelf commercial-grade sensors at five resolutions. Lastly, comprehensive benchmarks are also created by evaluating three types of representative methods namely classical image classification models, face anti-spoofing methods adapted from face domain and anomaly-detection-based approaches, while our experimental results reveal unique characteristics intrinsic only to palm vein spoof attacks, which will hopefully provide valuable guidance to researchers for further investigation. Furthermore, we also expanded PVASD by adding 20,000 spoof samples generated by artificial intelligence, and evaluated the vulnerability of the existing models to deepfake attacks. We will release our preprocessing pipeline, dataset, and benchmark codes at https://github.com/valhongli/PVASD to advance future reproducible studies and accelerate palm-vein anti-spoofing algorithmic research.**

***Index Terms*—Palm vein, anti-spoofing, presentation attacks, preprocessing pipeline, benchmark.**

Received 10 June 2025; revised 5 December 2025; accepted 24 December 2025. Date of publication 2 January 2026; date of current version 14 January 2026. The associate editor coordinating the review of this article and approving it for publication was Prof. Raghavendra Ramachandra. *(Corresponding authors: Hong Li; Zonglin Meng.)*

Caiping Yan is with the School of Information Science and Technology, Hangzhou Normal University, Hangzhou 311121, China (e-mail: yancp@hznu.edu.cn).

Zhi Lan, Hong Li, and Yuqi Li are with Hangzhou Insvision Technology Company Ltd., Hangzhou 311121, China (e-mail: hznujxbs@gmail.com; hong.li.leon@connect.um.edu.mo; yuqi001125@gmail.com).

Zonglin Meng is with the University of California, Los Angeles, CA 90095 USA (e-mail: meng925@g.ucla.edu).

Digital Object Identifier 10.1109/TIFS.2025.3650391

I. Introduction

PALM vein recognition technologies [1], [2], [3], [4] are now pervasive in various domains, including access control systems, military and confidential facilities, and financial payment verification. The adoption of palm vein recognition systems continues to grow due to their non-invasive nature and high security characteristics, making them particularly suitable for high-security applications.

However, these systems face significant security risks that cannot be ignored. Palm vein recognition systems [5], [6], [7] are vulnerable to various presentation attacks [8], [9], [10], [11], including print attacks (using printed vein patterns), video replay attacks (displaying recorded vein patterns) and fake hand presentations (using artificial hand models with simulated vein patterns). The consequences of successful attacks can be severe. In military and confidential facilities, security breaches could lead to compromise of classified information and potentially endanger military personnel and operations. In financial systems, malicious attackers may exploit vulnerabilities to commit identity theft and fraud. Even in less critical applications like attendance systems, successful spoofing attacks could disrupt personnel management and compromise organizational security. These vulnerabilities underscore the critical importance of developing robust PVAS (palm vein anti-spoofing) technologies.

Despite widespread adoption and security concerns, palm vein anti-spoofing remains largely unexplored compared with face anti-spoofing in both academia and industry. Although face anti-spoofing [12], [13], [14], [15], [16] have seen significant advances with numerous public datasets [17], [18], [19], [20], [21] and established benchmarks [22], [23], [24], palm vein anti-spoofing lacks fundamental research resources such as open-source data preprocessing pipeline, large-scale public datasets, and standardized benchmarking protocols for reliable performance evaluation.

The main steps of palm vein preprocessing pipeline are the localization of the palm and the extraction of ROI. Recently, most palm vein preprocessing [25], [26], [27], [28], [29], [30], [31] mainly focus on ROI extraction algorithms, which are design for palm vein recognition and have no open-source code. Therefore, contactless palm vein ROI location is still a challenge [10]. Besides, several public palm vein datasets are available, as shown in Table I, such as PUT Vein pattern Database [32], CASIA-MS-PalmprintV1

TABLE I
PALM VEIN DATASETS

Dataset	Subjects	Quantity	Pixels	Sensor	Format	PAs
PUT Vein Pattern Database [32]	100	2,400	1280x960	880nm	Image	None
CASIA-MS-PalmprintV1 [33]	200	7,200	768x576	460nm,630nm,700nm,850nm, 940nm and White Light	Image	None
PolyU Multispectral Palmprint Database [34]	500	6,000	128x128	NIR	Image	None
VERA Palmvein Database [35]	220	2,200	480x640	940nm	Image	None
FYO Palmvein Database [36]	150	640	800x600	NIR	Image	None
Our Dataset (PVASD)	5,515	1,187,519	640x400,640x480, 800x608,1024x720, 1140x720	NIR	Image	2D Prints, 3D

[33], PolyU Multispectral Palmprint Database [34], VERA Palmvein Database [35] and FYO Palmvein Database [36]. However, these datasets are primarily designed for identity recognition research and only contain live samples without any spoof attack samples. There are no publicly available datasets specifically designed for palm vein anti-spoofing research, and very limited literature addressing this critical security challenge. Additionally, the field lacks standardized benchmarks for assessing the effectiveness of palm vein liveness detection methods, which creates challenges for comparative analysis and advancement in this critical security domain. Therefore, this significant gap between practical needs and research progress calls for immediate attention from the research community.

To bridge this gap and establish a foundation for palm vein anti-spoofing research, an open-source data preprocessing pipeline, a comprehensive dataset with diverse presentation attacks and standardized evaluation protocols are urgently needed. Such a data preprocessing pipeline will enable researchers to enhance privacy protection and algorithm fairness when handling any palm vein anti-spoofing dataset for training or testing, and remove the background area while retaining only the areas rich in palm vein. Moreover, such a dataset would enable researchers to develop and evaluate anti-spoofing methods, understand the unique challenges in palm vein attacks, and ultimately enhance the security of palm vein recognition systems in actual applications. Furthermore, the establishment of benchmark protocols and baseline methods would provide a solid foundation for future research and facilitate fair comparisons between different approaches.

To address these challenges, we design a novel preprocessing pipeline that enhances both privacy protection and model performance for palm vein anti-spoofing, as shown in Fig. 4. By leveraging MediaPipe Hands solution [37] for landmark detection, then we extract only the vein-rich regions of palm images, removing all background elements that could contain sensitive information. The pipeline identifies six key palm landmarks to calculate the palm center and direction, establishing a standardized 128×128 Region-of-Interest (ROI). This preprocessing approach not only safeguards user privacy but significantly improves model effectiveness, demonstrated by a 53.3% reduction in error rates in our ablation studies. The standardized input created through this process provides a consistent foundation for reliable palm vein anti-spoofing detection.

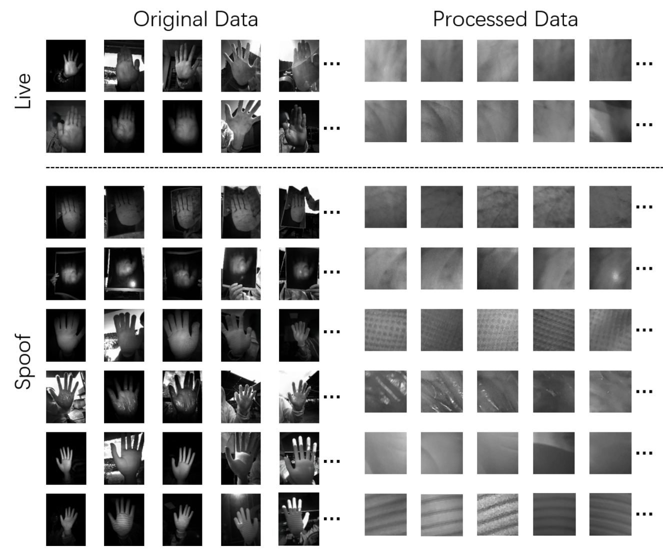

Fig. 1. The samples of PVASD. It is a large-scale palm vein anti-spoofing dataset containing 880,241 live images from 5352 real palms and 307,278 spoof images from 163 spoof subjects. Current publicly available palm vein recognition datasets suffer from insufficient spoof attack samples and the absence of standardized preprocessing pipeline.

In addition, we introduce PVASD (Palm Vein Anti-Spoofing Dataset), shown in Fig. 1, the first large-scale palm vein anti-spoofing dataset containing 1,187,519 images from 5,515 subjects. PVASD significantly advances the field through several key innovations in data collection and diversity. For live samples, we collect 880,241 real palm vein images from 5,352 different subjects under various natural conditions, ensuring the dataset reflects actual usage scenarios, as illustrated by the live samples in Fig. 2. For spoof samples, we systematically design and implement 16 different types of presentation attacks, including both 2D attacks (such as grayscale photograph, infrared photograph, color photograph and A4 paper hand-shaped cutouts) and 3D attacks (such as white cotton gloves and so on), resulting in 307,278 sophisticated spoof samples, as illustrated by the spoof samples in Fig. 3. The capture environment is carefully controlled to include different illumination conditions, collection angles, and distances to simulate various real-life scenarios. All samples were captured using a variety of commercial sensors to ensure image quality and consistency while maintaining practical applicability. We collected data at five different resolutions

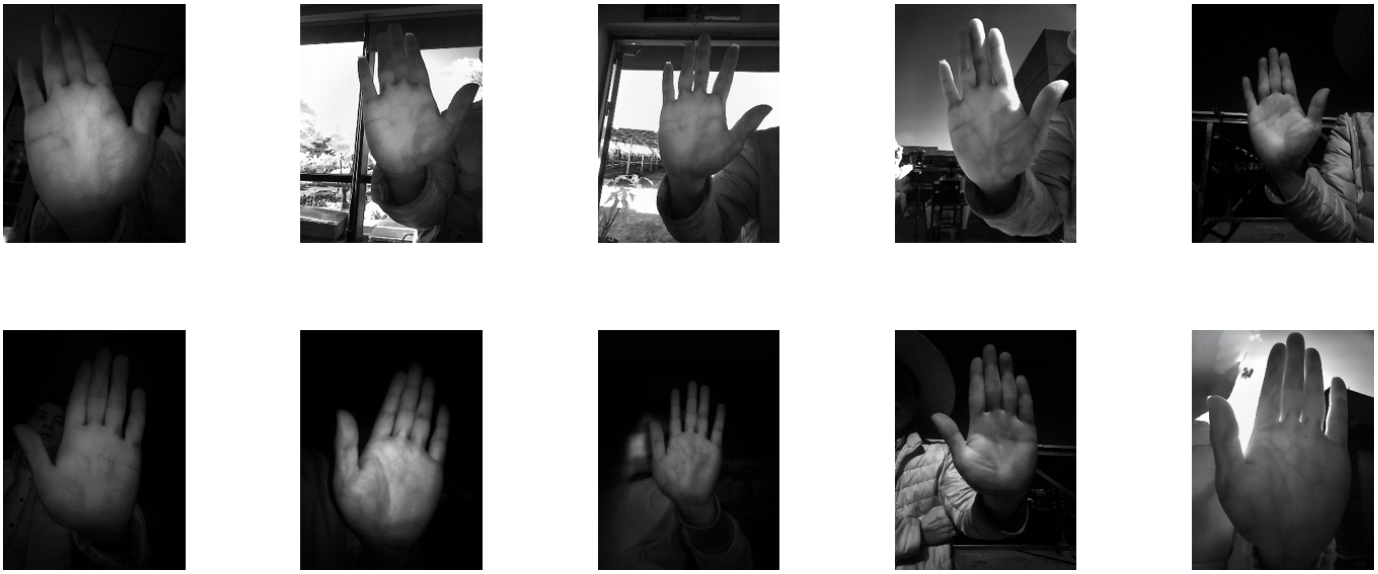

Fig. 2. Examples of live palm vein in PVASD. The dataset is categorized into two main groups: Indoor and Outdoor. The Indoor category includes scenes with no natural light and window-side conditions with varying lighting orientations (Front-lit, Back-lit, Side-lit). The Outdoor category comprises scenes with different lighting orientations (Front-lit, Back-lit, Side-lit).

(e.g., 640×400, 640×480, 800×608, 1024×720, 1140×720), aiming to simulate various device capture conditions and enhance the diversity of the dataset.

Moreover, to establish comprehensive benchmarks for PVASD, we evaluate three categories of representative approaches: classical image classification models (including state-of-the-art architectures like ConvnextLarge [38], DenseNet201 [39], EfficientNetB7 [40], EfficientVetX2L [41], MobileNetV3Large [42], RegnetX32gf [43], ResNet152 [44], and Vgg19 [45]), face anti-spoofing techniques (like BiFPNFAS [12] and EntryV1 [13]) adapted for palm vein, and anomaly detection algorithms (like DevNet [46] and SimpleNet [47]). These diverse baselines serve multiple purposes: they provide fundamental references for future research, highlight the unique challenges in palm vein anti-spoofing, and demonstrate the dataset's utility for developing specialized solutions. We implement standardized evaluation protocols using established metrics including Attack Presentation Classification Error Rate (APCER), Bonafide Presentation Classification Error Rate (BPCER), and Average Classification Error Rate (ACER) to ensure fair and comprehensive performance assessment [12], [13], [17], [24].

The five contributions of our paper are as follows:

(1) **Advanced data preprocessing pipeline.** To prevent palm vein spoofing, we propose a new preprocessing pipeline that employs standardized ROI extraction and quality evaluation after palm area keypoint localization using the MediaPipe Hands solution. Through the preprocessing pipeline we proposed, the error rate can be reduced by up to 53.3%, significantly improving the model performance and providing a solid foundation for accurate and efficient anti-spoofing methods.

(2) **Pioneering the creation of datasets.** With 1,187,519 photos from 5,515 subjects, we present PVASD, the first rich palm vein anti-spoofing dataset. Our PVASD contains real palm vein data in various scenarios and multiple spoof categories, and has collected a total of five different resolution data, ensuring the diversity and robustness of the technology.

(3) **Establishment of the first benchmark.** By assessing three types of representative methods—classical image classification models, face anti-spoofing methods, and anomaly detection algorithms—we create thorough benchmarks for palm vein anti-spoofing. Our work provides open-source code of standardized evaluation protocols and metrics to promote further research for palm vein anti-spoofing.

(4) **The foundation for upcoming studies.** Through comprehensive testing and analysis, we identify critical technical issues pertaining to palm vein anti-spoofing, assess the efficacy of various methodological approaches, pinpoint promising avenues for the advancement of specialized PVAS solutions, and offer valuable resources for further research in this domain.

(5) **First evaluation of deepfake vulnerabilities.** We extend PVASD with 20,000 AI-generated spoof samples (named PVASD-deepfake) and conduct controlled experiments on model robustness to generative attacks, providing quantitative evidence of current systems' fragility and establishing a new benchmark for deepfake defense research.

The preprocessing pipeline, PVASD, PVASD-deepfake, and benchmark code will be made publically at https://github.com/valhongli/PVASD available to promote research in this crucial but understudied field of biometric security.

The rest of the paper is organized as follows: Section II presents the preprocessing pipeline, Section III introduces the PVASD dataset, Section IV provides comprehensive benchmarks and evaluations (including deepfake vulnerability analysis), and Section V concludes with future work.

II. Dataset Preprocessing Pipeline

To enhance privacy protection and algorithmic fairness, we preprocessed the dataset by removing all non-palm background regions, retaining only the vein-rich area. This step

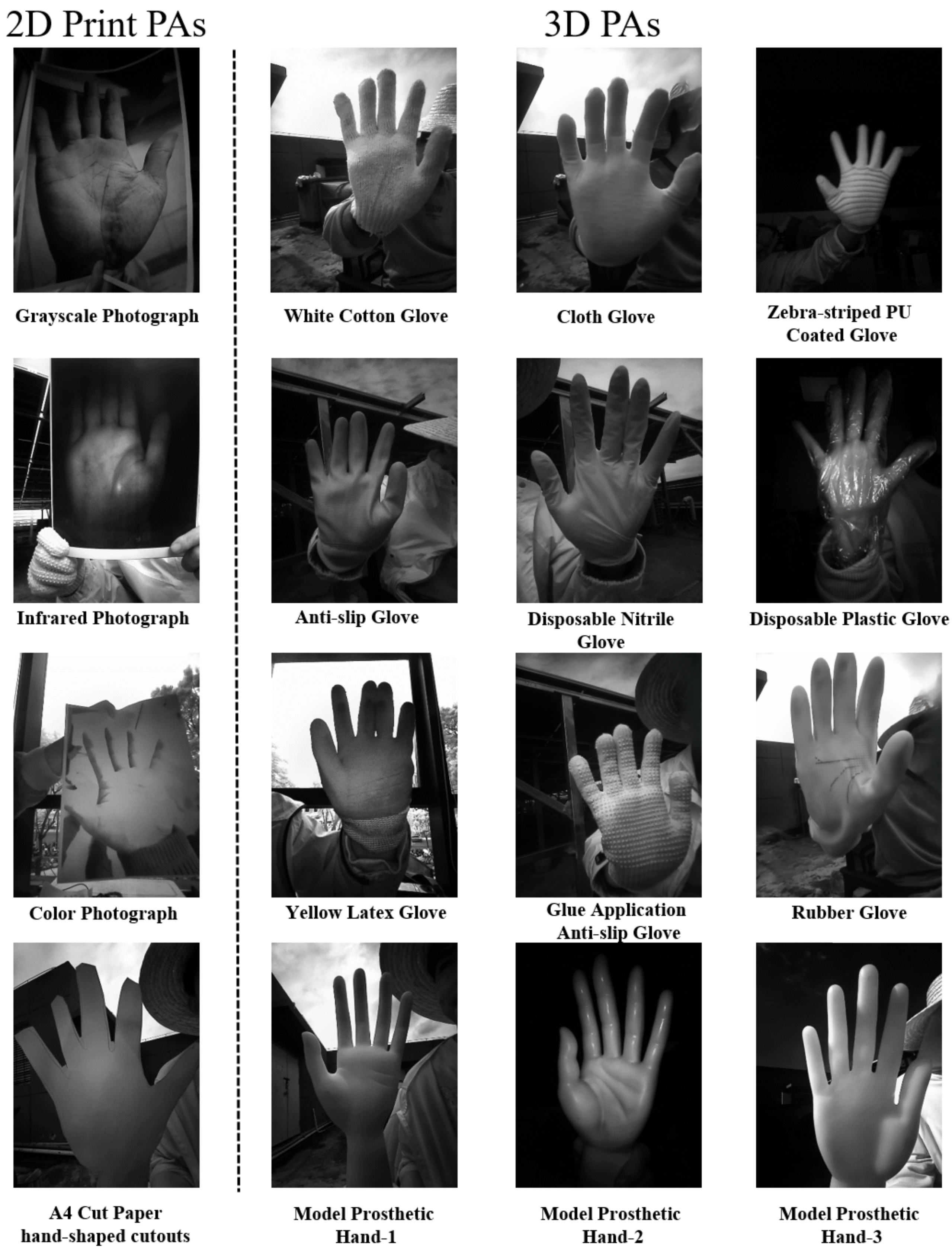

Fig. 3. Examples of spoof palm vein in PVASD. The dataset is categorized into two groups: 2D printed PAs and 3D PAs. 2D printing PAs includes four types, while 3D PAs includes twelve types. These PAs are all collected under various conditions such as different light intensities in multiple scenarios (indoor and outdoor).

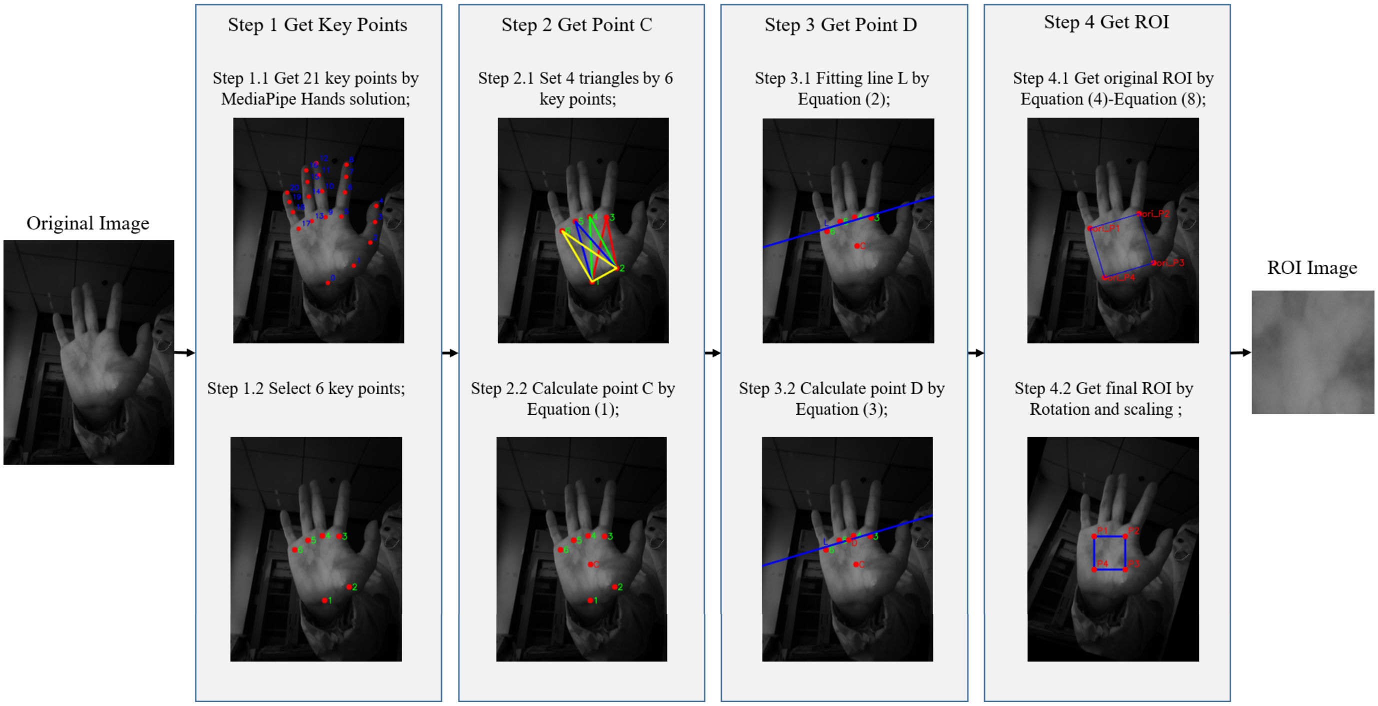

Fig. 4. The Preprocessing pipeline and details of the PVASD.

prevents potential exposure of sensitive information (e.g., skin texture or surrounding objects) while standardizing the input to consistent Region-of-Interest (ROI) sub-images, as illustrated in Fig. 4.

We use the MediaPipe Hands solution [37] to detect palm landmarks in the original image. MediaPipe is a framework developed by Google for building machine learning pipelines, particularly suited for tasks like pose estimation, face detection, and hand tracking. Moreover, MediaPipe Hands is open-source under Apache 2.0 with fixed model weights (v0.10.8 used), ensuring full reproducibility via our released code. Its geometric landmark detection minimizes texture bias, further mitigated by PVASD's diversity across 5,515 subjects and 5 sensors. The MediaPipe Hands solution includes a palm detection model that operates on the entire image to return an oriented hand bounding box, and a hand landmark model that operates on the cropped hand image region to return 21 3D hand landmarks MediaPipe Hands Documentation.

From the 21 landmarks provided by MediaPipe, we extract the following six key points: wrist (point 1), thumb CMC (point 2), index finger MCP (point 3), middle finger MCP (point 4), ring finger MCP (point 5), and pinky MCP (point 6). These points are used to determine the palm center and direction, which are then used to extract the ROI for palm vein analysis.

After obtaining the six key points, we calculate the center point (point C) of the palm. Through these six key points, we divide the palm area into four triangles. The point 1 and point 2 are two points of the triangles, and points 3 to 6 are the other points of the four triangles. We calculate the center points and areas of these four triangles, and then perform a weighted average calculation on them to obtain the center point (point C) of the entire palm, as shown in Equation (1):

$$C_x = \frac{\sum(center_{i,x} \cdot area_i)}{\sum area_i},$$
$$C_y = \frac{\sum(center_{i,y} \cdot area_i)}{\sum area_i}, \tag{1}$$

where C_x and C_y are the horizontal and vertical coordinates of point C, $center_{i,x}$ and $center_{i,y}$ are the horizontal and vertical coordinates of the center of triangle i, $area_i$ is the area of triangle i. This step ensures that the center point reflects the overall position of the hand, and triangles with larger weights (larger areas) contribute more to the center. This central point (point C) serves as a reference for locating the ROI.

Next, we determine the direction of the palm. Using the key points at the base of the fingers (points 3, 4, 5, 6), we fit a straight line (line L) through the least squares method, calculate its slope and intercept, as shown in Equation (2):

$$a_1 = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sum x_i^2 - \frac{(\sum x_i)}{n}}, a_0 = \frac{\sum y_i - a_1 \sum x_i}{n}, \tag{2}$$

where a_1 is the slope of the line L, a_0 is the intercept of the line L, x_i and y_i are the horizontal and vertical coordinates of point i, and n is number of the points. This straight line L represents the main direction of the palm.

Then, we determine the specific original location of the ROI. We find the point (point D) closest to the center point (point C) of the palm on the fitted straight line, as shown in Equation (3):

$$D_x = \frac{C_x + a_1 \cdot C_y - a_1 \cdot a_0}{a_1^2 + 1}, D_y = a_1 \cdot D_x + a_0, \tag{3}$$

where D_x and D_y are the horizontal and vertical coordinates of point D. Subsequently, we use Equation (4) to calculate the offsets d_x and d_y from point C to point D, as shown in Equation (4):

$$d_x = D_x - C_x, d_y = D_y - C_y. \tag{4}$$

where dx and dy are the offsets of the point C to point D. Then, we use Equation (5) to (8) to get the four original corner points (ori_P1 to ori_P4) of the original ROI, as shown in Equation (5) to (8):

$$ori\_P1_x = D_x + d_y, ori\_P1_y = D_y - d_x, \tag{5}$$
$$ori\_P2_x = D_x - d_y, ori\_P2_y = D_y + d_x, \tag{6}$$
$$ori\_P3_x = D_x + d_y - 2 \cdot d_x, ori\_P3_y = D_y - d_x - 2 \cdot d_y, \tag{7}$$
$$ori\_P4_x = D_x - d_y - 2 \cdot d_x, ori\_P4_y = D_y + d_x - 2 \cdot d_y, \tag{8}$$

where ori_Pi_x and ori_Pi_y are the horizontal and vertical coordinates of point ori_Pi. In order to ensure that the ROI area is as rich in palm veins as possible and the ROI area always faces upwards in the image, we rotated and scaled the four corner points (ori_P1 to ori_P4) of the original ROI to obtain the four corner points (P1 to P4) and ROI.

Finally, we extract the ROI and evaluate its quality. By cropping and resizing the original image, we obtain a 128×128 ROI image. After extraction, we check the grayscale quality of the image, including whether the brightness and contrast meet the preset standards. If qualified, we save the ROI image; otherwise, we mark it as failed and end the process.

Through the above steps, from key point detection to quality assessment, the palm vein ROI extraction process can efficiently obtain high-quality and standardized regions from palm vein images, providing a reliable foundation for subsequent palm vein anti-spoofing tasks.

Our preprocessing pipeline represents a significant contribution to palm vein anti-spoofing research, as it not only ensures data consistency and privacy protection but also directly enhances model performance by isolating discriminative vascular features, as demonstrated by a 53.3% reduction in error rates in our ablation studies.

III. Palm Vein Anti-Spoofing Dataset

Currently, several public palm vein datasets are available, such as the PUT Vein pattern Database [32], CASIA-MS-PalmprintV1 [33], PolyU Multispectral Palmprint Database [34], VERA Palmvein Database [35] and FYO Palmvein Database [36]. However, these datasets are primarily designed for identity recognition research and only contain live samples without any spoof attack samples. There are no publicly available datasets specifically designed for palm vein anti-spoofing research, and very limited literature addressing this critical security challenge. This significant gap between practical needs and research progress calls for immediate attention from the research community.

To bridge this gap and establish a foundation for palm vein anti-spoofing research, we present PVASD, the first large-scale dataset containing live data and spoof data, 1,187,519 images from 5,515 subjects. The PVASD significantly advances the field through several key innovations in data collection and diversity.

A. Dataset Overview

1) Live Data: The live samples was self-collected through a controlled acquisition process involving 2,676 distinct human participants. Each participant contributed high-resolution images of both their left and right palm veins, captured in various real-world scenarios and varying illumination conditions, as illustrated by the live samples in Fig. 2, resulting in a total of 880,241 palm vein images. A range of sensor types were used to capture data at five different resolutions in order to maintain image quality while guaranteeing diversity. Realism and utility are prioritized in the dataset, which features natural backdrops, realistic attitudes, and illumination variations that are present in daily life. Every participant provided their informed permission in compliance with ethical guidelines, and the information was anonymized to protect privacy.

2) Spoof Data: We built the spoof data of the PVASD using a self-collection mechanism to address the drawbacks of existing datasets, which basically lack fake data. By gathering data in a range of circumstances, we ensure comprehensive coverage of 2D and 3D presentation attacks (PAs) utilizing a variety of materials and real-life situations. The spoof samples' practicability and authenticity are greatly increased by this technique.

As illustrated by the spoof samples in Fig. 3, the PAs in our dataset encompass both 2D and 3D forms with diverse materials. The 2D print PAs has four styles (i.e., bending, folding, cutting, and plne) and four carriers (i.e., grayscale photograph, infrared photograph, color photograph and A4 paper hand-shaped cutouts). We collected 12 types of 3D Presentation Attacks (PAs) classified into four material categories. The first category consists of textile/fiber-based materials, including White Cotton Glove, Cloth Glove, and Zebra-striped PU Coated Glove. These materials feature soft and breathable characteristics commonly found in everyday use. The second category comprises elastomer/polymer-based materials, including Anti-slip Glove, Rubber Glove, Disposable Plastic Glove, and Disposable Nitrile Glove. These materials typically exhibit good elasticity and water resistance and are widely used in medical and industrial applications. The third category includes adhesive-coated materials, specifically the Glue Application Anti-slip Glove and Yellow Latex Glove. These materials have special adhesive properties on their surfaces that may cause unique interference with palm vein recognition systems. The fourth category consists of prosthetic prototypes: Model Prosthetic Hand-1, Model Prosthetic Hand-2, and Model Prosthetic Hand-3. These prosthetic models represent more complex and sophisticated 3D presentation attack methods that may simulate various characteristics of real human palm vein.

As shown in Table II, our dataset splitting protocol already implements a realistic unknown attack evaluation: 12 attack types are used for training, while 4 attack types (grayscale photograph and prosthetic prototypes) are exclusively reserved for validation and testing. This ensures that the model has never seen these two attack types during training, providing a rigorous evaluation of generalization to unseen spoof attacks. Although PVASD already includes data captured by multiple commercial near-infrared sensors with five distinct spatial

TABLE II
DATASET DETAILS INCLUDING SUBJECTS/IMAGES SETTINGS

Category	PAs	Subjects	Images	Usage
2D Print PAs	grayscale photograph	54	146,287	validation/test
	infrared photograph	54	106,669	train
	color photograph	35	15,509	train
	A4 paper hand-shaped cutouts	2	4,799	train
3D PAs	prosthetic prototypes	3	8,593	validation/test
	elastomer/polymer-based	6	11,050	train
	adhesive-coated	3	4,025	train
	textile/fiber-based	6	10,346	train

resolutions and different optical characteristics, these devices still belong to the same category of commercially available palm vein imagers. Therefore, the current release does not yet provide a formal unseen-device evaluation protocol. To address this limitation, we are actively extending the dataset with additional acquisition sessions using previously unseen NIR and multispectral sensors that differ significantly in wavelength response, illumination geometry, and optical design. These new devices will be introduced in the next major release of PVASD, enabling standardized cross-device and unseen-device benchmarks. In parallel, we will explore sensor-robust training strategies, including physics-informed data augmentation as well as test-time adaptation methods, to further enhance model generalization across diverse and unforeseen imaging hardware without requiring device-specific retraining.

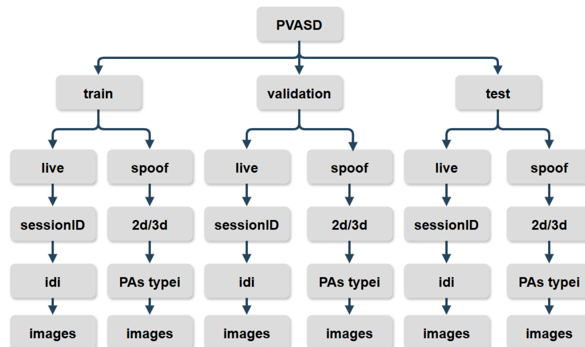

Fig. 5. Hierarchical organization of the PVASD.

B. Data Collection Methodology

Palm vein image acquisition was conducted using multiple commercial-grade near-infrared (NIR) sensors under carefully controlled conditions to simulate practical deployment scenarios while ensuring consistent image quality. Data were captured at five different resolutions, ranging from 640×400 to 1140×720 pixels, to enhance dataset diversity and support cross-resolution evaluations. The acquisition environment was systematically varied to encompass natural, artificial, and mixed illumination conditions, and incorporated multiple capture angles ($0°$, $\pm 15°$, and $\pm 30°$) and hand-to-sensor distances between 10 cm and 30 cm. Ambient temperature and humidity were regulated within the ranges of 10–35° C and 40–60%, respectively, to maintain physiological stability across sessions. All sensors were calibrated according to manufacturer guidelines prior to data collection to ensure uniform imaging quality.

For live sample acquisition, participants from diverse demographic backgrounds were recruited to ensure broad representativeness. Each subject contributed multiple sessions of palm vein images, captured under different environmental settings to document intra-subject variability. During acquisition, subjects were instructed to place their palms naturally on or above the sensor without excessive tension or deformation. Real-time monitoring was performed to verify correct hand placement, sufficient vein visibility, and appropriate focus. Post-capture quality control was employed to discard images exhibiting blur, overexposure, underexposure, or insufficient vascular contrast, thereby ensuring that only high-quality samples were retained in the dataset.

To comprehensively address potential security threats, spoof attack samples were systematically generated to cover both 2D and 3D presentation attack scenarios. For 2D attacks, various materials and techniques were utilized, including printed grayscale images, infrared photographs, color photographs, and A4 paper cutouts shaped as palms. For 3D attacks, diverse materials such as cloth gloves, rubber gloves, plastic gloves, latex gloves, and prosthetic hand models were employed. All attack samples were captured under the same environmental conditions and protocols as live samples to maintain consistency and realism.

Throughout the data collection process, detailed metadata including sensor type, resolution, environmental conditions, and capture parameters were recorded and associated with each image to facilitate subsequent analysis. Anonymization procedures were rigorously applied to protect participant privacy and duplicate or corrupted data were systematically removed. By combining technical consistency with environmental diversity, the resulting dataset provides a robust foundation for the development and evaluation of palm vein anti-spoofing algorithms under realistic operational conditions.

C. Dataset Organization and Structure

The PVASD is meticulously organized in a hierarchical structure designed to facilitate efficient access, comprehensive analysis, and standardized evaluation for researchers. Fig. 5 illustrates the overall organization of our dataset. The core dataset is structured into two primary categories (live samples and spoof samples) with further subdivisions based on anatomical, temporal, and attack-type considerations. Live samples are organized by subject ID, hand orientation (left/right), and capture session, allowing for systematic analysis of intra-subject variations, while spoof samples are categorized first by dimensionality (2D vs. 3D) and then by specific attack methodology and material type, enabling fine-grained analysis of different presentation attacks (PAs).

To ensure consistency and traceability, we implemented a standardized naming convention for all image files. Live samples and spoof samples follow the format: live_[ID][hand]_[scene]_[num]_[resolution].jpg, spoof_[ID][hand]_[scene]_[num]_[resolution].jpg or spoof_[PAs-type]_[hand] _[scene]_[num]_[resolution].jpg. This structured naming scheme enables rapid identification of key characteristics and facilitates automated parsing and classification of samples.

IV. Experiments

A. Evaluation Protocol

1) Dataset Partitioning: The PVASD is split into training (72%, 838,351 samples), validation (14%, 175,288 samples), and testing (14%, 173,880 samples) sets, ensuring no subject overlap between sets to prevent data leakage. A stratified sampling strategy was employed to maintain balanced distributions of attack types, resolutions, and environmental conditions across all subsets. Specifically, the training set includes approximately 685,953 live and 152,398 spoof samples, with similar proportional distributions in the validation and test sets.

2) Performance Metrics: We adopt standard metrics for comprehensive evaluation [48], [49]

$$APCER = \frac{FP}{FP + TN} \tag{9}$$

$$BPCER = \frac{FN}{FN + TP} \tag{10}$$

$$ACER = \frac{APCER + BPCER}{2} \tag{11}$$

where FP, FN, TP and TN represent false positives (spoof samples misclassified as live), false negatives (live samples misclassified as spoof), true positives and true negatives, respectively. APCER measures the error rate of accepting spoof attacks, BPCER evaluates the error rate of rejecting genuine samples, and ACER provides an overall performance indicator by averaging the two. Additionally, the Equal Error Rate (EER) [48] is used on the validation set to determine the threshold, which is then applied to compute APCER and BPCER on the test set.

B. Implementation Details

All models were trained using the Adam optimizer with an initial learning rate of 0.001, decayed via cosine annealing over 50 epochs. The batch size was set to 128× K, where K ranges from 1 to 10 based on model complexity and GPU memory constraints (e.g., K=1 for EfficientNetB7, K=2 for ResNet152, K=10 for MobileNetV3Large). Data augmentation included random horizontal flips (50% probability), rotations (±20°, uniform distribution), and brightness adjustments (±20%). Training was conducted on NVIDIA RTX3090 GPUs with 24GB memory. Early stopping was applied if validation loss did not improve for 5 consecutive epochs. Detailed configurations are provided in Table III.

C. Baseline Methods

To establish comprehensive benchmarks for palm vein anti-spoofing, we evaluated three categories of approaches, adapting all models for binary liveness classification (live vs. spoof) by replacing their classifier modules with a linear layer and sigmoid activation [50]. Detailed implementation parameters are provided in Table III. Batch sizes were selected empirically to maximize GPU utilization on a single NVIDIA RTX3090 (24GB), balancing training stability and convergence speed. Lightweight models like MobileNetV3Large used moderate batch sizes (e.g., 1280) to balance training efficiency and memory usage, while more complex models like ResNet152 required smaller batch sizes (e.g., 256) to avoid memory overflow on a single NVIDIA RTX3090 GPU. Deeper architectures like EfficientNetV2L typically needed even smaller batch sizes (e.g., 128) to accommodate their larger parameter counts and feature map dimensions while ensuring stable convergence.

TABLE III

Model Configuration Details

Parameter	Value	Notes
Input Resolution	128×128	3 channels
Batch Size	Model-dependent	EfficientNetV2L: 128, etc.
Optimizer	AdamW	β_1=0.9, β_2=0.999
Learning Rate	1e-3	weight decay
Training Epochs	50	Early stopping

(1) Classical Image Classification: We evaluated eight state-of-the-art CNN-based models: ConvnextLarge [38], DenseNet201 [39], EfficientNetB7 [40], EfficientNetV2L [41], MobileNetV3Large [42], RegNetX32gf [43], ResNet152 [44], and VGG19 [45]. These models range from lightweight architectures optimized for efficiency (e.g., MobileNetV3Large) to deep networks for robust feature extraction (e.g., ResNet152), adapted to capture palm vein patterns.

(2) Adapted Face Anti-Spoofing Methods: Two face anti-spoofing approaches were adapted for palm vein scenarios. BiFPNFAS [12] leverages an EfficientNet_b0 backbone and a bidirectional feature pyramid network to capture multi-scale features. EntryV1 [13] employs multi-scale spatial attention to focus on discriminative regions, fusing features for accurate liveness detection. Both were trained on PVASD to adapt to the characteristics of the palm vein.

(3) Anomaly Detection Approaches: We evaluated two anomaly detection methods. DevNet [46] uses deviation loss to establish normality baselines, re-trained on PVASD with adaptive thresholding for environmental variations. SimpleNet [47], adapted for global classification, employs multi-scale feature extraction to detect spoofing anomalies robustly.

D. Baseline Results

The experimental results for various baseline methods on the PVASD are summarized in Table IV, comparing computational complexity (FLOPs), validation performance (EER threshold and EER%), and testing metrics (APCER, BPCER, and ACER). Our comprehensive analysis reveals several key insights across the three methodological categories.

Among classical image classification models, MobileNetV3Large achieved the best overall performance with the lowest ACER of 3.620% and minimal computational requirements (0.154 G). This outstanding result suggests

TABLE IV
BASELINE PERFORMANCE ON THE PVASD. ↓ INDICATES LOWER VALUES ARE BETTER

Types	Baseline Methods	Flops	Validation		Test		
			EER threshold	EER(%)↓	APCER(%)↓	BPCER(%)↓	ACER(%)↓
Classical Image Classification	ConvnextLarge	22.5 G	0.9486	7.656	9.535	5.212	7.373
	DenseNet201	2.86 G	0.999	7.272	7.853	9.498	8.675
	EfficientNetB7	3.48 G	0.697	5.469	9.895	2.866	6.380
	EfficientNetV2L	8.1 G	0.838	4.711	8.357	2.373	5.365
	MobileNetV3Large	0.154 G	0.868	3.486	5.227	2.012	3.620
	RegNetX32gf	20.82 G	0.656	5.575	7.810	3.112	5.461
	ResNet152	7.58 G	0.767	10.513	7.670	5.926	6.798
	VGG19	13.02 G	0.9385	4.935	7.453	3.719	5.586
Face Anti-Spoofing Methods	BiFPNFAS	0.597 G	0.898	4.316	5.755	2.169	3.962
	EntryV1	1.84 G	0.868	2.454	6.061	1.793	3.927
Anomaly Detection Approaches	DevNet	4.84 G	5.038	12.300	10.600	6.439	8.520
	SimpleNet	6.04 G	0.989	5.636	4.908	3.088	3.998

that lightweight architectures with efficient building blocks are particularly well-suited for palm vein anti-spoofing tasks. The efficiency-performance balance demonstrated by MobileNetV3Large makes it exceptionally valuable for resource-constrained applications. EfficientNetV2L followed with 5.365% ACER (8.1 G), while VGG19 (5.586% ACER) outperformed deeper models like ResNet152 (6.798% ACER) and ConvnextLarge (7.373% ACER). This trend indicates that increased model complexity does not necessarily translate to better performance in palm vein anti-spoofing. For instance, EfficientNetB7, with 34.48 G FLOPs, achieved a higher ACER of 6.380% compared to MobileNetV3Large's 3.620%, underscoring the value of efficient architectural designs over simply increasing model depth or parameter count.

Face anti-spoofing methods, when adapted to palm vein scenarios, demonstrated strong but not superior performance. EntryV1 achieved an ACER of 3.927% and BiFPNFAS delivered 3.962% ACER, both slightly higher than MobileNetV3Large's leading result. However, BiFPNFAS's efficiency (0.597 G compared to EntryV1's 1.34 G) remains noteworthy, positioning it as a competitive option where computational resources are limited. EntryV1's particularly low BPCER of 1.793% indicates excellent ability to correctly authenticate genuine samples, though at the cost of slightly higher APCER compared to MobileNetV3Large.

In the anomaly detection category, SimpleNet (3.998% ACER) substantially outperformed DevNet (8.520% ACER), offering performance comparable to the face anti-spoofing methods but still trailing behind MobileNetV3Large. SimpleNet's balanced error distribution (APCER 4.908%, BPCER 3.088%) suggests good generalization capabilities, though not matching the leading classical model.

A clear trend emerges from these results: lightweight, efficient architectures specifically designed for mobile applications, like MobileNetV3Large, can outperform both specialized anti-spoofing methods and complex classification models in the palm vein domain. MobileNetV3Large's superior performance (3.620% ACER, 0.154G FLOPs) can be attributed to its inverted residual blocks with squeeze-and-excitation attention, which effectively capture subtle vascular patterns while maintaining computational efficiency. Compared to face anti-spoofing methods (e.g., EntryV1, 3.927% ACER), classical classification models benefit from general feature extraction capabilities, making them more adaptable to palm vein patterns. Anomaly detection methods like SimpleNet (3.998% ACER) show promise for detecting novel attacks but are less effective than lightweight classification models in this context. Moreover, more complex classification models like ResNet152 and ConvnextLarge, with significantly higher computational demands of 11.4 G and 15.5 G FLOPs, respectively, underperformed with ACERs of 6.798% and 7.373%.

This challenges the assumption that domain-specific methods would necessarily yield superior results, as MobileNetV3Large—a general image classification model—demonstrated better performance than both specialized and complex models in this context. Specifically, despite their tailored design, EntryV1 and BiFPNFAS failed to surpass MobileNetV3Large's ACER of 3.620%, with their own metrics lagging at 3.927% and 3.962%. The exceptional efficiency of MobileNetV3Large, achieving top performance with just 0.154 G FLOPs compared to models like EfficientNetV2L (5.365% ACER with 8.1 G FLOPs), highlights the importance of architectural efficiency in palm vein anti-spoofing. This represents a significant finding for practical deployment scenarios, demonstrating that sophisticated anti-spoofing capabilities can be achieved with minimal computational overhead.

These findings highlight three critical insights: (1) efficient architectural design can sometimes outweigh domain specialization for palm vein anti-spoofing, though specialized methods like EntryV1 and BiFPNFAS remain competitive; (2) lightweight models can achieve superior performance while minimizing computational requirements; and (3) MobileNetV3Large's inverted residual blocks with squeeze-and-excitation attention mechanisms appear particularly well-suited to capturing the subtle differences between genuine and spoofed palm vein patterns, likely due to their ability to focus on fine-grained features while maintaining computational efficiency. Future research should further explore the optimization of efficient architectures specifically for palm vein anti-spoofing applications.

E. Ablation Experiments

To thoroughly evaluate the effectiveness of our dataset and better understand the factors affecting palm vein anti-spoofing performance, we conducted four comprehensive ablation studies. These experiments investigate the impact of: (1) training

TABLE V

EFFECT OF LIVE SAMPLE QUANTITY ON ANTI-SPOOFING PERFORMANCE ACROSS DIFFERENT MODEL ARCHITECTURES. ↓ INDICATES LOWER VALUES ARE BETTER

Methods	LiveIdNum	Validation		Test		
		EER threshold	EER(%)↓	APCER(%)↓	BPCER(%)↓	ACER(%)↓
MobileNetV3Large	100	0.041	21.083	19.772	17.544	18.658
	500	0.092	14.597	18.631	7.372	13.001
	1,000	0.253	8.296	9.352	3.553	6.453
	2,000	0.816	6.085	9.348	2.223	5.785
	3,000	0.767	5.379	7.753	2.820	5.287
	4,000	0.888	4.052	5.858	2.353	4.105
EntryV1	100	0.001	13.635	6.124	18.131	12.128
	500	0.011	7.733	10.937	3.299	7.118
	1,000	0.384	5.581	10.473	2.167	6.320
	2,000	0.686	2.464	9.001	2.568	5.784
	3,000	0.939	3.605	6.783	2.282	4.532
	4,000	0.485	3.900	6.722	1.630	4.176
SimpleNet	100	0.001	15.484	14.753	11.203	12.978
	500	0.041	11.784	13.029	5.786	9.407
	1,000	0.545	9.275	9.112	3.178	6.145
	2,000	0.949	6.917	6.378	3.520	4.949
	3,000	0.969	5.340	6.247	2.847	4.547
	4,000	0.959	4.386	6.124	2.732	4.428

TABLE VI

IMPACT OF PRESENTATION ATTACK TYPE DIVERSITY ON ANTI-SPOOFING PERFORMANCE METRICS. ↓ INDICATES LOWER VALUES ARE BETTER

Methods	Spoof Type	Validation		Test		
		EER threshold	EER(%)↓	APCER(%)↓	BPCER(%)↓	ACER(%)↓
MobileNetV3Large	2d	0.797	5.037	9.904	3.619	6.761
	3d	0.979	5.388	6.529	5.497	6.013
	2d+3d	0.868	3.486	5.227	2.012	3.620
EntryV1	2d	0.918	6.632	8.801	3.842	6.322
	3d	0.999	7.938	5.916	8.094	7.005
	2d+3d	0.868	2.454	6.061	1.793	3.927
SimpleNet	2d	0.939	4.342	7.541	3.031	5.286
	3d	0.999	9.123	9.188	4.927	7.058
	2d+3d	0.989	5.636	4.908	3.088	3.998

data volume, (2) presentation attack diversity, (3) environmental conditions, and (4) dataset preprocessing. The first three ablation experiments used the identical test set as the Baseline Results (and partly reused those reported values).

(1) Effect of Training Data Volume. Table V demonstrates how the quantity of live samples in the training set affects model performance across our three representative approaches. For each method, we progressively increased the number of unique live subjects from 100 to 4,000 while maintaining consistent testing conditions. The results reveal a clear correlation between training data volume and anti-spoofing performance. For MobileNetV3Large, ACER decreased dramatically from 18.658% with 100 subjects to just 4.105% with 4,000 subjects—a relative improvement of 78.0%. Similar patterns emerged with EntryV1 (12.128% to 4.176%, 65.6% improvement) and SimpleNet (12.978% to 4.428%, 65.9% improvement). This consistent performance improvement across architectures underscores the critical importance of large-scale datasets for developing robust palm vein anti-spoofing solutions. Notably, the improvement rate diminishes as data volume increases, suggesting an asymptotic limit to performance gains from data volume alone. The relationship appears to follow a logarithmic pattern, with the most substantial improvements occurring in the early stages of data augmentation (from 100 to 1,000 subjects). Interestingly, EntryV1 required fewer samples to achieve comparable performance levels, reaching an ACER of 5.784% with just 2,000 subjects—performance that MobileNetV3Large could only approach with 3,000 subjects. This suggests that architectures incorporating attention mechanisms may utilize training data more efficiently for palm vein anti-spoofing tasks.

(2) Impact of Presentation Attack Diversity. Table VI examines how the diversity of presentation attacks in training data affects model robustness. We evaluated each model when trained on: (i) only 2D presentation attacks, (ii) only 3D presentation attacks, and (iii) both attack types combined. The results consistently demonstrate that models trained on both attack types outperform those trained on either type alone. MobileNetV3Large achieved an ACER of 3.620% when trained on combined attacks, compared to 6.761% with only 2D attacks and 6.013% with only 3D attacks. EntryV1 showed even more dramatic improvements, with the combined training yielding an ACER of 3.927% versus 6.322% (2D only) and 7.005% (3D only)—a relative improvement of 37.88% and 43.94%, respectively. Notably, models trained exclusively on 2D or 3D attacks exhibited asymmetric error distributions. For example, SimpleNet trained on 2D attacks showed higher APCER (7.541%) than BPCER (3.031%), indicating greater

TABLE VII

INFLUENCE OF ENVIRONMENTAL CONDITIONS ON PALM VEIN ANTI-SPOOFING PERFORMANCE. ↓ INDICATES LOWER VALUES ARE BETTER

Methods	Scene	Validation		Test		
		EER threshold	EER(%)↓	APCER(%)↓	BPCER(%)↓	ACER(%)↓
MobileNetV3Large	indoor	0.686	3.943	6.558	2.231	4.394
	outdoor	0.182	24.553	18.556	20.207	19.381
	indoor+outdoor	0.868	3.486	5.227	2.012	3.620
EntryV1	indoor	0.747	3.154	7.381	1.320	4.350
	outdoor	0.999	16.225	9.631	12.934	11.282
	indoor+outdoor	0.868	2.454	6.061	1.793	3.927
SimpleNet	indoor	0.828	6.670	7.987	4.414	6.201
	outdoor	0.707	17.073	14.888	9.327	12.107
	indoor+outdoor	0.989	5.636	4.908	3.088	3.998

difficulty in detecting novel 3D attacks. Conversely, when trained solely on 3D attacks, the same model showed more balanced errors but higher overall ACER (7.058%). These findings highlight the critical importance of presentation attack diversity in training data. Models exposed to a wider variety of attack vectors develop more robust feature representations that generalize better to unseen spoofing attempts. The complementary nature of 2D and 3D attack recognition appears to create synergistic effects when combined, suggesting that comprehensive anti-spoofing systems require diverse training across multiple attack dimensions.

(3) Influence of Environmental Conditions. Table VII investigates how environmental factors affect model performance by comparing training on: (i) indoor-only data, (ii) outdoor-only data, and (iii) combined indoor and outdoor data. The results reveal dramatic performance differences across environmental conditions. All three models performed substantially better on indoor data than outdoor data, with MobileNetV3Large showing a 4.394% ACER indoors versus 19.381% outdoors—a 4.4× performance gap. Similarly, EntryV1 achieved 4.350% ACER indoors compared to 11.282% outdoors (2.6× difference), and SimpleNet recorded 6.201% indoors versus 12.107% outdoors (2.0× difference). Most significantly, models trained on the combined indoor and outdoor dataset consistently outperformed those trained on either condition alone. EntryV1 achieved its best performance (3.927% ACER) with combined training, surpassing both indoor-only (4.350%) and outdoor-only (11.282%) training. This pattern held across all evaluated architectures. The substantial performance degradation in outdoor environments highlights specific challenges in palm vein anti-spoofing under variable lighting conditions. Natural illumination introduces complex variations in contrast, shadows, and infrared interference that complicate the distinction between genuine and spoofed presentations. However, exposure to these challenging conditions during training appears to enhance model robustness overall, as evidenced by the superior performance of models trained on combined environments.

(4) Effect of Data Preprocessing. To evaluate the impact of preprocessing on palm vein anti-spoofing performance, we compared two approaches: (i) no preprocessing (using original images) and (ii) our proposed ROI extraction pipeline. To enable a fair comparison, we constructed a new dataset by randomly sampling a subset of original images from the PVASD. Specifically, we extracted 114,107 images, including 84,554 live samples and 29,553 spoof samples. The dataset was partitioned into a training set 97,953 samples, 85.8%), validation set (8,973 samples, 7.9%), and test set (7,181 samples, 6.3%), ensuring no subject overlap across subsets to prevent data leakage. Random sampling was performed with a fixed seed (seed=42) to ensure reproducibility, maintaining a representative distribution of attack types, resolutions, and environmental conditions.

The new dataset was constructed to provide a unified baseline for comparing preprocessing methods, as the original PVASD includes images preprocessed with our ROI pipeline. By sampling original, unprocessed images, we ensured that all preprocessing methods were applied consistently to the same input images. The subset was drawn to balance computational feasibility with representativeness, prioritizing sufficient coverage of subjects and attack types. The live-to-spoof ratio (74:26) is the same as the original dataset (74:26).

For each subset (training, validation, test), we applied two preprocessing methods to the same set of original images: (i) No preprocessing: Original images were used directly without modification. (ii) Proposed ROI pipeline: MediaPipe Hands solution were used for palm keypoint localization, followed by ROI extraction and quality assessment, producing standardized 128 × 128 ROI images. The training set was preprocessed separately for each method to train three distinct models (MobileNetV3Large, EntryV1, SimpleNet). The validation and test sets were preprocessed correspondingly to evaluate model performance, ensuring consistency across all experiments.

As shown in Table VIII, MobileNetV3Large achieved an ACER of 0.310% with our proposed ROI preprocessing pipeline, compared to 0.644% (no preprocessing), representing a 51.9% improvement. Similar improvements were observed for EntryV1 (0.406% and 0.630%) and SimpleNet (0.399% and 0.855%). The lower error rates compared to baseline experiments (e.g. 3.620% ACER for MobileNetV3Large) are attributed to the smaller but more controlled dataset (114,107 vs. 1,187,519 images), which improved the sensitivity to spoof detection. Notably, our ROI pipeline significantly reduced BPCER (0.363% for MobileNetV3Large, 0.401% for EntryV1, 0.592% for SimpleNet), demonstrating its effectiveness in isolating discriminative vascular features for spoof detection. SimpleNet benefited most (53.3% ACER reduction), suggesting that simpler architectures rely heavily on optimized preprocessing.

TABLE VIII
INFLUENCE OF DATA PREPROCESSING ON PALM VEIN ANTI-SPOOFING PERFORMANCE. ↓ INDICATES LOWER VALUES ARE BETTER

Methods	Preprocessing Type	Validation		Test		
		EER threshold	EER(%)↓	APCER(%)↓	BPCER(%)↓	ACER(%)↓
MobileNetV3Large	No Preprocessing	0.535	0.007	0.257	1.032	0.644
	Our Methods	0.636	2.435	0.257	0.363	0.310
EntryV1	No Preprocessing	0.999	0.170	0.0	1.261	0.630
	Our Methods	0.757	0.474	0.411	0.401	0.406
SimpleNet	No Preprocessing	0.989	0.007	0.411	1.299	0.855
	Our Methods	0.969	3.106	0.206	0.592	0.399

To ensure fairness, all preprocessing methods were applied to the same set of original images using standardized parameters. The random sampling process used a fixed seed (seed=42) for reproducibility. Performance was evaluated using the same metrics (APCER, BPCER, ACER) and threshold (EER from the validation set) as in the baseline experiments, ensuring comparability. The preprocessing scripts and dataset splits will be included in the public PVASD repository to facilitate replication.

An overview of the studies on ablation. The ablation studies conducted on the PVASD provide several valuable insights for further palm vein anti-spoofing research. Our experimental study demonstrates that, for all evaluated designs, larger training sets lead to stronger anti-spoofing systems. Data scale is a critical factor in model performance. The fact that models trained on 4,000 participants have error rates up to 81.9% lower than those trained on restricted data provides strong empirical support for our strategy of building a large-scale dataset with thousands of subjects.

Additionally, our tests clearly demonstrate that a variety of presentation attack types are necessary to create systems that can handle a range of spoofing efforts. Models trained on heterogeneous attack types (both 2D and 3D) demonstrated superior performance compared to those specialized for a single attack category, with improvements in ACER ranging from 12.5% to 44.0%. This confirms the value of PVASD's comprehensive attack simulation methodology, which systematically incorporates multiple presentation attack instruments across different material categories and production techniques.

Despite the inherent difficulties caused by variable lighting conditions in outdoor settings, models trained in both indoor and outdoor environments demonstrated significantly improved generalization capabilities, achieving lower overall error rates. This finding supports our data collection strategy of purposefully including a variety of live environmental conditions to reflect actual usage scenarios. This is the third important finding from our ablation studies.

Fourth, we created an advanced workflow for prepping datasets that is especially tailored for palm vein anti-spoofing. This pipeline improves data quality and model performance by using the MediaPipe Hands solution for accurate palm region keypoint localization and a new ROI extraction method. It achieves an error rate reduction of up to 53.3%, making it an essential tool for reliable anti-spoofing research.

All things considered, these findings offer convincing empirical support for the PVASD's comprehensive design principles and preparation processes. PVASD provides a strong basis for the creation and assessment of palm vein anti-spoofing solutions that can be successfully implemented in actual-life scenarios by purposefully combining extensive subject participation, various demonstration attack types, and various environmental conditions with the data preprocessing strategy we proposed.

TABLE IX
ZERO-COST MITIGATION OF THE CATASTROPHIC OUTDOOR COLLAPSE ON MOBILENETV3LARGE (THE MOST SEVERELY AFFECTED MODEL IN TABLE VII)

Method	APCER (%)↓	BPCER (%)↓	ACER (%)↓
Original	18.53	**20.22**	19.37
+ CLAHE	**12.14**	24.17	**18.16**
Relative improvement	**34.5%**	+19.5%	**6.2%**

While the current preprocessing pipeline includes image quality control, minor blur and defocus may still affect palm vein visibility in practical scenarios. Future work will explore integrating image restoration models such as MB-TaylorFormer V2 [51], MC-Blur [52], and DBLRNet [53] to enhance the robustness of preprocessing under degraded conditions. Moreover, the proposed pipeline can be further complemented by palm vein enhancement modules to improve robustness under degraded image quality.

Illumination Normalization Baseline. To further analyze whether the illumination discrepancy contributes to the degradation of performance in outdoor environments, we tested a simple baseline photometric normalization (from Table VII) by applying CLAHE [54] to the test images without retraining. As shown in Table IX, CLAHE reduces APCER from 18.53% to 12.14%, indicating improved spoof detection under strong illumination. Although BPCER slightly increases (20.21% to 24.17%), the overall ACER decreases from 19.37% to 18.16%. This confirms that illumination shift is a key factor affecting outdoor performance and motivates future research on the extraction of learnable illumination-invariant characteristics.

F. Visualization Analysis of Model Attention

To interpret the superior performance of MobileNetV3Large (ACER=3.620%, APCER=5.227%) over complex models like ResNet152 (ACER=5.163%), we apply Grad-CAM [55] on PVASD samples (Fig. 6). On **live samples**, MobileNetV3Large generates focused activation on vein bifurcations and density variations (c), while ResNet152 produces diffuse activation across the entire palm (e), indicating overfitting to irrelevant

TABLE X
VULNERABILITY TO AI-GENERATED (DEEPFAKE) PALM-VEIN ATTACKS ON MOBILENETV3LARGE. ↓ INDICATES LOWER VALUES ARE BETTER

Training Contains AI-Spoofs	APCER (%) ↓	BPCER (%) ↓	ACER (%) ↓	ΔACER vs Original
Physical attacks only (original)	5.227	2.012	3.620	–
Yes (seen during training)	6.641	3.009	4.825	+33.3% (relative)
No (completely unseen)	**9.679**	2.012	**5.846**	**+61.5% (relative)**

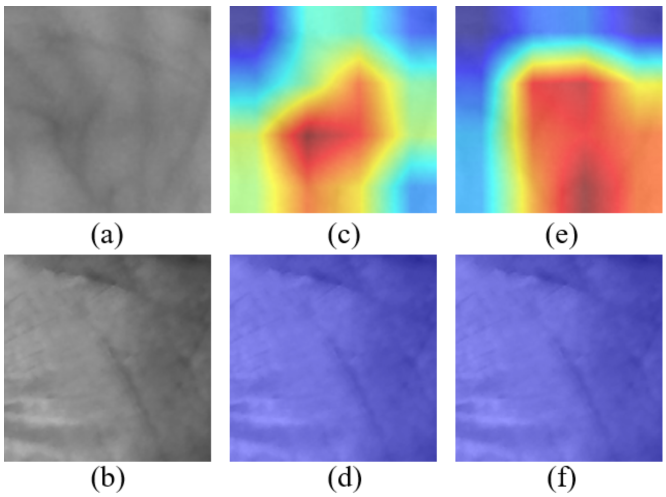

Fig. 6. Grad-CAM visualization on PVASD samples. (a) Live palm image. (b) Spoof (printed) image. (c) MobileNetV3Large generates focused activation on vein bifurcations in live samples. (d) MobileNetV3Large shows dispersed activation on spoof samples, avoiding overfitting to artifacts. (e) ResNet152 exhibits diffuse activation across the entire palm in live samples. (f) ResNet152 produces near-zero activation on spoof samples.

textures. On **spoof samples**, both models show weak or dispersed activation (d, f), but this is *expected and beneficial*: spoof images lack consistent vascular patterns, and MobileNetV3Large avoids being misled by printing artifacts.

These results provide strong visual evidence that the inverted residual blocks with squeeze-and-excitation (SE) attention enable lightweight models to prioritize genuine vascular cues while maintaining robustness to spoof variations, explaining their dominance in both ACER and APCER. This analysis indicates that the superior performance of MobileNetV3Large is not merely architectural, but emerges from the specific characteristics of PVASD — where fine-graned vascular patterns dominate over background cues — highlighting a dataset-driven insight for future model design.

G. Vulnerability to Deepfake Presentation Attacks

To address emerging threats from AI-generated (deepfake) spoofs, we extended PVASD with 20,000 high-fidelity synthetic spoof images generated using PVTree [56]. We evaluated vulnerability in two controlled settings: (i) **seen during training**: 12,000 synthetic spoofs added to the training set (4,000 each to validation/test), with models retrained from scratch; (ii) **completely unseen during training**: no synthetic data in training, only added to validation/test. The best-performing model (MobileNetV3Large) was used with hyperparameters identical to those of the baseline experiments.

Results are summarized in Table X. Even when the model is exposed to 12,000 synthetic spoofs during training, performance still degrades substantially, with ACER rising from 3.620% to 4.825% — a relative increase of 33.3%. More alarmingly, when AI-generated spoofs are completely unseen during training, APCER nearly doubles from 5.227% to 9.679%, pushing ACER to 5.846% and yielding a catastrophic 61.5% relative increase. This sharp deterioration demonstrates that current palm-vein anti-spoofing systems, including the best-performing lightweight architectures, remain extremely fragile against high-fidelity deepfake presentation attacks. Analysis of the confusion matrices further reveals that the vast majority of false acceptances originate from synthetic samples that successfully replicate realistic vein textures and skin appearance, underscoring the urgent need for specialized generative-robust defenses in this domain.

The dataset of 20,000 AI-generated spoof images will be publicly released as the **PVASD-deepfake** upon acceptance, establishing the first standardized benchmark for deepfake defending research in palm-vein anti-spoofing.

V. CONCLUSION AND DISCUSSION

In this paper, for palm vein anti-spoofing, we presented and released advance dataset preprocessing pipeline, a large-scale dataset (PVASD) and first benchmark code. Our preprocessing pipeline significantly enhances model performance by reducing error rates by up to 53.3%. Moreover, the PVASD is the first large-scale resource specifically designed for palm vein anti-spoofing research. It includes 16 different types of spoofing attack (both 2D and 3D) collected under varied conditions. Besides, we have taken a significant step beyond traditional physical attacks by generating 20,000 high-fidelity AI-generated spoof images using PVTree. Additionally, our benchmark evaluations indicate that the challenges in palm vein anti-spoofing differ from those in face anti-spoofing. This difference stems from near-infrared imaging and the complex vascular patterns. The experimental results demonstrated the suboptimal outcome of directly applying conventional face recognition techniques to spoof detection tasks. In the meantime, anomaly detection-based approaches showed promising detection capabilities against novel attack vectors. By increasing the data volume, enriching spoof type variations, and enhancing environmental conditions to expand the diversity of the training dataset, there was a significant reduction in error rates compared to conventional baseline implementations.

Despite its comprehensive nature, the PVASD has several limitations. The dataset relies on existing commercial sensors without incorporating newer sensor technologies, and extreme environmental conditions remain underrepresented. While rigorous anonymization reduces privacy risks, concerns about potential biometric data misuse persist. Future research should focus on developing specialized neural network architectures tailored for palm vein anti-spoofing using vascular pattern analysis and near-infrared imaging properties. Key directions include examining temporal dynamics of vein patterns, detecting novel spoofing attacks, and incorporating additional physiological indicators such as blood flow patterns. The dataset should be expanded with new sensor technologies, increased geographic diversity. Besides, we will integrate the entire acquisition–preprocessing–detection process into an end-to-end evaluation framework to assess real-world system performance. Additionally, leveraging geometry-aware foundation models tailored for vulnerable object categories [57] and advanced multi-sensor alignment strategies developed in cooperative perception [58], [59] could significantly improve generalization across unseen devices, novel attack types, and extreme environmental conditions.

PVASD also reveals several concrete directions for future baseline development. First, the severe performance degradation in outdoor scenes suggests integrating illumination-robust preprocessing (e.g., infrared histogram alignment, spectral normalization, or photometric domain randomization) as lightweight baseline modules. Second, the surprisingly strong performance of MobileNetV3Large indicates that SE-style channel attention is particularly effective for capturing fine-grained vascular textures; thus, incorporating SE-like attention into deeper architectures may yield more robust baselines. Finally, the dataset shows that background artifacts strongly affect non-attention models, implying that background suppression or texture-aware auxiliary losses can serve as principled baseline strategies. These directions demonstrate that PVASD not only provides data and benchmarks, but also exposes design principles that can guide future palm-vein anti-spoofing research.

The PVASD, preprocessing pipeline, and benchmarks provide a foundation for systematic palm vein anti-spoofing research, enabling reproducibility and fair method comparisons in academia while informing more secure recognition systems in industry. These advances support stronger biometric authentication, critical infrastructure protection, and non-invasive security solutions. We will release the data preprocessing pipeline, PVASD, and benchmark code to accelerate progress in palm vein anti-spoofing and enable researchers to develop enhanced security solutions.

REFERENCES

[1] D. Luo, Y. Qiao, D. Xie, S. Zhang, and W. Kang, “Palm vein recognition under unconstrained and weak-cooperative conditions,” *IEEE Trans. Inf. Forensics Security*, vol. 19, pp. 4601–4614, 2024.

[2] D. Luo, J. Huang, W. Yang, M. Saad Shakeel, and W. Kang, “RSNet: Region-specific network for contactless palm vein authentication,” *IEEE Trans. Inf. Forensics Security*, vol. 20, pp. 2734–2747, 2025.

[3] T. Wu, L. Leng, M. K. Khan, and F. A. Khan, “Palmprint-palmvein fusion recognition based on deep hashing network,” *IEEE Access*, vol. 9, pp. 135816–135827, 2021.

[4] F. O. Babalola, Ö. Toygar, and Y. Bitirim, “Boosting hand vein recognition performance with the fusion of different color spaces in deep learning architectures,” *Signal, Image Video Process.*, vol. 17, no. 8, pp. 4375–4383, Nov. 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:259887001

[5] T. Xie and W. Kang, “A random-binding-based bio-hashing template protection method for palm vein recognition,” *IEEE Trans. Inf. Forensics Security*, vol. 20, pp. 4243–4255, 2025.

[6] G. Ananthi, J. R. Sekar, and S. Arivazhagan, “Human palm vein authentication using curvelet multiresolution features and score level fusion,” *Vis. Comput.*, vol. 38, no. 6, pp. 1901–1914, Jun. 2022, doi: 10.1007/s00371-021-02253-9.

[7] F. O. Babalola, Y. Bitirim, and Ö. Toygar, “Palm vein recognition through fusion of texture-based and CNN-based methods,” *Signal, Image Video Process.*, vol. 15, no. 3, pp. 459–466, Apr. 2021.

[8] P. Tome and S. Marcel, “On the vulnerability of palm vein recognition to spoofing attacks,” in *Proc. Int. Conf. Biometrics (ICB)*, May 2015, pp. 319–325.

[9] J. Schuiki, M. Linortner, G. Wimmer, and A. Uhl, “Attack detection for finger and palm vein biometrics by fusion of multiple recognition algorithms,” *IEEE Trans. Biometrics, Behav., Identity Sci.*, vol. 4, no. 4, pp. 544–555, Oct. 2022.

[10] W. Wu, S. J. Elliott, S. Lin, S. Sun, and Y. Tang, “Review of palm vein recognition,” *IET Biometrics*, vol. 9, no. 1, pp. 1–10, Jan. 2020.

[11] H. Wang, L. Su, H. Zeng, P. Chen, R. Liang, and Y. Zhang, “Anti-spoofing study on palm biometric features,” *Expert Syst. Appl.*, vol. 218, May 2023, Art. no. 119546.

[12] K. Roy et al., “Bi-FPNFAS: Bi-directional feature pyramid network for pixel-wise face anti-spoofing by leveraging Fourier spectra,” *Sensors*, vol. 21, no. 8, p. 2799, Apr. 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/8/2799

[13] N. Sergievskiy, R. Vlasov, and R. Trusov, “Generalizable method for face anti-spoofing with semi-supervised learning,” 2022, *arXiv:2206.06510*.

[14] Y. He, F. Peng, R. Cai, Z. Yu, M. Long, and K.-Y. Lam, “Category-conditional gradient alignment for domain adaptive face anti-spoofing,” *IEEE Trans. Inf. Forensics Security*, vol. 19, pp. 10071–10085, 2024.

[15] H. Li, W. Li, H. Cao, S. Wang, F. Huang, and A. C. Kot, “Unsupervised domain adaptation for face anti-spoofing,” *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 7, pp. 1794–1809, Jul. 2018.

[16] W. Wang, F. Wen, H. Zheng, R. Ying, and P. Liu, “Conv-MLP: A convolution and MLP mixed model for multimodal face anti-spoofing,” *IEEE Trans. Inf. Forensics Security*, vol. 17, pp. 2284–2297, 2022.

[17] D. Wang et al., “Wild face anti-spoofing challenge 2023: Benchmark and results,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2023, pp. 6380–6391.

[18] X. Tan, Y. Li, J. Liu, and L. Jiang, “Face liveness detection from a single image with sparse low rank bilinear discriminative model,” in *Proc. ECCV*, 2010, pp. 504–517.

[19] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li, “A face antispoofing database with diverse attacks,” in *Proc. 5th IAPR Int. Conf. Biometrics (ICB)*, Mar. 2012, pp. 26–31.

[20] A. Liu, Z. Tan, J. Wan, S. Escalera, G. Guo, and S. Z. Li, “CASIA-SURF CeFA: A benchmark for multi-modal cross-ethnicity face anti-spoofing,” in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2021, pp. 1178–1186.

[21] Y. Zhang et al., “CelebA-spoof: Large-scale face anti-spoofing dataset with rich annotations,” in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 70–85.

[22] S. Zhang et al., “CASIA-SURF: A large-scale multi-modal benchmark for face anti-spoofing,” *IEEE Trans. Biometrics, Behav., Identity Sci.*, vol. 2, no. 2, pp. 182–193, Apr. 2020.

[23] Z. Yu, A. Liu, C. Zhao, K. H. M. Cheng, X. Cheng, and G. Zhao, “Flexible-modal face anti-spoofing: A benchmark,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2023, pp. 6346–6351.

[24] S. Zhang et al., “A dataset and benchmark for large-scale multi-modal face anti-spoofing,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 919–928.

[25] Y. Zhou and A. Kumar, “Human identification using palm-vein images,” *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 4, pp. 1259–1274, Dec. 2011.

[26] Y.-P. Lee, “Palm vein recognition based on a modified $(2D)^2$LDA,” *Signal, Image Video Process.*, vol. 9, no. 1, pp. 229–242, 2015.

[27] G. K. O. Michael, T. Connie, L. S. Hoe, and A. T. B. Jin, “Design and implementation of a contactless palm vein recognition system,” in *Proc. Symp. Inf. Commun. Technol. - SoICT*, 2010, pp. 92–99.

[28] W. Kang and Q. Wu, “Contactless palm vein recognition using a mutual foreground-based local binary pattern,” *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 11, pp. 1974–1985, Nov. 2014.

[29] J.-C. Lee, “A novel biometric system based on palm vein image,” *Pattern Recognit. Lett.*, vol. 33, no. 12, pp. 1520–1528, Sep. 2012.

[30] Y. Zhou and A. Kumar, "Contactless palm vein identification using multiple representations," in *Proc. 4th IEEE Int. Conf. Biometrics, Theory, Appl. Syst. (BTAS)*, Sep. 2010, pp. 1–6.
[31] X. Sun, Q. Xu, C. Wang, W. Dong, and Z. Zhu, "ROI extraction for online touchless palm vein based on concavity analysis," in *Proc. 32nd Youth Academic Annu. Conf. Chin. Assoc. Autom. (YAC)*, May 2017, pp. 1123–1126.
[32] R. Kabaciński and M. Kowalski, "Vein pattern database and benchmark results," *Electron. Lett.*, vol. 47, no. 20, pp. 1127–1128, Sep. 2011.
[33] Institute of Automation, Chinese Academy of Sciences (CASIA). (2019). *CASIA Multi-Spectral Palmprint Database (CASIA-MS-PalmPrintV1)*. [Online]. Available: https://biometrics.idealtest.org
[34] D. Zhang, Z. Guo, G. Lu, L. Zhang, and W. Zuo, "An online system of multispectral palmprint verification," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 2, pp. 480–490, Feb. 2010.
[35] P. Tome and S. Marcel, "Palm vein database and experimental framework for reproducible research," in *Proc. Int. Conf. Biometrics Special Interest Group (BIOSIG)*, Sep. 2015, pp. 1–7.
[36] Ö. Toygar, F. O. Babalola, and Y. Bitirim, "FYO: A novel multimodal vein database with palmar, dorsal and wrist biometrics," *IEEE Access*, vol. 8, pp. 82461–82470, 2020.
[37] C. Lugaresi et al., "MediaPipe: A framework for building perception pipelines," 2019, *arXiv:1906.08172*.
[38] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 11976–11986.
[39] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.
[40] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 6105–6114. [Online]. Available: https://proceedings.mlr.press/v97/tan19a.html
[41] M. Tan and Q. V. Le, "EfficientNetv2: Smaller models and faster training," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 10096–10106. [Online]. Available: https://proceedings.mlr.press/v139/tan21a.html
[42] A. Howard et al., "Searching for MobileNetV3," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 1314–1324.
[43] I. Radosavovic, R. P. Kosaraju, R. Girshick, K. He, and P. Dollar, "Designing network design spaces," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 10425–10433.
[44] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[45] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
[46] G. Pang, C. Ding, C. Shen, and A. van den Hengel, "Explainable deep few-shot anomaly detection with deviation networks," 2021, *arXiv:2108.00462*.
[47] Z. Liu, Y. Zhou, Y. Xu, and Z. Wang, "SimpleNet: A simple network for image anomaly detection and localization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 20402–20411.
[48] S. Marcel, M. S. Nixon, J. Fierrez, and N. Evans, *Handbook of Biometric Anti-Spoofing: Presentation Attack Detection*, vol. 2. Cham, Switzerland: Springer, 2019.
[49] R. Ramachandra and C. Busch, "Presentation attack detection methods for face recognition systems: A comprehensive survey," *ACM Comput. Surveys*, vol. 50, no. 1, pp. 1–37, Jan. 2018.
[50] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, Oct. 1986.
[51] Z. Jin, Y. Qiu, K. Zhang, H. Li, and W. Luo, "MB-TaylorFormer v2: Improved multi-branch linear transformer expanded by Taylor formula for image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 47, no. 7, pp. 5990–6005, Jul. 2025.
[52] K. Zhang et al., "MC-blur: A comprehensive benchmark for image deblurring," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 5, pp. 3755–3767, May 2024.
[53] K. Zhang, W. Luo, Y. Zhong, L. Ma, W. Liu, and H. Li, "Adversarial spatio-temporal learning for video deblurring," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 291–301, Jan. 2019.
[54] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics Gems*, 1994, pp. 474–485.
[55] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 618–626.
[56] S. Shang et al., "PVTree: Realistic and controllable palm vein generation for recognition tasks," in *Proc. 39th AAAI Conf. Artif. Intell. 37th Conf. Innov. Appl. Artif. Intell. 15th Symp. Educ. Adv. Artif. Intell.*, Apr. 2025, vol. 39, no. 7, pp. 6767–6775, doi: 10.1609/aaai.v39i7.32726.
[57] Z. Meng, X. Xia, R. Xu, W. Liu, and J. Ma, "HYDRO-3D: Hybrid object detection and tracking for cooperative perception using 3D LiDAR," *IEEE Trans. Intell. Vehicles*, vol. 8, no. 8, pp. 4069–4080, Aug. 2023.
[58] Z. Meng, X. Xia, and J. Ma, "Toward foundation models for inclusive object detection: Geometry- and category-aware feature extraction across road user categories," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 54, no. 11, pp. 6570–6580, Nov. 2024.
[59] Z. Meng, Y. Zhang, Z. Zheng, Z. Zhao, and J. Ma, "AgentAlign: Misalignment-adapted multi-agent perception for resilient inter-agent sensor correlations," 2024, *arXiv:2412.06142*.

Caiping Yan (Member, IEEE) received the Ph.D. degree in computer science from the University of Macau, China, in 2017. She is currently with Hangzhou Normal University, Hangzhou, China. Her current research interests include digital forensics and image processing.

Zhi Lan received the M.S. degree from the Department of Computer Science, Hangzhou Normal University, Hangzhou, China, in 2023. His current research interests include image processing.

Hong Li received the Ph.D. degree in computer science from the University of Macau, China, in 2017. His current research interests include image processing, semi-supervised learning, and computer vision.

Yuqi Li received the B.S. degree from the School of Artificial Intelligence, Guangxi Minzu University, Nanning, China, in 2023. Her current research interests include image processing.

Zonglin Meng received the bachelor's degree in computer science and the Ph.D. degree from UCLA. He is currently a Post-Doctoral Researcher with the UCLA Samueli School of Engineering. His work has been published in major journals, including IEEE TRANSACTIONS ON INTELLIGENT VEHICLES, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS, and IEEE INTERNET OF THINGS. He also published recent work in computer vision conferences, including the Computer Vision and Pattern Recognition (CVPR) and European Conference and Computer Vision (ECCV). His current research interests include autonomous driving systems, including perception, connected vehicles, multi-agent sensor fusion, and multi-modal sensor fusion.