

# LSAP-PV: High-Fidelity Palm Vein Image Synthesis via Layered Spectral Absorption Projection-Guided Diffusion Model

Sheng Shang<sup>1\*</sup>, Chenglong Zhao<sup>2\*</sup>, Ruixin Zhang<sup>2</sup>, Jianlong Jin<sup>1</sup>, Jingyun Zhang<sup>3</sup>, Jun Wang<sup>3</sup>, Yang Zhao<sup>1</sup>, Shouhong Ding<sup>2†</sup>, Wei Jia<sup>1†</sup>

<sup>1</sup> School of Computer Science and Information Engineering, Hefei University of Technology, China

<sup>2</sup> Tencent Youtu Lab

<sup>3</sup> Tencent WeChat Pay Lab33

shengshang@mail.hfut.edu.cn, {lornezhao, ruixinzhang}@tencent.com, jianlong@mail.hfut.edu.cn, {naskyzhang, earljwang}@tencent.com, yzhao@hfut.edu.cn, ericshding@tencent.com, jiawei@hfut.edu.cn

## Abstract

Palm vein recognition has emerged as a promising biometric technology, yet its development remains constrained by the scarcity of large-scale publicly available datasets. Several methods of palm vein image generation have been proposed to address this issue. These methods usually focus on the anatomical realism of palm vein patterns, but overlook the biophysical correlation between identities and vein patterns, particularly in simulating identity-specific vein contrast. To tackle this limitation, we propose a novel biophysics-driven synthesis method. Our method constructs a 3D palm vascular tree via an established modeling method. Then, a projection model is proposed to map the 3D tree into 2D space to derive palm vein patterns. The projection model is based on skin spectral absorption and simulates the natural attenuation of light passing through the skin using a layer integration method. For different identities, we sample different skin parameters, resulting in varying degrees of attenuation. This method effectively simulates the variation in vein contrast across different identities. Furthermore, we introduce a conditional diffusion model that uses the projected patterns as identity conditions to generate palm vein images. To the best of our knowledge, this is the first palm vein generation method based on the diffusion model. Experimental results demonstrate that our method not only outperforms existing methods, but also enables a recognition model trained on our synthetic data to achieve superior performance compared to a model trained on real-world data at a scale of 2,000 IDs under an open-set protocol with a  $TAR@FAR = 1 : 1$  of  $1e - 4$ .

**Code** — [https://github.com/Sunniva-Shang/LSAP\\_PV-Layered-Spectral-Absorption-Projection](https://github.com/Sunniva-Shang/LSAP_PV-Layered-Spectral-Absorption-Projection)

## Introduction

As a highly promising biometric technology, palm vein recognition has garnered significant attention from both academia and industry in recent years. Palm veins are hidden under the skin and can be collected in a non-contact way.

\*These authors contributed equally.

†Corresponding authors.

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Therefore, palm vein recognition has better hygienic safety and privacy protection. Several leading technology companies, such as Amazon (guide 2020) and Tencent (Tencent 2024), have released various palm vein recognition products, which indicates that the era of large-scale industrialization of palm vein recognition has gradually arrived. However, the research on palm vein recognition is currently constrained by a critical challenge, that is, the lack of large-scale publicly available datasets for training recognition models. Generating large-scale synthetic datasets is one of the feasible solutions to address this challenge.

Existing palm vein generation methods mainly focus on modeling anatomically realistic palm vein patterns. These methods typically employ 2D modeling-based methods (Salazar et al. 2021a; Ou et al. 2022; Salazar-Jurado et al. 2024) to model palm veins with a stereoscopic structure, producing unrealistic results. PVTree (Shang et al. 2025), as a 3D modeling-based method, constructs 3D palm vascular tree and projects it onto a 2D plane, achieving visually realistic palm vein patterns. However, it fails to explain the biophysical correlation between individual identities and vein patterns. As a result, generating palm vein images with identity-specific vein contrast becomes impossible.

In reality, palm veins are located beneath the skin, which means that the collected palm vein images contain not only the inherent vein patterns distribution but also the optical interaction relationship between the skin components and incident light. The interaction relationship is a crucial biophysical correspondence that determines vein contrast in collected images. Vein contrast refers to the luminance difference between veins and surrounding skin tissues, with higher contrast indicating clearer vein visualization. Real-world palm vein data exhibit varying levels of clarity across different identities. However, PVTree’s depth-to-intensity normalization projection method ignores this variation, unifying the vein contrast of different identities into the same range. This unification compromises the discriminability of identity and disrupts the biophysical correspondence between individual identities and vein patterns.

To address this issue, we improve the projection method of PVTree and propose a novel biophysics-driven layered

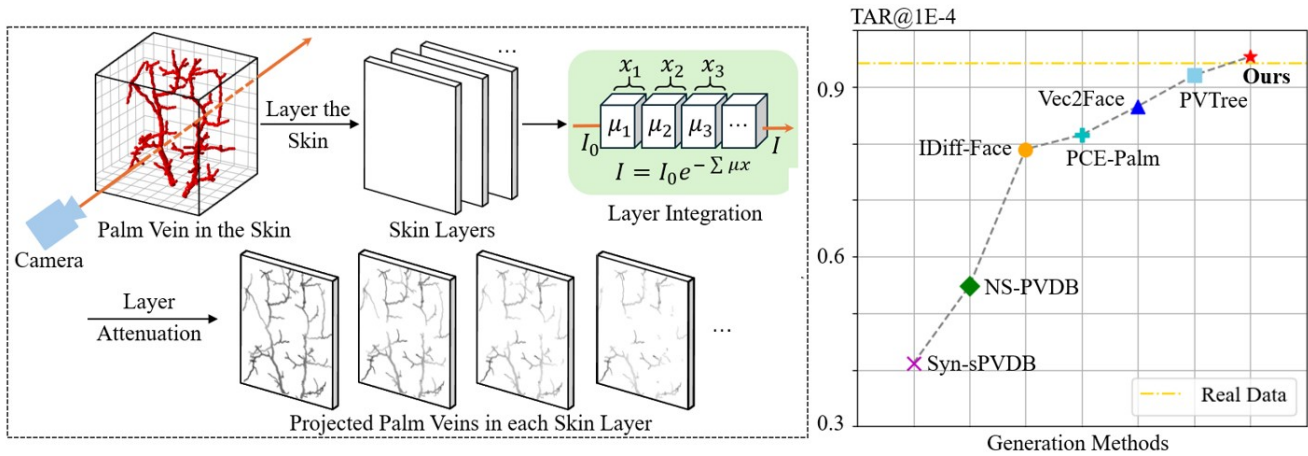


Figure 1: Layered spectral absorption projection (LSAP) and comparison of different generation methods in recognition tasks. **Left:** LSAP model. We model the 3D palm vein projection as light attenuation in a multi-layered medium. **Right:** The average accuracy of recognition models trained with synthetic data generated by different methods on publicly available datasets.

spectral absorption projection (LSAP), which aims to bridge the authenticity gap in 3D-to-2D projection and improve the physiological fidelity of the projected palm vein patterns. When light penetrates human skin, the light-absorbing media in the skin, such as water, melanin, collagen, and blood, absorb the light, resulting in attenuation of the light. The degree of attenuation depends on both skin thickness and concentration of light-absorbing media, with greater thickness and concentration inducing a greater degree of attenuation. Differences in these parameters account for observed variations in vein contrast. By parameterizing the skin model with the skin thickness and concentrations of light-absorbing media, we establish a physiologically meaningful identity representation that directly controls inter-identity variations in vein contrast. The Beer-Lambert law (Swinehart 1962) defines the light attenuation intensity as it passes through a homogeneous medium. Based on this law, human skin is modeled as a multi-layer homogeneous medium, allowing the calculation of light attenuation by using a layer integration method, as illustrated in Fig.1. For different identities, we randomly sample skin parameters from clinical distributions, thereby generating palm vein patterns with identity-specific vein contrast.

After obtaining the palm vein patterns that represent identity features, we introduce diffusion model (Ho, Jain, and Abbeel 2020) to achieve the domain transfer from patterns to images. The image-to-image translation method relies on paired data for training. Existing methods use palm crease energy extraction module (PCEM) (Jin et al. 2024) to extract the corresponding pattern from palm vein image, which is the binarized image and exhibits a significant domain gap with the synthetic patterns. To mitigate the domain gap, we propose a vein pattern extraction module (VPEM) trained on paired pseudo-palm vein data. It is able to achieves domain transfer from real palm vein images to synthetic patterns, effectively reducing the domain gap between extracted and synthetic patterns.

Our method effectively solves the problem of the inabil-

ity to simulate identity-specific vein contrast in the field of palm vein generation, further improving the authenticity of generated images. In summary, our contributions are as follows:

- Our projection model is the first to incorporate biophysical priors for controlling vein contrast. Based on skin spectral absorption, the model can obtain palm vein patterns with identity-specific vein contrast.
- We propose a vein pattern extraction module trained on synthetic data and obtain patterns from real-world data that are similar to the synthetic patterns, which effectively bridge the domain gap between extracted and synthetic patterns.
- This is the first palm vein generation method based on diffusion model. Compared to existing GAN-based generation methods, our method achieves more realistic and higher-quality palm vein images.
- We conduct extensive recognition experiments on several public palm vein datasets and synthetic datasets. The experimental results demonstrate that our method significantly outperforms existing methods.

## Related Work

### Palm Vein Recognition Methods

Traditional palm vein recognition methods rely on hand-crafted feature extraction algorithms to capture both local and global features (Wang et al. 2016; Abed et al. 2020; Sun et al. 2021). These methods limit expressive power across different datasets and exhibit poor generalization. With the rapid development of deep learning, many deep learning-based methods have been introduced into the field of biometric recognition. These methods aim to optimize deep neural networks to extract discriminative and robust palm vein features (Wang, Sun, and Sowmya 2019; Li et al. 2022; Htet and Lee 2023; Qin et al. 2023). Compared to traditional methods, deep learning-based methods more accu-

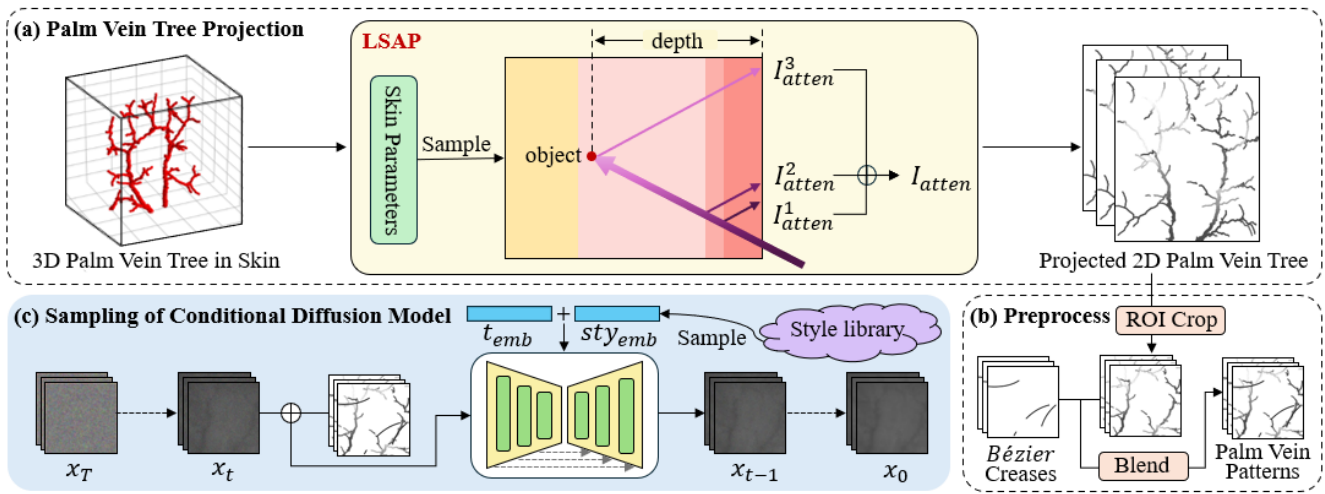


Figure 2: Pipeline of the proposed method. First, the PVTree model is used to generate a 3D palm vascular tree. Then, we project it onto a 2D space using our LSAP model. Next, Bézier creases (Zhao et al. 2022) are used to simulate the palmprint, which are then blended with the projected palm vein patterns to obtain the final patterns. Finally, we employ a condition diffusion model to generate palm vein images. A discrete style library provides randomly sampled style conditions, while the patterns act as identity conditions, yielding palm vein images with specified styles and identities.

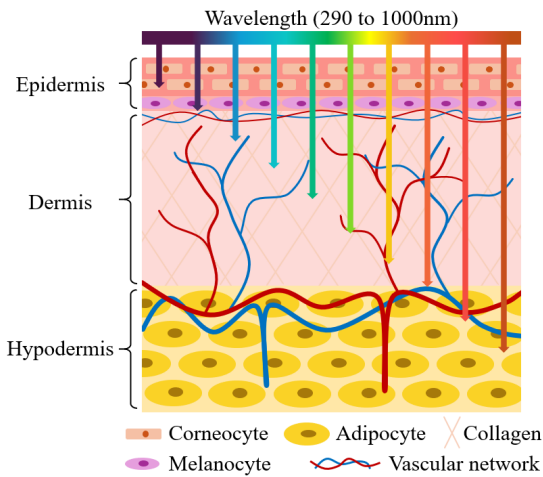


Figure 3: Multi-layered skin model.

rately represent palm vein characteristics. However, they are often constrained by the limited scale of publicly available datasets.

### Palm Vein Generation Methods

Given the scarcity of publicly available palm vein datasets, employing generative models like StyleGAN (Karras et al. 2020) to directly sample noise from the latent space for palm vein image generation (Salazar et al. 2021b) often results in less-than-ideal outcomes. To mitigate the complexity of generation, existing methods usually divide the task of palm vein image generation into two subtasks: the modeling of palm vein patterns and the subsequent rendering of palm vein images. Existing methods for palm

vein modeling include agent-based growth algorithms of the Physarum (Salazar et al. 2021a), the random block composition method (RBC) (Ou et al. 2022), mathematical modeling method (Salazar-Jurado et al. 2024), and 3D modeling method PVTree. Following this, GAN-based models are introduced to facilitate the rendering of palm vein images. However, these methods overlook the biophysical correspondence between individual identities and vein patterns, making it impossible to simulate the variations in vein contrast among different identities.

### Image-to-Image Translation Methods

Image-to-image translation methods aim to transform one representation of an image into another while preserving the main features and content of the image. With advancements in deep learning, particularly in GAN and diffusion models, this field has made significant strides. GAN-based methods, such as Pix2Pix (Isola et al. 2017) and CycleGAN (Zhu et al. 2017), achieve domain-to-domain image translation by leveraging the adversarial relationship between the generator and the discriminator, along with various constraint loss functions. Diffusion model-based methods (Rombach et al. 2022; Yang et al. 2025) employ image encoders to process the input images and then use diffusion models to progressively denoise the images, resulting in high-quality outputs. In recent years, diffusion models have demonstrated powerful generation capabilities and have garnered increasing research and application interest.

## Methods

### Overall Framework

Our generation method is a two-stage process: first, palm vein patterns are modeled to obtain identity conditions, and

then a diffusion model is introduced to synthesize palm vein images. The overall framework is illustrated in Fig.2.

In this section, we first discuss the anatomy of the skin and derive a multi-layered skin model. Following this, we propose a projection method based on the skin model. Finally, we introduce the diffusion model for generating palm vein images.

### Multi-Layered Skin Model

To better simulate light attenuation in the skin, we propose a multi-layered skin model. As shown in Fig.3, the typical skin model consists of three layers: the epidermis, dermis, and hypodermis (Footner et al. 2023; Nunez 2009). This paper focuses solely on the skin components that exhibit light-absorbing properties.

The epidermis primarily consists of the upper basal tissue and melanocytes located in the stratum basale. The upper basal tissue protects the skin and prevents water evaporation, while melanocytes secrete melanin, which protects the skin from ultraviolet damage. The dermis contains an extensive vascular network and collagen. The vascular network supplies essential nutrients and oxygen to skin cells, and the collagen provides structural integrity and support to the skin. The hypodermis is mainly composed of adipocytes and large blood vessels. Adipocytes store energy for the body, and the large blood vessels extend upward to form the vascular network of the dermis. The multi-layered skin model enables a more precise description of light absorption in different skin layers, which is crucial for understanding the interaction between light and skin.

### Layered Spectral Absorption Projection

Based on the proposed multi-layered skin model, we propose a layered spectral absorption projection model (LSAP). This model combines the skin spectral absorption to project the 3D palm vascular tree modeled by PVTree into a 2D space.

According to the Beer-Lambert law, when a beam of light passes through a homogeneous medium of a certain thickness, the attenuated light intensity  $I_{atten}$  follows the law:

$$I_{atten} = I_0 e^{-\mu_\lambda x}, \quad (1)$$

where  $I_0$  is the initial light intensity,  $\mu_\lambda$  is the absorption coefficient of the medium for light with a wavelength of  $\lambda$ , and  $x$  is the thickness of the medium. Therefore, for non-uniform media such as human skin, we use path integration to calculate the attenuated light intensity:

$$I_{atten} = I_0 e^{-\int \mu_\lambda(x) dx}, \quad (2)$$

where  $\mu_\lambda(x)$  represents the absorption coefficient of the medium at a depth of  $x$  for light with a wavelength of  $\lambda$ .

Given the impracticality of acquiring the absorption coefficient at each skin depth, we approximate the path integral using a discretized summation over segmented skin layers. Based on the multi-layered skin model, we represent the skin as a multi-layer homogeneous medium. Specifically, for the skin layer containing only a single medium, its absorption coefficient is determined by that medium. For the skin layer

Layer	Thickness( $\mu m$ )	Tissues	Concentration(%)
1	85 to 145	water	$1 - C_{me}$
2	15	melanin	0.8 to $43(C_{me})$
		water	$1 - C_b - C_{co}$
3	1310 to 3310	blood	0.25 to $2(C_b)$
		collagen	15 to $30(C_{co})$

Table 1: Sampling range of parameters for skin models (Vyas, Banerjee, and Burlina 2013).

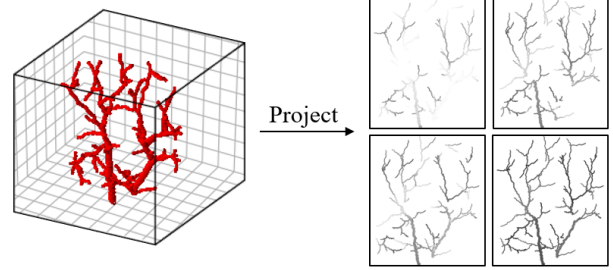


Figure 4: Projection results of the same palm vascular tree under different skin parameters.

containing multiple media, its absorption coefficient is related to the concentration of these media. In practice, we focus on the epidermis and dermis, as these are the regions where near-infrared light penetrates most easily.

The epidermis contains water and melanin, which exhibit relatively distinct stratified structures. Therefore, it is modeled as two skin layers containing a single medium. The dermis contains water, blood, and collagen, which are mixed without distinct stratification. Hence, it is modeled as a single skin layer containing multiple media. Ultimately, we get a three-layered skin model comprising the upper epidermis, the melanin layer, and the dermis layer. The absorption coefficient for each layer is calculated as follows:

$$\begin{cases} \mu_1 = (1 - C_{me})\mu_{wt}; \\ \mu_2 = C_{me}\mu_{me}; \\ \mu_3 = (1 - C_b - C_{co})\mu_{wt} + C_b\mu_b + C_{co}\mu_{co}, \end{cases} \quad (3)$$

where  $\mu_i, i = 1, 2, 3$  represent the absorption coefficients of each layer, and  $\mu_{wt}, \mu_{me}, \mu_b, \mu_{co}$  represent the absorption coefficients of water, melanin, blood, and collagen, respectively. Additionally,  $C_{me}, C_{co},$  and  $C_b$  denote the concentrations of melanin, collagen, and blood, respectively. Thus, the layered spectral absorption projection is modeled as:

$$I_{atten} = I_0 exp\left(-\sum_i \mu_i(\lambda)x_i\right), \quad (4)$$

where  $\mu_i(\lambda)$  represents the absorption coefficient of the  $i$ -th skin layer for light with a wavelength of  $\lambda$ , and  $x_i$  represents the thickness of the  $i$ -th skin layer.

After establishing the projection model, we need to determine the model parameters. Aligning with public benchmark datasets that predominantly use 850 nm near-infrared

illumination, we standardize  $\lambda = 850$  for deriving absorption coefficients (Nunez 2009). Notably, our method supports generalization to arbitrary wavelengths. Additionally, for different identities, the concentrations of the media within the skin and the thickness of each layer vary, which results in diverse vein contrast in the collected palm vein images. Therefore, we set sampling regions for the relevant parameters. By sampling different media concentrations and skin thicknesses for different identities, we effectively simulate the variations in vein contrast. Table 1 lists the sampling regions of the relevant parameters, and Fig.4 illustrates the projection results of the same 3D palm vascular tree by sampling various skin parameters.

Additionally, real-world palm vein images of the same identity exhibit local variations due to the influence of the collection environment, which specifically manifests as the absence or lack of clarity of certain palm vein patterns. To simulate this, we further refine the projection method of PVTree by randomly masking portions of the patterns during projection. Besides, during palm vein imaging, the results are significantly affected by palmprints, exhibiting prominent palmprint lines. Therefore, we integrate state-of-the-art B’ezier creases to model prominent palmprint lines. Following this, we adhere to the PVTree settings, cropping the region of interest(ROI), and blending it with the Bézier creases to obtain the final palm vein patterns.

## Palm Vein Images Generation

**Vein Pattern Extraction Module** Upon obtaining the palm vein pattern, we input it as an identity condition into the generation model to produce the palm vein image. This image-to-image generation model relies on paired data for training, necessitating the acquisition of corresponding patterns from real-world data. Existing methods employ PCEM for pattern extraction, but its resulting binarized images have a substantial domain gap compared to our depth-aware vein patterns that preserve rich anatomical information. To bridge this gap, we propose a vein pattern extraction module (VPEM) trained on synthetic data. Our pipeline first generates synthetic image-pattern pairs using the PVTree. This paired data enables supervised training of arbitrary image-to-image translation models. For VPEM’s architecture, we adopt CycleGAN due to its demonstrated effectiveness in image translation tasks. After training, VPEM is able to perform style transfer from images to patterns, effectively acquiring the ability to extract patterns from real-world data.

The extraction results of VPEM compared to the existing method are shown in Fig.5. The results demonstrate that, compared to the existing method, VPEM not only effectively captures the palm vein patterns in real-world data but also includes depth information. Furthermore, since VPEM is trained on synthetic patterns, the results extracted from real-world data exhibit a distribution similar to the synthetic patterns, thereby significantly minimizing the domain gap.

**Conditional Diffusion Model** The diffusion model is a generative model composed of two processes: the forward diffusion process and the reverse denoising process. During the forward diffusion process, the model incrementally corrupts the data by adding noise over a series of time steps  $t$ .

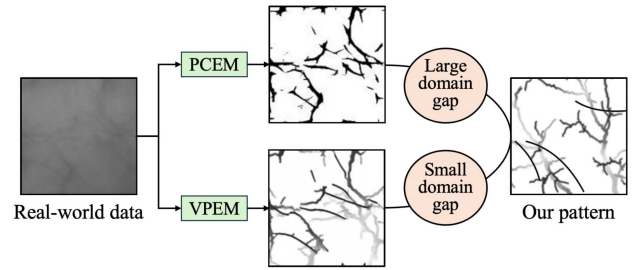


Figure 5: Visual comparison between palm vein patterns extracted by PCEM and VPEM, alongside our patterns.

At each step, a certain amount of noise is added, gradually transforming the data into complete noise by the final step. The process is formulated as:

$$q(x_t|x_{t-1}) = N(x_t; \sqrt{\alpha_t}x_{t-1}, (1 - \alpha_t)I), \quad (5)$$

where  $x_t$  represents the noisy image at step  $t$ , and  $\alpha_t$  is a pre-defined parameter. During the reverse denoising process, a noise prediction model  $\epsilon_\theta(x_t, t)$  is employed to estimate the added noise at each timestep, enabling stepwise reconstruction of the original data distribution through mean computation  $\mu_\theta(x_t, t, \epsilon_\theta)$ . The variance in this process is typically fixed and determined by predefined parameters.

**Style Library** To address style discrepancies caused by acquisition devices and environmental factors across public palm vein datasets, we introduce a discrete style library corresponding to distinct datasets. These style conditions are transformed into dense vector representations  $sty$  via an embedding layer and subsequently fused with the diffusion model’s timestep embeddings to modulate noise prediction  $\epsilon_\theta(x_t, t, sty)$ . This composite embedding modulates the denoising process at every timestep, enabling the diffusion model to generate realistic palm vein structures while exhibiting style characteristics aligned with the target domain. The process is formulated as:

$$p_\theta(x_t - 1|x_t) = N(x_{t-1}; \mu_\theta(x_t, t, \epsilon_\theta), (1 - \alpha_t)I). \quad (6)$$

During training, palm vein patterns extracted from real-world data by VPEM are used as conditions and concatenated with the input images for model training. In the inference stage, synthetic patterns provide an identity condition, while randomly sampled style code from the style library deliver stylistic control. This dual-conditioning framework synthesizes palm vein images with specified identity and target-domain stylistic attributes.

## Experiment

### Experimental Setup

**Datasets** We utilize four publicly available palm vein datasets: CASIA (Hao et al. 2008), PolyU (Zhang et al. 2009), HFUT(Shang et al. 2025), and TongJi (Zhang et al. 2018), comprising a total of 1,502 identities and 24,048 images. Table 2 provides detailed information about these datasets. Among these datasets, CASIA is a multispectral dataset, and we only use the images captured at a wavelength of 840nm. Following the same open-set evaluation

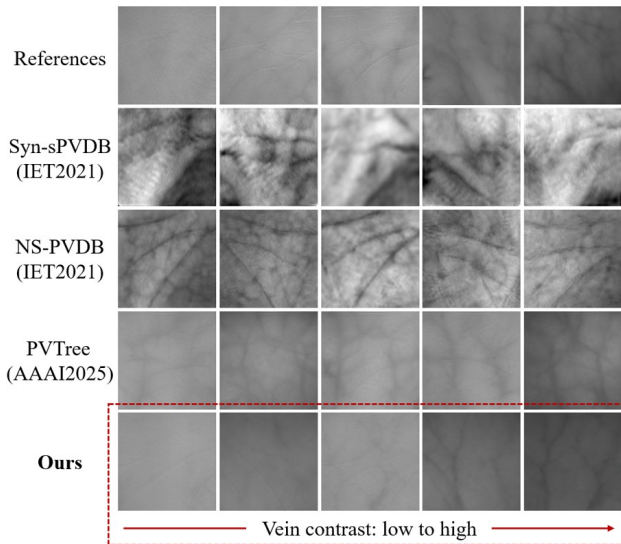


Figure 6: A visual comparison between existing generative methods and our method is presented. The red box highlights that our synthetic data exhibits clear identity-specific vein contrast, consistent with real-world data. The references are sourced from public datasets (Hao et al. 2008; Zhang et al. 2009; Shang et al. 2025).

protocol in PVTTree, we split each dataset into training and testing sets with a 1:1 ratio. For both the generative model and the recognition model, we use the training sets from the four public datasets, comprising a total of 751 identities and 12,024 images. The testing sets are employed to evaluate the performance of the recognition models.

Datasets	#ID	Samples	Total images	Sessions interval
CASIA	200	6	1,200	30days
HFUT	202	24	4,848	7days
PolyU	500	12	6,000	9days
TongJi	600	20	12,000	61days

Table 2: Details of the palm vein datasets used in our experiments.

**Generation Model Training Setups** The generation model uses an open-source implementation of ID-DPM(Nichol and Dhariwal 2021). The resolution of  $128 \times 128$  pixels is a commonly adopted size in public palm vein datasets and widely used in biometric fields like face and palmprint. We maintain this convention for our synthetic data. Besides, we employ a cosine noise schedule and set the number of diffusion steps to 1,000, with a learning rate of  $2e - 4$ . All experiments are conducted on 8 GPUs with a batch size of 32.

**Recognition Model Training Setups** The recognition model employs ArcFace (Deng et al. 2019) with a scale factor  $s = 64$  and a margin  $m = 0.5$ , implemented on a

Datasets	Syn-sPVDB	NS-PVDB	PVTTree	Ours
FID	31.62	28.63	20.07	<b>18.50</b>
Wang17	0.48	0.60	0.63	<b>0.73</b>

Table 3: Scores of quality assessment on real and synthetic datasets.

ResNet50 (He et al. 2016) backbone. In the training process, we adopt the SGD optimizer with 0.9 momentum and  $5 \times 10^{-4}$  weight decay for 30 epochs. The learning rate is initialized to 0.1 and decayed by  $10 \times$  at epochs 20 and 25. All experiments are conducted on 4 GPUs with a batch size of 32.

### Assessment of Image Quality

To evaluate the quality of our synthetic images, we employ both visual comparison and quantitative assessment methods. Fig.6 presents a visual comparison between our synthetic samples and those generated by open-sourced palm vein generation methods, including Syn-sPVDB (Salazar et al. 2021b), NS-PVDB (Salazar et al. 2021a), PVTTree, as well as real-world data. It can be observed that our data not only exhibit high visual similarity to real-world data but also effectively simulate significant variation in vein contrast across different identities. It demonstrates the effectiveness of our proposed projection model. For quantitative assessment, we use the commonly employed image quality metrics: Fréchet inception distance (FID) and the vein image-specific quality measure Wang17 (Wang et al. 2017), where a lower FID and a higher Wang17 score indicate better image quality. As shown in Table 3, the experimental results demonstrate that our synthetic data achieves the lowest FID score, highlighting their distribution similarity to real-world data. Furthermore, the comparison of palm vein image quality scores using Wang17 further emphasizes the high quality of our synthetic data.

### Assessment of Recognition Performance

In downstream recognition tasks, we compare the existing generation methods with our method in terms of their impact on the accuracy of the recognition model. For palm vein generation methods, we select the open-source methods Syn-sPVDB, NS-PVDB, and PVTTree. Additionally, to provide a more comprehensive evaluation, we also compare advanced methods in the fields of palmprint and face generation, including PCE-Palm, IDiff-Face (Boutros et al. 2023), and Vec2Face (Wu et al. 2024). The comparative experiments among different methods use the same dataset size of 4,000 IDs with 7 samples per ID, consistent with the open-source synthetic palm vein datasets. The experimental results are presented in Table 4, where we use equal error rate (EER) and true acceptance rate at a  $1e - 4$  false acceptance rate (TAR@ $1e-4$ ) as evaluation metrics.

Based on the experimental results presented in Table 4, we observe that, compared to real-world data, existing generation methods are generally unable to effectively enhance

Datasets			CASIA		PolyU		HFUT		TongJi		Average	
Name	IDs	Imgs	EER	TAR@ 1e-4	EER	TAR@ 1e-4	EER	TAR@ 1e-4	EER	TAR@ 1e-4	EER	TAR@ 1e-4
Real data	751	12024	0.0259	0.8980	0.0008	0.9948	0.0049	0.9321	0.0130	0.9306	0.0112	0.9389
NS-PVDB	4000	28000	0.1297	0.4259	0.0250	0.8140	0.0381	0.7145	0.1718	0.2371	0.0912	0.5479
Syn-sPVDB	4000	28000	0.2466	0.2025	0.0199	0.8597	0.0836	0.4831	0.2794	0.0978	0.1574	0.4107
PCE-Palm	4000	28000	0.0456	0.6804	0.0120	0.8831	0.0143	0.8478	0.0264	0.8479	0.0246	0.8148
IDiff-Face	4000	28000	0.0402	0.6710	0.0040	0.9725	0.0182	0.8420	0.0444	0.6724	0.0267	0.7895
Vec2Face	4000	28000	0.0402	0.6710	0.0040	0.9725	0.0182	0.8420	0.0444	0.6724	0.0267	0.8653
PVTree	4000	28000	0.0241	0.8749	0.0040	0.9699	0.0100	0.9001	0.0117	0.9358	0.0125	0.9202
Ours	4000	28000	0.0194	0.8914	<b>0.0007</b>	<b>0.9984</b>	0.0044	0.9402	0.0115	0.9432	0.0090	0.9433
	1000	20000	0.0289	0.8547	0.0010	0.9928	0.0077	0.9390	0.0154	0.9202	0.0133	0.9267
	2000	40000	0.0239	0.8858	0.0009	0.9932	<b>0.0028</b>	<b>0.9526</b>	0.0144	0.9307	0.0105	0.9406
	3000	60000	0.0195	0.8930	0.0008	0.9971	0.0054	0.9432	0.0114	0.9366	0.0093	0.9425
	4000	80000	<b>0.0181</b>	<b>0.9031</b>	0.0009	0.9968	0.0060	0.9517	<b>0.0106</b>	<b>0.9459</b>	<b>0.0089</b>	<b>0.9494</b>
Mix <sup>1</sup>	4751	92024	<b>0.0151</b>	<b>0.9385</b>	<b>0.0006</b>	<b>0.9988</b>	<b>0.0017</b>	<b>0.9559</b>	<b>0.0073</b>	<b>0.9771</b>	<b>0.0062</b>	<b>0.9676</b>

<sup>1</sup> Mixing the real data with our 4,000 IDs dataset.

Table 4: Recognition performance on real datasets and synthetic datasets under different experimental settings.

Settings			TAR@1e-4				
L	V	S	CASIA	PolyU	HFUT	TongJi	Average
✗	✗	✗	0.8525	0.9726	0.8718	0.8846	0.8954
✗	✗	✓	0.8611	0.9836	0.8925	0.8949	0.9080
✗	✓	✗	0.8642	0.9840	0.8935	0.8963	0.9095
✓	✗	✗	0.8824	0.9962	0.9244	0.9285	0.9329
✓	✓	✓	<b>0.8914</b>	<b>0.9984</b>	<b>0.9402</b>	<b>0.9432</b>	<b>0.9433</b>

Table 5: Recognition performance on synthetic datasets with different generation components.

the performance of recognition models. In contrast, our synthetic data surpasses the recognition performance of existing methods with a small number of IDs. As the number of IDs increases, the recognition performance further improves, surpassing real data at a scale of 2,000 IDs and achieving optimal performance at a scale of 4,000 IDs. Additionally, we find that the performance of the recognition model achieves greater improvement when supported by a combination of synthetic and real data. This demonstrates that our method achieves a balance between identity consistency and intra-class diversity, consequently yielding superior performance.

### Ablation Study

In the ablation experiments, we evaluate the impact of different components in the generation process on recognition performance, including the LSAP model, the VPDM, and

the style library. For convenience, we denote these components as L, V, and S, respectively. Table 5 presents the experimental results for different component configurations. All experimental settings remain consistent with previous experiments, utilizing a synthetic dataset size of 4,000 IDs with 7 samples per ID. The experimental results show that the inclusion of each component has a positive effect on improving the recognition performance. Notably, our projection model significantly enhances the recognition performance compared to the original method.

## Conclusion

In this paper, we propose a novel palm vein generation method that is able to simulate identity-specific vein contrast. Using the projected vein patterns as identity conditions, we introduce a conditional diffusion model to generate high-fidelity palm vein images. In our method, we propose VPDM to extract palm vein patterns from real-world data, thereby providing paired data for the training of the diffusion model. The experimental results show that the image quality of palm vein generated by our method is obviously better than that generated by other existing generation methods. Furthermore, the recognition accuracy of the model trained on data generated by our method is obviously higher than that of the models trained on data generated by other existing generation methods, and even exceeds that of the model trained on real data at a scale of 2,000 IDs. This validates that our work significantly promotes the study of vein image generation. In future work, exploring how to better define biophysical priors and effectively incorporate them into deep models will be a valuable research direction.

## Acknowledgments

This work is partly supported by the grants of the National Natural Science Foundation of China under Nos.62476077, 62272142, and 62076086.

## References

- Abed, M. H.; Alsaeedi, A. H.; Alfoudi, A. D.; Otebolaku, A. M.; and Razooghi, Y. S. 2020. Palm vein identification based on hybrid features selection model. *arXiv preprint arXiv:2007.16195*.
- Boutros, F.; Grebe, J. H.; Kuijper, A.; and Damer, N. 2023. Idiff-face: Synthetic-based face recognition through fizzy identity-conditioned diffusion model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 19650–19661.
- Deng, J.; Guo, J.; Xue, N.; and Zafeiriou, S. 2019. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 4690–4699.
- Footner, E.; Firipis, K.; Liu, E.; Baker, C.; Foley, P.; Kapsa, R. M.; Pirogova, E.; O’Connell, C.; and Quigley, A. 2023. Layer-by-Layer Analysis of In Vitro Skin Models. *ACS Biomaterials Science & Engineering*, 9(11): 5933–5952.
- guide, P. 2020. Amazon Announces New Contactless Payment System Using Just Your Hand. <https://www.pcguides.com/news/amazon-announces-new-contactless-payment-system-using-just-your-hand/>. [Online; accessed: 2025-03-08].
- Hao, Y.; Sun, Z.; Tan, T.; and Ren, C. 2008. Multispectral palm image fusion for accurate contact-free palmprint recognition. In *2008 15th IEEE International Conference on Image Processing*, 281–284. IEEE.
- He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.
- Ho, J.; Jain, A.; and Abbeel, P. 2020. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33: 6840–6851.
- Htet, A. S. M.; and Lee, H. J. 2023. Contactless palm vein recognition based on attention-gated residual U-Net and ECA-ResNet. *Applied Sciences*, 13(11): 6363.
- Isola, P.; Zhu, J.-Y.; Zhou, T.; and Efros, A. A. 2017. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1125–1134.
- Jin, J.; Shen, L.; Zhang, R.; Zhao, C.; Jin, G.; Zhang, J.; Ding, S.; Zhao, Y.; and Jia, W. 2024. PCE-Palm: Palm Crease Energy Based Two-Stage Realistic Pseudo-Palmprint Generation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 2616–2624.
- Karras, T.; Laine, S.; Aittala, M.; Hellsten, J.; Lehtinen, J.; and Aila, T. 2020. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 8110–8119.
- Li, Y.; Lu, H.; Wang, Y.; Gao, R.; and Zhao, C. 2022. ViT-Cap: a novel vision transformer-based capsule network model for finger vein recognition. *Applied Sciences*, 12(20): 10364.
- Nichol, A. Q.; and Dhariwal, P. 2021. Improved denoising diffusion probabilistic models. In *International conference on machine learning*, 8162–8171. PMLR.
- Nunez, A. S. 2009. *A physical model of human skin and its application for search and rescue*. Air Force Institute of Technology.
- Ou, W.-F.; Po, L.-M.; Zhou, C.; Xian, P.-F.; and Xiong, J.-J. 2022. Gan-based inter-class sample generation for contrastive learning of vein image representations. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 4(2): 249–262.
- Qin, H.; Gong, C.; Li, Y.; Gao, X.; and El-Yacoubi, M. A. 2023. Label enhancement-based multiscale transformer for palm-vein recognition. *IEEE Transactions on Instrumentation and Measurement*, 72: 1–17.
- Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; and Ommer, B. 2022. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 10684–10695.
- Salazar, E.; Hernández-García, R.; Barrientos, R. J.; Vilches, K.; Mora, M.; and Vásquez, A. 2021a. Automatic Generation of Synthetic Palm Vein Images: a Nature-based Approach. In *11th International Conference of Pattern Recognition Systems (ICPRS 2021)*, volume 2021, 38–43.
- Salazar, E.; Hernández-García, R.; Barrientos, R. J.; Vilches, K.; Mora, M.; and Vásquez, A. 2021b. Generating Style-based Palm Vein Synthetic Images for the Creation of Large-Scale Datasets. In *11th International Conference of Pattern Recognition Systems (ICPRS 2021)*, volume 2021, 182–187.
- Salazar-Jurado, E. H.; Vilches-Ponce, K.; Hernández-García, R.; and Barrientos, R. J. 2024. Palm vein modeling for generating synthetic images with biometric purposes: a geometrical approach. *Computational and Applied Mathematics*, 43(3): 108.
- Shang, S.; Zhao, C.; Zhang, R.; Jin, J.; Zhang, J.; Guo, R.; Ding, S.; Wu, Y.; Zhao, Y.; and Jia, W. 2025. PVTree: Realistic and Controllable Palm Vein Generation for Recognition Tasks. *arXiv, AAAI-25:2503.02547*.
- Sun, S.; Cong, X.; Zhang, P.; Sun, B.; and Guo, X. 2021. Palm vein recognition based on npe and kelm. *IEEE access*, 9: 71778–71783.
- Swinehart, D. F. 1962. The beer-lambert law. *Journal of chemical education*, 39(7): 333.
- Tencent. 2024. WeChat Palm Swipe Payment: Easily wave your hand to complete the payment. <https://www.tencent.com/zh-cn/articles/2201785.html>. [Online; accessed: 2025-03-08].
- Vyas, S.; Banerjee, A.; and Burlina, P. 2013. Estimating physiological skin parameters from hyperspectral signatures. *Journal of biomedical optics*, 18(5): 057008–057008.

- Wang, C.; Peng, M.; Xu, L.; and Chen, T. 2016. A single scale retinex based method for palm vein extraction. In *2016 IEEE Information Technology, Networking, Electronic and Automation Control Conference*, 75–78. IEEE.
- Wang, C.; Zeng, X.; Sun, X.; Dong, W.; and Zhu, Z. 2017. Quality assessment on near infrared palm vein image. In *2017 32nd Youth academic annual conference of Chinese association of automation (YAC)*, 1127–1130. IEEE.
- Wang, G.; Sun, C.; and Sowmya, A. 2019. Multi-weighted co-occurrence descriptor encoding for vein recognition. *IEEE Transactions on Information Forensics and Security*, 15: 375–390.
- Wu, H.; Singh, J.; Tian, S.; Zheng, L.; and Bowyer, K. W. 2024. Vec2Face: Scaling face dataset generation with loosely constrained vectors. *arXiv preprint arXiv:2409.02979*.
- Yang, S.; Wang, Y.; Li, H.; Meng, J.; Wu, Y.; Meng, X.; and Zhang, J. 2025. Hybrid Fourier score distillation for efficient one image to 3D object generation. *Visual Intelligence*, 3(1): 1–13.
- Zhang, D.; Guo, Z.; Lu, G.; Zhang, L.; and Zuo, W. 2009. An online system of multispectral palmprint verification. *IEEE transactions on instrumentation and measurement*, 59(2): 480–490.
- Zhang, L.; Cheng, Z.; Shen, Y.; and Wang, D. 2018. Palmprint and palmvein recognition based on DCNN and a new large-scale contactless palmvein dataset. *Symmetry*, 10(4): 78.
- Zhao, K.; Shen, L.; Zhang, Y.; Zhou, C.; Wang, T.; Zhang, R.; Ding, S.; Jia, W.; and Shen, W. 2022. Bézierpalm: A free lunch for palmprint recognition. In *European Conference on Computer Vision*, 19–36. Springer.
- Zhu, J.-Y.; Park, T.; Isola, P.; and Efros, A. A. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, 2223–2232.