ASYMMETRIC GENERATIVE ADAPTATION FOR FEDERATED MULTISPECTRAL PALMPRINT RECOGNITION UNDER SPECTRAL HETEROGENEITY

Anonymous authors
Paper under double-blind review

ABSTRACT

Multispectral palmprint recognition captures both surface ridge textures and subcutaneous vein patterns across multiple spectral bands, offering improved robustness to presentation attacks in high-security authentication scenarios. However, deploying such systems on distributed consumer-grade devices via Federated Learning (FL) introduces a critical practical barrier: *Spectral Heterogeneity*. Variations in local sensor hardware and ambient illumination produce severe Non-IID spectral distributions across clients, causing conventional personalized FL (pFL) to overfit on scarce local samples. To address these challenges, we present **AsyGA** (Asymmetric Generative Adaptation), a privacy-first pFL framework designed for heterogeneous deployment on consumer-grade devices. Each client trains a lightweight diffusion model entirely on-device to capture its local spectral profile, while only the downstream multispectral classifier is shared with the server for communication-efficient aggregation. After deployment, clients leverage these local generators to synthesize single-domain surrogate images for on-device personalization, eliminating raw biometric data exposure while minimizing communication overhead. To counteract *Asymmetric Modality Collapse*, a phenomenon where biased single-domain gradients overwrite representations of unseen bands during local optimization, we implement a hardware-friendly **Fidelity-Preserving Structural Constraint** consisting of four progressive mechanisms: (1) supervised contrastive optimization on synthetic identities for robust open-set separation; (2) feature-anchored cosine distillation against a frozen global teacher to stabilize the student embedding space; (3) batch normalization statistics locking to suppress synthetic distribution shift; and (4) targeted convolutional backbone freezing to shield globally learned multi-band filters from biased gradients while truncating backpropagation to slash on-device memory overhead. Evaluated under a strict cross-dataset open-set protocol spanning 287 million verification pairs across four spectral domains, AsyGA reduces the Equal Error Rate by over 50% against state-of-the-art FL baselines and more than doubles the True Acceptance Rate under a strict FAR $= 10^{-6}$ threshold. Furthermore, the on-device personalization stage adds only 44.6% of the FedAvg training time (61.5% including one-time generator training) and executes seamlessly on a consumer-grade laptop GPU without specialized infrastructure.

1 INTRODUCTION

Multispectral palmprint recognition has emerged as a highly secure biometric modality (Jain et al., 2004; Gao et al., 2026; Zhao et al., 2025b). Unlike standard RGB imaging, multispectral systems capture multiple wavelengths (e.g., Red, Green, Blue, and Near-Infrared/NIR) (Zhang et al., 2010; Hao et al., 2008) to jointly encode surface friction ridges and subcutaneous palmar vein networks (Wu et al., 2021; Kumar & Shekhar, 2011), demonstrating enhanced defense against presentation attacks (Liu et al., 2025). However, the sensitive nature of anatomical data demands stringent privacy protection, precluding centralized data collection.

Federated Learning (FL) (McMahan et al., 2017; Wu et al., 2022) enables collaborative model training without transmitting raw biometric data. Yet applying FL to multispectral palmprints

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

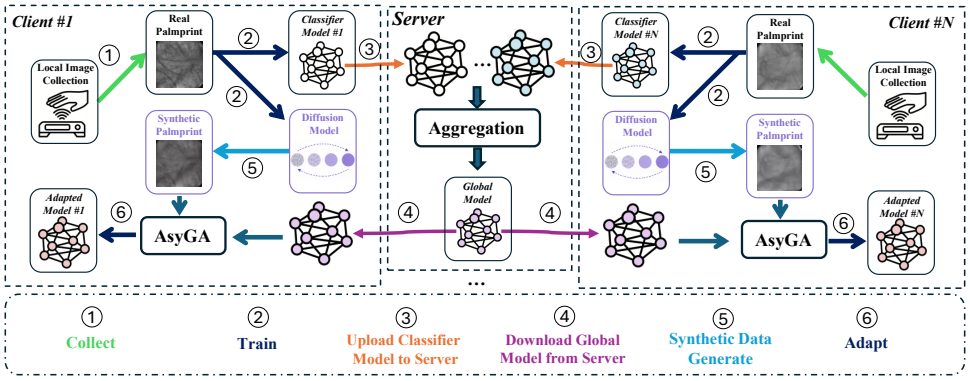


Figure 1: Overview of the proposed **AsyGA** framework for privacy-preserving personalized palmprint recognition across six collaborative execution stages.

exposes a fundamental bottleneck: *Spectral Heterogeneity (Spectral Non-IID)*. Hardware sensor discrepancies and ambient illumination across distributed consumer terminals induce profound multi-band distribution shifts, causing severe client drift and elevated Equal Error Rates (EER). Personalized FL (pFL) (Arivazhagan et al., 2019; Ma et al., 2022) mitigates drift via local fine-tuning, while generative strategies synthesize client-specific samples for Non-IID relief (Zhao et al., 2025a), including realistic palmprint diffusion (Jin et al., 2025). Nonetheless, such approaches cannot be straightforwardly applied to multispectral settings.

Multispectral models rely on joint correlations between visible textures and subcutaneous NIR patterns. Fine-tuning with synthetic data from only one observed domain introduces gradient interference that overwrites representations of unobserved modalities, collapsing cross-spectral verification. Because consumer-grade sensing hardware captures only its local spectral band, fully aligned multispectral synthetic generation remains infeasible on-device. This formalizes the **Asymmetric-Modality** personalization problem: *how can a four-domain global model be adapted using single-domain synthetic data while preserving representations of all unseen modalities?*

To resolve this dilemma, we propose **AsyGA** (**Asy**mmetric **G**enerative **A**daptation), a decoupled pFL framework: lightweight diffusion models operate purely on local consumer-grade devices to capture private spectral distributions, while only the multispectral classifier participates in federated aggregation. After aggregation, clients synthesize single-domain samples for personalized adaptation. To prevent modality collapse, AsyGA integrates a **Fidelity-Preserving Structural Constraint** with four progressive mechanisms: supervised contrastive optimization for open-set discriminative boundaries; feature-anchored cosine distillation to stabilize student embeddings against a global teacher; batch normalization locking to curb synthetic distribution drift; and targeted backbone freezing to isolate globally learned multi-band filters from biased gradients.

The main contributions of this paper are:

- **Asymmetric-Modality Problem Formalization.** We identify and formally characterize the modality collapse induced by single-domain generative fine-tuning in multispectral FL, establishing Asymmetric-Modality personalization as a vital challenge for heterogeneous consumer-grade biometrics.
- **AsyGA Framework.** We present AsyGA, a decoupled generative pFL framework where local diffusion models remain strictly private while the classifier is collaboratively aggregated, achieving zero-trust personalization without exposing raw data or generative weights.
- **Fidelity-Preserving Structural Constraint.** We introduce a four-mechanism constraint (supervised contrastive optimization, feature-anchored distillation, BN statistics locking, and backbone freezing) that preserves pretrained cross-spectral knowledge during asymmetric adaptation.
- **Performance and Consumer-Grade Practicality.** Cross-dataset evaluations demonstrate that AsyGA cuts EER by over 50% relative to SOTA FL baselines and more than doubles TAR under FAR $= 10^{-6}$, with profiling confirming feasibility on commodity laptop GPUs.

2 Related Work

Multispectral Palmprint Recognition. Palmprint recognition is widely favored in biometric authentication due to rich line textures and high discriminability (Zhang et al., 2003; Fei et al., 2019; Su et al., 2024). Deep architectures substantially enhance spatial feature extraction for palmprint verification (Yang et al., 2023a;b). To counter sophisticated presentation attacks (Liu et al., 2025), multispectral palmprints (Zhang et al., 2010; Hao et al., 2008) fuse surface friction ridges with subcutaneous palmar vein structures (Wu et al., 2021). However, centralizing such sensitive anatomical data explicitly breaches privacy regulations, motivating decentralized paradigms.

Federated Learning in Biometrics. FL (McMahan et al., 2017) enables privacy-preserving collaborative model training by sharing parameter updates instead of raw data (Lian et al., 2023). For multispectral palmprints, PSFed-Palm (Yang et al., 2024) introduces collaborative cross-spectral alignment, and DPFed-Palm (Li et al., 2025) advances this paradigm. FedPalm (Yang et al., 2026) leverages dual global-local expert modules for verification. Despite these advances, standard FL remains vulnerable to severe Non-IID domain shifts. Conventional pFL methods (Arivazhagan et al., 2019) fine-tune locally but require continuous access to real on-device data, triggering overfitting and privacy risks under local data scarcity.

Generative Personalization in FL. Generative models can synthesize client-specific samples to mitigate Non-IID distributions in FL (Zhao et al., 2025a). Diff-Palm (Jin et al., 2025) demonstrates high-fidelity palmprint synthesis via diffusion models. However, extending this to multispectral settings faces a barrier: resource constraints on consumer-grade devices make multi-band joint synthesis prohibitive, restricting generators to single-spectrum outputs. Naïvely fine-tuning with asymmetric synthetic samples triggers destructive gradient interference that overrides globally learned cross-spectral correlations—a failure mode we term *Asymmetric Modality Collapse*.

3 Proposed Methodology

As illustrated in Fig. 1, AsyGA achieves multispectral personalized FL via two decoupled phases: (1) global cross-spectral collaborative pre-training on decentralized real biometrics, followed by (2) localized on-device personalization driven entirely by private single-spectrum synthesis.

3.1 Phase I: Global Cross-Spectral FL & Decoupled Privacy

Let $\mathcal{C} = \{1, \ldots, N\}$ denote N distributed clients. Each client $k \in \mathcal{C}$ possesses a biometric sensor capturing single-band palmprints $X^k \subset \mathcal{X}^{m_k}$, where $m_k \in \mathcal{M}_{\text{spec}} = \{\text{NIR}, \text{Red}, \text{Green}, \text{Blue}\}$ is the client's spectral modality. Under zero-trust constraints, raw images X^k never leave local storage. Clients optimize local classifiers on their private spectra and transmit parameter updates to the central server for iterative federated aggregation (e.g., FedAvg). Global FL is indispensable because: (1) single-spectrum consumer-grade generators cannot capture subcutaneous NIR venous structures; (2) server aggregation synthesizes population and hardware diversity across clients; and (3) local raw data is permanently sealed offline after local generative training, ensuring an uncompromising zero-trust boundary.

3.2 Asymmetric Modality Collapse and the Four AsyGA Constraints

Let W_G denote the globally aggregated parameters from Phase I. Following deployment, client k leverages its local lightweight diffusion model G^k_{diff} to synthesize private single-spectrum surrogates $\mathcal{D}^k_{\text{syn}} = \{(x_j, y_j)\}_{j=1}^{N_{\text{syn}}}$, where $x_j \in \mathcal{X}^{m_k}$ and $y_j \in \{1, \ldots, K\}$ is the identity label. Naïvely fine-tuning W_G solely on $\mathcal{D}^k_{\text{syn}}$ induces biased local gradients $\nabla_W L_{\text{local}}$ that overwrite unobserved spectral feature bases. As formalized in Appendix B, cross-spectral alignment $S(W)$ decays at least linearly: $\mathbb{E}[S(W_{T_{\text{ad}}})] \leq S(W_0) - T_{\text{ad}}\eta\delta/2$, i.e., *Asymmetric Modality Collapse*.

To resolve this, we partition the network as $W = \{w_{\text{conv}}, w_{\text{proj}}, w_{\text{cls}}\}$: the convolutional backbone, projection layers outputting embedding $f(x; W)$, and the ArcFace classification head. Under hard freezing, only $\theta = \{w_{\text{proj}}, w_{\text{cls}}\}$ is updated. Four structural constraints are introduced (Fig. 2, Algorithm 1).

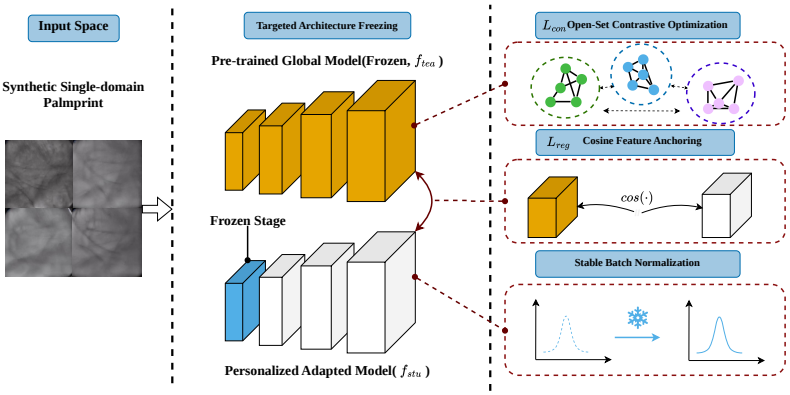


Figure 2: Workflow of the AsyGA personalization module driven by single-domain synthetic data via four progressive structural constraints.

3.2.1 Constraint 1: Open-Set Semantic and Contrastive Optimization

Let $\mathcal{B} = \{(x_j, y_j)\}_{j=1}^{B} \subset \mathcal{D}_{\text{syn}}^{k}$ denote a mini-batch. Let $f_{\text{stu}}(\cdot) \equiv f(\cdot; W_{\text{stu}}) : \mathcal{X}^{m_k} \to \mathbb{R}^d$ be the student feature extractor, and $\mathcal{H}(\cdot; w_{\text{head}}) : \mathbb{R}^d \to \mathbb{R}^K$ the classification head yielding logits $h_j = \mathcal{H}(f_{\text{stu}}(x_j); w_{\text{head}})$. Identity classification loss is computed via categorical cross-entropy:

$$L_{\text{ce}} = -\frac{1}{B}\sum_{j=1}^{B} \log \frac{\exp(h_{j,y_j})}{\sum_{c=1}^{K} \exp(h_{j,c})}. \tag{1}$$

Because open-set biometric verification evaluates unseen identities by feature metric distance rather than closed-set logits, L_{ce} alone cannot prevent radial scattering. We thus draw two augmented views of each synthetic image, obtaining $2B$ normalized embeddings $z_i = f_{\text{stu}}(\tilde{x}_i)/\|f_{\text{stu}}(\tilde{x}_i)\|_2 \in \mathbb{S}^{d-1}$ with labels $\tilde{y}_i$, and optimize via Supervised Contrastive Loss:

$$L_{\text{con}} = \frac{1}{2B}\sum_{i=1}^{2B} \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{\exp(z_i^\top z_p/\tau)}{\sum_{a \in A(i)} \exp(z_i^\top z_a/\tau)}, \tag{2}$$

where $\tau > 0$ is temperature, $A(i) \equiv \{1, \ldots, 2B\} \setminus \{i\}$, and $P(i) \equiv \{p \in A(i) : \tilde{y}_p = \tilde{y}_i\}$.

3.2.2 Constraint 2: Feature-Anchored Regularization via Knowledge Distillation

To prevent single-domain gradients from distorting the globally aligned metric space, a static replica of W_G is frozen as a teacher anchor $f_{\text{tea}}(\cdot) \equiv f(\cdot; W_G)$. Using the first augmented view $z_j = f_{\text{stu}}(x_j^{(1)})/\|f_{\text{stu}}(x_j^{(1)})\|_2$ and $z_{\text{tea}}(x_j) = f_{\text{tea}}(x_j^{(1)})/\|f_{\text{tea}}(x_j^{(1)})\|_2$, we enforce cosine distillation:

$$L_{\text{reg}} = 1 - \frac{1}{B}\sum_{j=1}^{B} z_j^\top z_{\text{tea}}(x_j) = 1 - \frac{1}{B}\sum_{j=1}^{B} \frac{f_{\text{stu}}(x_j)^\top f_{\text{tea}}(x_j)}{\|f_{\text{stu}}(x_j)\|_2 \|f_{\text{tea}}(x_j)\|_2}. \tag{3}$$

The total objective is formulated as:

$$L_{\text{total}} = L_{\text{ce}} + \lambda L_{\text{con}} + \gamma L_{\text{reg}}, \tag{4}$$

where $\lambda \geq 0$ controls hyperspherical clustering and $\gamma > 0$ penalizes deviation from the global manifold.

3.2.3 Constraint 3: Anti-Drift Batch Normalization Stabilization

Calculating mini-batch statistics $\mu_{\mathcal{B}}$ and $\sigma_{\mathcal{B}}^2$ over single-band synthetic inputs corrupts global multi-band running statistics for unseen bands. To eliminate covariate shift, all BN layers are fixed in `eval()` mode:

$$\mu \leftarrow \mu_{\text{global}}, \qquad \sigma^2 \leftarrow \sigma^2_{\text{global}}. \tag{5}$$

This operation is independent of hard freezing: BN buffers are preserved against updates even when backbone weights are being trained. When the backbone is not hard-frozen, its parameters use $\eta_{\text{base}} = 10^{-6}$, while θ uses $\eta_{\text{head}} = 10^{-4}$.

Algorithm 1 Fidelity-Preserving Adaptation on Consumer-Grade Client

Require: Global model W_G, generator G_{diff}, epochs E, batch size B, budget N_{syn}, weights λ, γ, temperature τ, rates η_{head} (and η_{base} if backbone is unfrozen).

Offline Generative Sampling: Sample $\mathcal{D}_{\text{syn}} = \{(x_i, y_i)\}_{i=1}^{N_{\text{syn}}} \sim G_{\text{diff}}()$
Architectural Initialization & Freezing:
Initialize student $W_{\text{stu}} \leftarrow W_G$; instantiate frozen teacher $W_{\text{tea}} \leftarrow W_G$
Freeze backbone w_{conv} in W_{stu} ($\Delta w_{\text{conv}} \leftarrow 0$); set trainable parameters $\theta = \{w_{\text{proj}}, w_{\text{cls}}\}$
Lock all Batch Normalization layers in W_{stu} to `eval()` mode
On-Device Fine-Tuning:
for epoch $e = 1$ **to** E **do**
  for mini-batch $\mathcal{B} = \{(x_j, y_j)\}_{j=1}^{B} \subset \mathcal{D}_{\text{syn}}$ **do**
    Generate two augmented views $x_j^{(1)}, x_j^{(2)}$ for each x_j
    $f_{\text{stu}}^{(v)} \leftarrow f(x_j^{(v)}; W_{\text{stu}})$, $v \in \{1, 2\}$; $\quad f_{\text{tea}} \leftarrow f(x_j^{(1)}; W_{\text{tea}})$
    $z_j^{(v)} \leftarrow f_{\text{stu}}^{(v)} / \|f_{\text{stu}}^{(v)}\|_2$; $\quad z_{\text{tea}} \leftarrow f_{\text{tea}} / \|f_{\text{tea}}\|_2$
    Compute ArcFace logits $h_j \leftarrow \mathcal{H}(f_{\text{stu}}^{(1)}; w_{\text{cls}}, y_j)$
    Evaluate L_{ce} on view 1 via equation 1, L_{con} on both views via equation 2, and L_{reg} via equation 3
    $L_{\text{total}} \leftarrow L_{\text{ce}} + \lambda L_{\text{con}} + \gamma L_{\text{reg}}$
    Update θ via $\nabla_\theta L_{\text{total}}$ using Adam
  end for
end for

Ensure: Personalized client model W_{stu}

Table 1: The Privacy-Utility Dilemma and Generative Utilization Strategies on Cross-Spectral EER.

(a) Privacy-Utility Dilemma

Training Paradigm	Privacy	Avg EER	NIR-Blue
(1) Isolated Local	Preserved	27.42%	48.05%
(2) Centralized Joint	Violated	8.95%	24.81%

(b) Synthetic Data Utilization Strategies

Strategy	Training Pipeline	Avg EER	NIR-Blue
(0) Global Base	Real only	9.20%	24.33%
(1) Aug. FL	Real + Synth	7.17%	21.27%
(2) **AsyGA**	FL on Real, FT Synth	**3.97%**	**12.63%**

3.2.4 Constraint 4: Targeted Architectural Hard-Freezing

To eliminate gradient interference on spatial filters, we freeze the entire convolutional backbone w_{conv} ($\Delta w_{\text{conv}} = 0$) and restrict optimization to projection and classification parameters $\theta = \{w_{\text{proj}}, w_{\text{cls}}\}$:

$$\theta^{(t+1)} \leftarrow \theta^{(t)} - \eta_{\text{head}} \cdot \text{Adam}\big(\nabla_\theta L_{\text{total}}\big). \tag{6}$$

This isolates globally learned multi-band spatial filters from single-domain gradient distortion and eliminates backbone backward passes, avoiding the need to retain spatial activations for backpropagation.

4 Experiments

4.1 Experimental Setup

Datasets and Strict Cross-Dataset Protocol. We partition the CASIA Multispectral Database (Hao et al., 2008) (Red, Green, Blue, NIR; $6{\times}100{\times}4 = 2{,}400$ images) across clients under a dual Non-IID setup (disjoint identities and disjoint spectral domains per client). Open-set biometric verification is evaluated exclusively on the unseen PolyU Database (Zhang et al., 2010) ($12{\times}500{\times}4 = 24{,}000$ images), yielding 287,988,000 verification pairs (432,000 genuine / 287,424,000 impostor) without capture shortcuts.

Implementation Details. CCNet (Yang et al., 2023a) is the default backbone. Local diffusion models synthesize 2,400 single-spectrum surrogates ($N_{\text{syn}} = 2{,}400$). Optimization uses Adam ($\text{lr}_{\text{head}} = 10^{-4}$, $\text{lr}_{\text{trans}} = 10^{-6}$), $\lambda = 0.5$, $\gamma = 50$, $\tau = 0.07$, $B = 32$, and $E = 10$ local epochs across $T = 50$ FL rounds. Results are averaged over three independent runs (std. $< 0.05\%$).

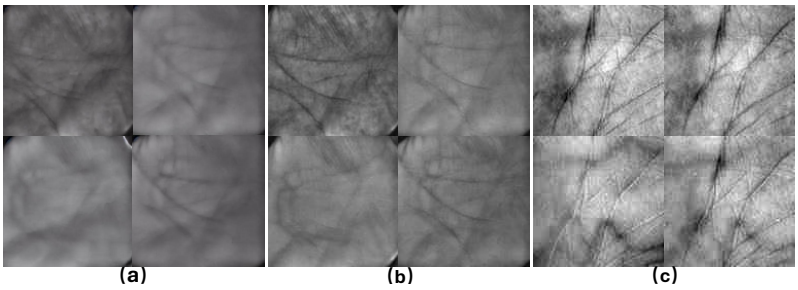


Figure 3: Visual comparison of palmprint samples: (a) synthetic single-spectrum palmprints from client diffusion models; (b) real multi-spectral palmprints from CASIA (Hao et al., 2008); (c) unseen PolyU dataset (Zhang et al., 2010) for zero-trust evaluation.

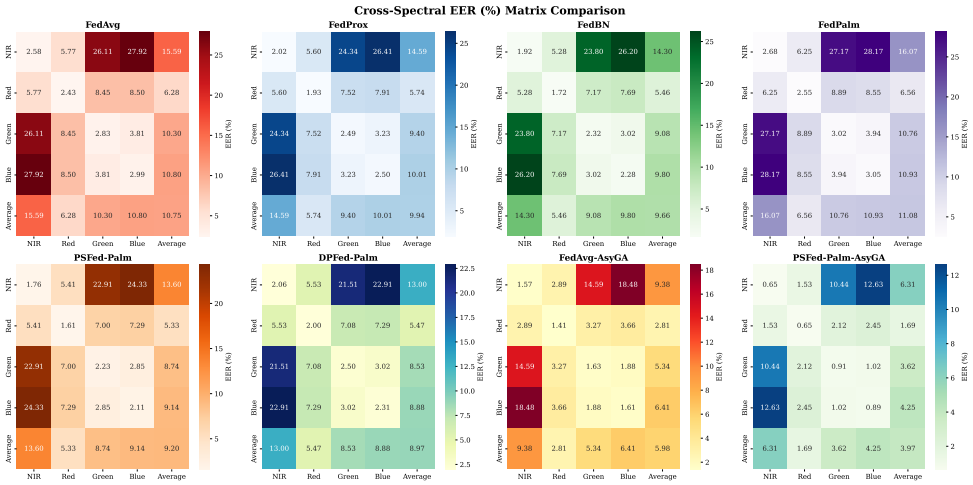


Figure 4: Cross-Spectral EER Heatmap Comparison between Baselines and AsyGA. Lower is superior.

4.2 Motivation: The Privacy-Utility Dilemma

Table 1a contrasts two paradigms: isolated local training thoroughly protects privacy but suffers from severe spectral shifts, leaving models blind to unseen spectra during evaluation and pushing NIR-vs.-Blue EER to 48.05%. While advanced FL baselines like DPFed-Palm approach centralized accuracy (8.97% vs. 8.95% EER), standard FL baselines plateau between 8.97% and 11.08% (Fig. 4). AsyGA operates strictly within privacy-preserving boundaries while utilizing generative personalization to break these global limits, achieving 3.97% EER.

4.3 Macro Comparison Across FL Baselines

We integrate AsyGA into representative FL aggregation schemes (Fig. 4). Conventional FL yields high average EERs from 8.97% to 11.08% due to cross-domain spectral shifts. Despite using only single-domain synthetic data locally, AsyGA consistently boosts robustness: DPFed-Palm (Li et al., 2025) with AsyGA drops EER from 8.97% to 6.62%; FedAvg-AsyGA achieves 5.98% (33.3% relative gain); and PSFed-Palm-AsyGA achieves 3.97% (a 55.7% relative error reduction).

4.4 Generative Augmentation vs. AsyGA

Mixing synthetic samples directly into FL batches (Table 1b) reduces EER from 9.20% to 7.17%, but cross-spectral performance remains degraded (21.27% NIR–Blue EER) because single-spectrum batches disrupt multi-band feature extraction. AsyGA decouples these phases: Phase I learns cross-spectral topologies from decentralized real data, and Phase II adapts locally via synthetic samples under distillation anchors, yielding 3.97% average EER and 12.63% NIR–Blue EER.

Table 2: Progressive ablation of AsyGA components on cross-spectral EER.

Fine-Tuning Strategy	Overall Avg EER	NIR vs. Blue EER
(0) Global Baseline	9.20%	24.33%
(1) Traditional FT (All Off)	12.10%	27.59%
(2) + Supervised Contrastive (L_{con})	8.72%	22.51%
(3) + Teacher Anchor ($\gamma = 50$)	4.47%	14.13%
(4) + BN Statistics Freezing	4.15%	13.55%
(5) + Hard Freezing (All On, **AsyGA**)	**3.97%**	**12.63%**

Table 3: Performance and Resource Overhead Comparison across Execution Tasks on an RTX 4060 Laptop GPU.

Task Configuration	Total Images	Time / 100 Img.	Throughput	Total Time	Primary Hardware Bottleneck (Ratio)
Diffusion Gen. (Inference)	600	61.20 s	1.64 img/s	∼6.1 min	UNet Reverse Sampling (71.0%)
Diffusion Train (Client 1)	1,000	21.63 s	4.62 img/s	∼3.6 min	UNet Forward & Backward (93.7%)
FedAvg (50 Aggregations)	307,200	**0.42 s**	**240.40 img/s**	∼21.3 min	Backward & Optimizer Update (59.5%)
Fine-tune (Hard Freezing ON)	12,800	1.59 s	63.00 img/s	**∼3.4 min**	Dual-path Forward (53.1%) / I/O (42.6%)
Fine-tune (Hard Freezing OFF)	12,800	2.02 s	49.50 img/s	∼4.3 min	Forward (41.5%) / I/O (39.5%) / Backward (19.0%)

4.5 Micro Deep Dive on Cross-Spectral Fidelity

Fig. 4 reveals how synthetic fine-tuning enhances multispectral verification without modality collapse. Intra-spectrally, diagonal EERs are compressed, validating adaptation to local sensors. Cross-spectrally, anchor-based distillation aligns representations across dissimilar bands capturing distinct physiological structures. Compared to the global baseline (9.20% / 24.33%) and DPFed-Palm (8.97% / 22.91%), PSFed-Palm-AsyGA cuts average EER to **3.97%** and NIR-vs.-Blue EER to **12.63%** (a **55.7%** and **44.8%** relative drop, respectively).

4.6 Ablation Study and Synergistic Mechanics

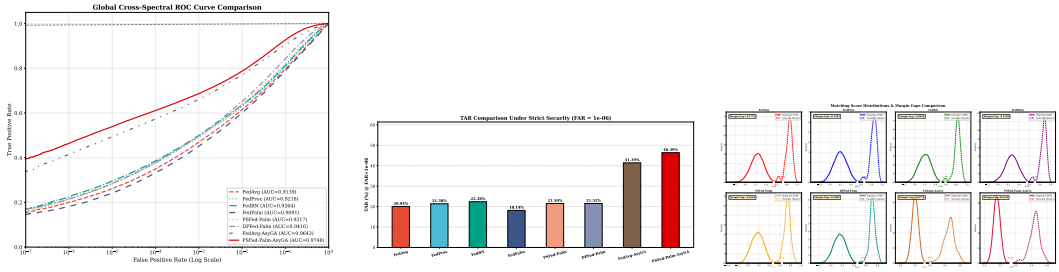
Table 2 analyzes the contribution of each design component. Naive fine-tuning (Strategy 1) on synthetic data degrades multi-band features, driving overall and NIR–Blue EERs to 12.10% and 27.59%. Progressively integrating our constraints systematically counters this collapse: (1) L_{con} (Strategy 2) recovers EER to 8.72% via metric structuring; (2) teacher distillation ($\gamma = 50$, Strategy 3) slashes EER to 4.47% by stabilizing cross-spectral manifolds; (3) locking BN statistics (Strategy 4) mitigates synthetic shift, reaching 4.15%; and (4) Targeted Hard-Freezing (Strategy 5) isolates multi-band filters, achieving optimal EERs of 3.97% (overall) and 12.63% (NIR–Blue). Additional leave-one-out ablations in Appendix A corroborate these complementary roles.

4.7 Resource Efficiency and Computational Overhead Analysis

Table 3 benchmarks execution stages on an NVIDIA RTX 4060 Laptop GPU. Relative to 21.3 min of FedAvg training, personalization (diffusion inference ∼6.1 min + adaptation ∼3.4 min) adds 44.6% overhead, or 61.5% including one-time generator training (∼3.6 min). Targeted hard-freezing boosts adaptation throughput from 49.50 to 63.00 img/s by eliminating backbone backward passes. Runtime is dominated by forward inference (53.1%) and I/O (42.6%), reducing per-step complexity to $\mathcal{O}(P_{\text{head}})$ and activation memory to $\mathcal{O}(1)$ relative to spatial layers.

4.8 Verification Limits and Open-Set Robustness

Fig. 5a displays log-scale ROC curves: AsyGA reaches an AUC of **0.9748** on PSFed-Palm. Under strict industrial security thresholds (FAR $= 10^{-6}$, Fig. 5b), baselines yield 18%–22% TAR, whereas AsyGA more than doubles TAR to 46.39% (vs. 21.50% for PSFed-Palm and 18.14% for FedPalm). Matching score distributions in Fig. 5c explain this gain: while DPFed-Palm exhibits substantial genuine-impostor overlap (margin gap of –0.1094), PSFed-Palm-AsyGA achieves a net positive margin gap (+0.0138), marking an absolute separation recovery of +0.1232 and validating superior cross-spectral discrimination.



(a) Log-scale ROC curves (b) TAR at FAR $= 10^{-6}$ (c) Matching score distributions

Figure 5: Open-set biometric verification limits and margin analysis: (a) Log-scale ROC curves showing AsyGA achieving AUC of 0.9748; (b) True Acceptance Rate under extreme security constraints (FAR $= 10^{-6}$); (c) Cosine similarity distributions indicating resolved feature overlap.

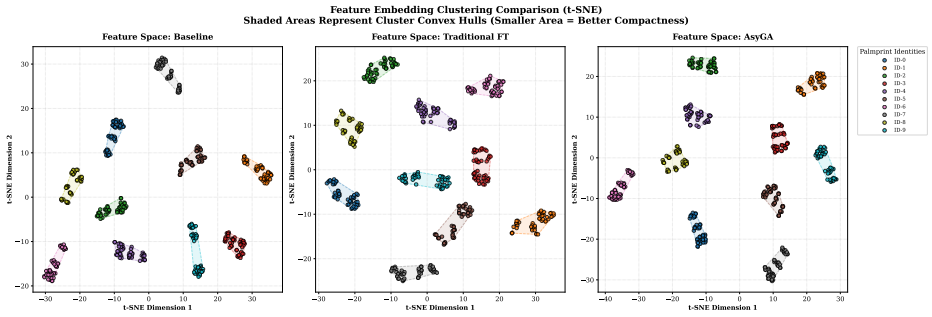


Figure 6: t-SNE visualization on PolyU (Zhang et al., 2010). (a) Baseline: loosely dispersed clusters; (b) Traditional FT: severe topological distortion and overlap; (c) AsyGA: highly compact, separable identity islands.

4.9 Qualitative Latent Space Visualization via t-SNE

Fig. 6 presents t-SNE visualizations (van der Maaten & Hinton, 2008) of ten identities on PolyU. Clusters are loosely scattered under the Global Baseline (Fig. 6a), while Traditional Fine-Tuning (Fig. 6b) triggers severe modality collapse and overlapping hulls. In contrast, AsyGA (Fig. 6c) forms compact, well-separated identity clusters, demonstrating that structural constraints preserve cross-spectral manifolds while enforcing wide-margin separation.

4.10 Architectural Generalization and Hyperparameter Sensitivity

Table 4 (Left) evaluates closed-set identification: AsyGA maintains 94.16% Rank-1 accuracy, while generative mixing (84.91%) and naive fine-tuning (86.55%) degrade severely. Fig. 7 shows that $\gamma = 50$ provides an optimal stability-plasticity balance across $[50, 80]$. Furthermore, AsyGA generalizes across backbones (Table 4, Right), lowering EER to 4.99% on CO3Net (Yang et al., 2023b) and 6.96% on ViT. AsyGA yields larger improvements on CNNs than ViTs because ViTs lack localized spatial inductive biases; fine-tuning on single-domain tokens modifies global self-attention across spectral channels, making localized structural alignment less targeted than in convolutional architectures.

5 Conclusion

This paper presented AsyGA, a pFL framework for multispectral palmprint recognition under severe spectral heterogeneity. By decoupling global cross-spectral representation learning from local diffusion-based distribution modeling, AsyGA achieves robust client personalization without post-FL access to real biometric data. We characterized *Asymmetric Modality Collapse*, where unconstrained local adaptation using single-domain synthetic data degrades representations of unseen spectral domains. To resolve this, AsyGA implements a Fidelity-Preserving Structural Constraint anchoring pre-trained features through supervised contrastive learning, feature distillation, BN stabilization, and targeted backbone freezing. Evaluations across two multispectral benchmarks confirm that AsyGA

Table 4: Comprehensive Robustness Analysis. **Left:** Closed-set identification performance on PolyU. **Right:** Open-set verification across alternative architectures.

Strategy	Avg Rank-1
(0) Global Baseline	99.19%
(1) **AsyGA (Ours)**	**94.16%**
(2) Traditional FT	86.55%
(3) Augmented FL	84.91%

Backbone	Strategy	Avg EER (↓)
CO3Net (Yang et al., 2023b)	Global Base	9.60%
	AsyGA	**4.99%**
ViT	Global Base	7.39%
	AsyGA	**6.96%**

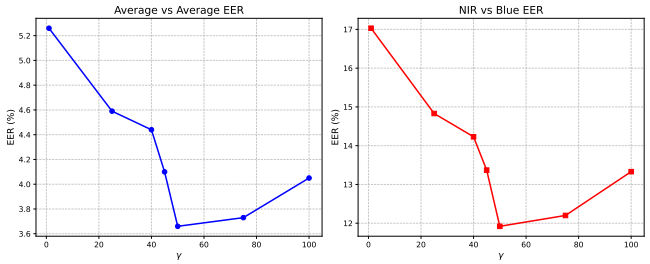


Figure 7: Sensitivity analysis of structural constraint weight γ, reaching optimum at $\gamma \approx 50$.

cuts EER by over 50% relative to state-of-the-art baselines while ensuring practical efficiency on consumer-grade hardware.

AI USE STATEMENT

Generative AI tools (e.g., ChatGPT, GitHub Copilot) were utilized strictly for language polishing, formatting refinement, and auxiliary code generation. Specifically, AI was used to improve the grammar, clarity, and academic tone of the manuscript, assist with LaTeX formatting and reference management, and aid in the implementation of boilerplate code, debugging, and scripting for experimental setup. The authors have rigorously reviewed, verified, and edited all AI-generated outputs. The authors take full responsibility for all content, scientific claims, experimental design, and implementation artifacts in this manuscript.

REPRODUCIBILITY STATEMENT

The paper details federated training, local synthesis, personalization constraints, evaluation protocols, datasets, and hardware profiling required to reproduce the reported results. Anonymous source code and configuration details will be provided with the submission.

REFERENCES

Manoj Ghuhan Arivazhagan, Vinay Aggarwal, Aaditya Kumar Singh, and Sunav Choudhary. Federated learning with personalization layers. *arXiv preprint arXiv:1912.00818*, 2019.

Lunke Fei, Guangming Lu, Wen Jia, Shaohua Teng, and David Zhang. Feature extraction methods for palmprint recognition: A survey and evaluation. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 49(2):346–363, 2019.

Chengrui Gao, Ziyuan Yang, Wei Jia, Lu Leng, Bob Zhang, and Andrew Beng Jin Teoh. Deep learning in palmprint recognition: A comprehensive survey. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 56(3):2143–2162, 2026. doi: 10.1109/TSMC.2025.3649416.

Ying Hao, Zhenan Sun, Tieniu Tan, and Chao Ren. Multispectral palm image fusion for accurate contact-free palmprint recognition. In *2008 15th IEEE International Conference on Image Processing*, pp. 281–284, 2008. doi: 10.1109/ICIP.2008.4711746.

Anil K. Jain, Arun Ross, and Salil Prabhakar. An introduction to biometric recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):4–20, 2004.

Jianlong Jin, Chenglong Zhao, Ruixin Zhang, Sheng Shang, Jianqing Xu, Jingyun Zhang, ShaoMing Wang, Yang Zhao, Shouhong Ding, Wei Jia, and Yunsheng Wu. Diff-palm: Realistic palmprint generation with polynomial creases and intra-class variation controllable diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 26367–26376, June 2025.

James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A. Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, Demis Hassabis, Claudia Clopath, Dharshan Kumaran, and Raia Hadsell. Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences*, 114(13):3521–3526, 2017.

A. Kumar and S. Shekhar. Personal identification using multibiometrics rank-level fusion. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 41(5):743–752, 2011.

Shuyi Li, Jianian Hu, Bob Zhang, Xin Ning, and Lifang Wu. Dynamic personalized federated learning for cross-spectral palmprint recognition. *IEEE Transactions on Image Processing*, 34: 4885–4895, 2025. doi: 10.1109/TIP.2025.3592508.

Zhizhong Li and Derek Hoiem. Learning without forgetting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(12):2935–2947, 2017.

F.-Z. Lian, J.-D. Huang, J.-X. Liu, G. Chen, J.-H. Zhao, and W.-X. Kang. FedFV: A personalized federated learning framework for finger vein authentication. *Machine Intelligence Research*, 20(5): 683–696, 2023.

C. Liu, H. Shao, and D. Zhong. Learning discriminative palmprint anti-spoofing features via high-frequency spoofing regions adaptation. *IET Image Processing*, 19(1):70029, 2025.

X. Ma, J. Zhang, S. Guo, and W. Xu. Layer-wised model aggregation for personalized federated learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10092–10101, 2022.

Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*, pp. 1273–1282. PMLR, 2017.

Lin Su, Lunke Fei, Bob Zhang, Shuping Zhao, Jie Wen, and Yong Xu. Complete region of interest for unconstrained palmprint recognition. *IEEE Transactions on Image Processing*, 33:3662–3675, 2024.

Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9:2579–2605, 11 2008.

C. Wu, F. Wu, L. Lyu, Y. Huang, and X. Xie. Communication-efficient federated learning via knowledge distillation. *Nature Communications*, 13(1):2032, 2022.

Ting Wu, Lu Leng, Muhammad Khurram Khan, and Faizan Ahmed Khan. Palmprint-palmvein fusion recognition based on deep hashing network. *IEEE Access*, 9:135816–135827, 2021.

Z. Yang, H. Huangfu, L. Leng, B. Zhang, A. B. J. Teoh, and Y. Zhang. Comprehensive competition mechanism in palmprint recognition. *IEEE Transactions on Information Forensics and Security*, 18:5160–5170, 2023a.

Ziyuan Yang, Wen Xia, Yu Qiao, Zetian Lu, Bob Zhang, Lu Leng, and Yi Zhang. Co3Net: Coordinate-aware contrastive competitive neural network for palmprint recognition. *IEEE Transactions on Instrumentation and Measurement*, 72:3276506, 2023b.

Ziyuan Yang, Andrew Beng Jin Teoh, Bob Zhang, Lu Leng, and Yi Zhang. Physics-driven spectrum-consistent federated learning for palmprint verification. *International Journal of Computer Vision*, 132(10):4253–4268, 2024.

Ziyuan Yang, Yingyu Chen, Chengrui Gao, Andrew Beng Jin Teoh, Bob Zhang, and Yi Zhang. Fedpalm: A general federated learning framework for closed- and open-set palmprint verification. *IEEE Transactions on Information Forensics and Security*, pp. 1–1, 2026. doi: 10.1109/TIFS.2026.3671064.

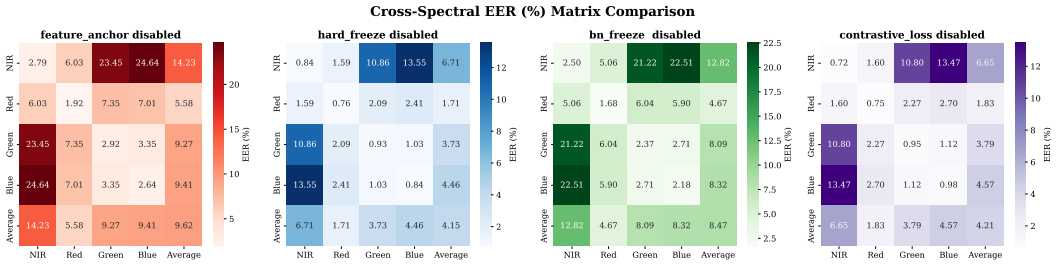
David Zhang, Wai-Kin Kong, Jane You, and Michael Wong. Online palmprint identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(9):1041–1050, 2003.

David Zhang, Zhenhua Guo, Guangming Lu, Lei Zhang, and Wangmeng Zuo. An online system of multispectral palmprint verification. *IEEE Transactions on Instrumentation and Measurement*, 59 (2):480–490, 2010. doi: 10.1109/TIM.2009.2028772.

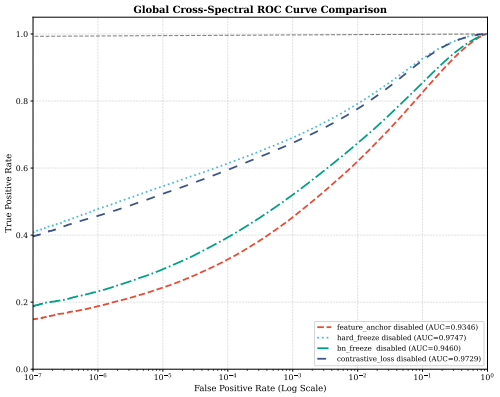
Peng Zhao, Shaocong Guo, Yanan Li, Shusen Yang, and Xuebin Ren. Fedgen: Personalized federated learning with data generation for enhanced model customization and class imbalance. *Future Generation Computer Systems*, 164:107595, 2025a. doi: https://doi.org/10.1016/j.future.2024.107595.

Shuping Zhao, Lunke Fei, Bob Zhang, Jie Wen, and J. Cui. Toward mobile palmprint recognition via multi-view hierarchical graph learning. *IEEE Transactions on Information Forensics and Security*, 20:101–113, 2025b.

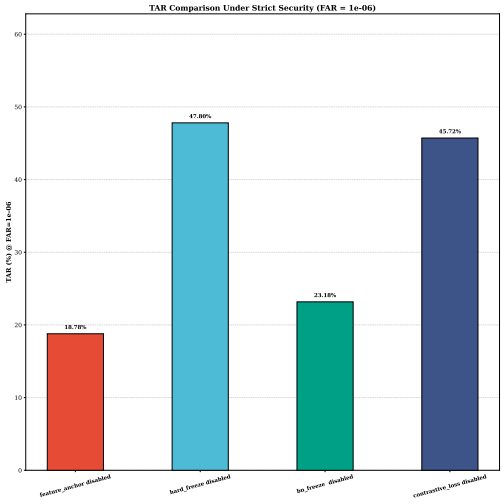
A LEAVE-ONE-OUT ABLATION: COMPONENT ISOLATION AND FAILURE MODE ANALYSIS



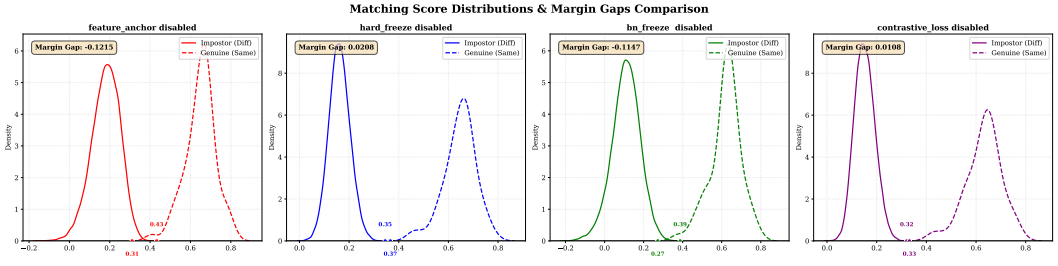
(a) Cross-spectral EER (%) matrix comparisons



(b) Log-scale global cross-spectral ROC curves



(c) TAR (%) under strict security threshold (FAR = 10^{-6})



(d) Score distributions and genuine-impostor margin gaps

Figure 8: Leave-one-out ablation analysis across four diagnostic perspectives: (a) Cross-spectral verification EER matrices; (b) Global ROC curves under varying security regimes; (c) Verification TAR at an industrial-grade strict threshold (FAR = 10^{-6}); (d) Matching score density distributions and decision margin gaps.

To systematically isolate the individual responsibility of each architectural mechanism in preventing *Asymmetric Modality Collapse*, we evaluate a leave-one-out ablation setting where exactly one component is disabled while all other three remain active:

- **feature_anchor disabled**: Removing the topological cosine distillation loss ($\gamma = 0$).
- **hard_freeze disabled**: Unfreezing convolutional feature backbones during personalization ($\Delta w_{\text{conv}} \neq 0$).
- **bn_freeze disabled**: Allowing running mean and variance of Batch Normalization layers to update on synthetic single-band data.
- **contrastive_loss disabled**: Removing the intra-class hyperspherical contrastive objective ($\lambda = 0$).

A.1 Quantitative & Qualitative Dissection

1. The Primary Pillar: Feature Anchoring. Disabling the feature anchor (`feature_anchor disabled`) inflicts the most devastating performance collapse across all four evaluation axes:

- **EER Explosion**: Overall EER surges from the intact AsyGA level ($\sim$3.97%) to **9.62%** (Fig. 8a). Crucially, cross-spectral pairs involving unseen spectrums collapse catastrophically (NIR–Blue EER deteriorates to **24.64%**, and NIR–Green to **23.45%**).
- **Extreme-Security Failure**: TAR at $\mathrm{FAR} = 10^{-6}$ plummets to an unacceptable **18.78%** (Fig. 8c), a loss of nearly 30% in absolute verification capability compared to intact configurations.
- **Severe Distribution Collapse**: As shown in Fig. 8d, the genuine-impostor score distributions exhibit catastrophic overlap, yielding a heavily negative margin gap of **–0.1215** (with the genuine distribution left tail extending to 0.43 against an impostor peak at 0.31).

This conclusively demonstrates that without the global feature anchor (L_{reg}), unconstrained single-modality adaptation rapidly drives the model into the *Asymmetric Modality Collapse* regime.

2. Channel Distribution Anchor: BN Statistics Freezing. Disabling Batch Normalization freezing (`bn_freeze disabled`) represents the second-most severe point of failure:

- When running statistics are updated on synthetic single-band (Blue-dominant) tokens, the global channel-wise mean and variance drift irreversibly away from NIR distributions.
- Consequently, overall average EER climbs to **8.47%** (Fig. 8a), with the NIR–Blue pair deteriorating to **22.51%**.
- Under extreme security ($\mathrm{FAR} = 10^{-6}$), the model completely loses discriminability, collapsing to **23.18%** TAR and a global AUC of only **0.9460** (Fig. 8b).
- The margin gap collapses to **–0.1147** (Fig. 8d), proving that batch statistics freezing is vital for preserving multi-band domain invariant representations against synthetic distribution shifts.

3. Metric Compaction: Contrastive Loss. Omitting the hyperspherical contrastive loss (`contrastive_loss disabled`):

- Overall EER remains relatively controlled at **4.21%**, and NIR–Blue EER stays at **13.47%**, because the backbone and global geometry remain anchored by L_{reg} and frozen BN.
- However, fine-grained class compactness is visibly compromised: TAR at $\mathrm{FAR} = 10^{-6}$ drops from 47.80% to **45.72%** (Fig. 8c), and the margin gap narrows by 48% from $+0.0208$ down to $+0.0108$ (Fig. 8d).
- This confirms that while L_{con} is not the primary line of defense against global catastrophic forgetting, it is strictly necessary for enforcing tight intra-class clustering and maximizing boundary margins under extreme operational thresholds.

4. Structural Parameter Isolation: Hard Freezing. Disabling targeted hard-freezing (`hard_freeze disabled`):

- In terms of open-set metrics, retaining the feature anchor and frozen BN prevents acute degradation (average EER achieves **4.15%**, ROC AUC achieves **0.9747**, and TAR @ $\mathrm{FAR} = 10^{-6}$ reaches **47.80%**).
- The primary advantage of Hard-Freezing is computational and systemic rather than solely metric-driven: as established in Table 3, unlocking the backbone introduces an additional backward pass, causing a 21.4% drop in throughput (from 63.00 down to 49.50 img/s) and increasing local personalization time by $\sim$26%.
- Furthermore, leaving the backbone unfrozen exposes the shallow convolutional filters to potential long-term drift if local iterations continue, whereas hard-freezing provides an absolute mathematical guarantee ($\Delta w_{\mathrm{conv}} = 0$) against filter degradation.

A.2 Summary of Mechanistic Insights

The leave-one-out findings reveal a hierarchical dependency among AsyGA components:

$$\underbrace{\mathcal{L}_{\text{reg}}\ (\text{Distillation Anchor})}_{\text{Prevents Feature Collapse}} \gg \underbrace{\text{BN Freezing}}_{\text{Prevents Covariate Shift}} \gg \underbrace{\mathcal{L}_{\text{con}}}_{\text{Enforces Wide Margin}} \approx \underbrace{\text{Hard Freezing}}_{\text{Guarantees Spatial Invariance \& Efficiency}}. \tag{7}$$

The feature anchor and BN freezing act as non-negotiable structural anchors against multi-band amnesia, while contrastive tuning and parameter freezing optimize operational margin separability and real-time execution throughput on consumer-grades.

B Theoretical Derivation and Proof Details for Asymmetric Modality Collapse

In this section, we analyze the optimization dynamics of naïve single-modality adaptation and formally prove how the AsyGA structural constraints preserve cross-spectral manifold topology.

B.1 Mathematical Formulation and Problem Setup

Let $\mathcal{X} = \bigcup_{m \in \mathcal{M}_{\text{spec}}} \mathcal{X}^m$ be the unified multispectral palmprint input space across spectral modalities $\mathcal{M}_{\text{spec}} = \{\text{NIR}, \text{Red}, \text{Green}, \text{Blue}\}$. Let $f(\cdot\,; W) : \mathcal{X} \to \mathbb{R}^d$ denote the deep feature extractor parameterized by weights W, and let $z(x; W) = f(x; W)/\|f(x; W)\|_2 \in \mathbb{S}^{d-1}$ be its projection onto the unit hypersphere $\mathbb{S}^{d-1}$.

Let W_G denote the Phase I global model obtained after federated convergence. We assume that W_G encodes a *cross-spectral invariant manifold* $\mathcal{M}$. Let $\mathcal{D}_{\text{pair}}$ denote the joint distribution of cross-spectral positive pairs sharing identical biometric identities, $(x^B, x^N) \sim \mathcal{D}_{\text{pair}}$, where $x^B \in \mathcal{X}^{\text{Blue}}$ and $x^N \in \mathcal{X}^{\text{NIR}}$. The pre-trained global representation satisfies the alignment condition:

$$\mathbb{E}_{(x^B, x^N) \sim \mathcal{D}_{\text{pair}}}\big[1 - \langle z(x^B; W_G), z(x^N; W_G)\rangle\big] \le \epsilon, \tag{8}$$

where $\langle \cdot, \cdot \rangle$ denotes the standard Euclidean inner product on $\mathbb{S}^{d-1}$, and $\epsilon \in (0, 1)$ is an upper bound on initial cross-spectral misalignment.

To quantify cross-spectral verification fidelity under continuous weight updates W_t, we define the cross-spectral similarity functional:

$$S(W) \triangleq \mathbb{E}_{(x^B, x^N) \sim \mathcal{D}_{\text{pair}}} \langle z(x^B; W), z(x^N; W)\rangle. \tag{9}$$

From equation 8, the initial global model satisfies $S(W_G) \ge 1 - \epsilon$.

B.2 Monotonic Degradation Under Naïve Single-Modality Adaptation

Consider an edge client adapting model W using synthetic data $\tilde{X}^B \subset \mathcal{X}^{\text{Blue}}$ restricted entirely to the Blue modality. Under standard gradient descent with learning rate $\eta > 0$, the parameter update rule at step t is:

$$W_{t+1} = W_t - \eta \nabla_W L_{\text{local}}(W_t; \tilde{X}^B). \tag{10}$$

B.3 Monotonic Degradation Under Naïve Single-Modality Adaptation

Consider an edge client adapting model W using synthetic data $\tilde{X}^B \subset \mathcal{X}^{\text{Blue}}$ restricted entirely to the Blue modality. Under gradient descent with learning rate $\eta > 0$, the update at step t is

$$W_{t+1} = W_t - \eta \nabla_W L_{\text{local}}(W_t; \tilde{X}^B). \tag{11}$$

We assume S is twice differentiable with $\|\nabla^2 S(W)\|_2 \le M$ and $\|\nabla_W L_{\text{local}}\|_2 \le G$. For any update $\Delta = W_{t+1} - W_t$, a second-order Taylor bound gives $\big|S(W_t + \Delta) - S(W_t) - \langle \nabla_W S(W_t), \Delta\rangle\big| \le \frac{M}{2}\|\Delta\|_2^2$. Setting $\Delta = -\eta \nabla_W L_{\text{local}}$ yields

$$S(W_{t+1}) \le S(W_t) - \eta \langle \nabla_W L_{\text{local}}(W_t; \tilde{X}^B), \nabla_W S(W_t)\rangle + \frac{M\eta^2 G^2}{2}. \tag{12}$$

Because the synthetic domain $\tilde{X}^B$ exhibits surface-ridge textures but lacks the subcutaneous venous topologies present in $\mathcal{X}^{\mathrm{NIR}}$, the local objective can reduce empirical risk through visible-band shortcuts, which biases its gradient relative to the multi-band manifold $\mathcal{M}$ (Li & Hoiem, 2017; Kirkpatrick et al., 2017). We formalize this bias as follows (note the sign convention: a *positive* inner product means that a descent step on L_{local} *decreases* S).

Assumption 1 (Gradient Misalignment Under Spectral Asymmetry)**.** *Let the training support be restricted to a single observed spectral domain $\mathcal{X}^B \subsetneq \mathcal{X}$. For every step $t \in [0, T_{\mathrm{ad}} - 1]$,*

$$\mathbb{E}_{\tilde{X}^B}\big[\langle \nabla_W L_{\mathrm{local}}(W_t; \tilde{X}^B), \nabla_W S(W_t)\rangle \,\big|\, W_t\big] \geq \delta, \quad \textit{for some constant } \delta > 0. \tag{13}$$

Assumption 1 is a modeling assumption. It is consistent with the degradation of Strategy 1 in Table 2 (overall EER $9.20\% \to 12.10\%$, NIR–Blue EER $\to 27.59\%$), but Table 2 does not measure the gradient inner product directly.

Theorem 1 (Asymmetric Modality Collapse)**.** *Under Assumption 1, let the step size satisfy $\eta \leq \frac{\delta}{MG^2}$. Then unconstrained local personalization over T_{ad} iterations satisfies*

$$\mathbb{E}[S(W_{T_{\mathrm{ad}}})] \leq S(W_0) - \frac{T_{\mathrm{ad}}\,\eta\,\delta}{2}. \tag{14}$$

Since $S \geq -1$, the bound is informative only for $T_{\mathrm{ad}} \leq 2\big(S(W_0) + 1\big)/(\eta\delta)$.

Proof. Taking conditional expectations in equation 12 and applying Assumption 1,

$$\mathbb{E}[S(W_{t+1}) \mid W_t] \leq S(W_t) - \eta\delta + \frac{M\eta^2 G^2}{2}. \tag{15}$$

For $\eta \leq \delta/(MG^2)$ we have $\frac{M\eta^2G^2}{2} \leq \frac{\eta\delta}{2}$, hence $\mathbb{E}[S(W_{t+1}) \mid W_t] \leq S(W_t) - \frac{\eta\delta}{2}$. Summing over $t = 0, \ldots, T_{\mathrm{ad}} - 1$ and taking total expectations gives equation 14. □

Note that $S(W_G) \geq 1 - \epsilon$ (Eq. equation 8) is a *lower* bound on the starting point; Theorem 1 therefore bounds the *drop* relative to $S(W_0)$ and does not by itself imply an absolute upper bound of the form $1 - \epsilon - T_{\mathrm{ad}}\eta\delta$. Since $S \leq 1$ always, one may write $\mathbb{E}[S(W_{T_{\mathrm{ad}}})] \leq 1 - T_{\mathrm{ad}}\eta\delta/2$.

B.4 Manifold Preservation via AsyGA Constraints

Let θ denote the trainable parameters under AsyGA and W_{frozen} the rest (see Sec. 3 for the partition). With $\Delta W_{\mathrm{frozen}} = 0$, the first-order change of S caused by frozen parameters is exactly zero, i.e. $\langle \nabla_{W_{\mathrm{frozen}}} S, \Delta W_{\mathrm{frozen}}\rangle = 0$; the update is confined to θ. (This is a statement about the *update*, not about the gradient $\nabla_{W_{\mathrm{frozen}}} L_{\mathrm{local}}$, which is not zero but is simply not applied.) The analysis below requires that S depends on θ, i.e. that θ contains layers that contribute to the embedding f.

Feature-anchored gradient realignment. The distillation loss is

$$L_{\mathrm{reg}}(\theta) = 1 - \frac{1}{B}\sum_{j=1}^{B} \langle z(x_j; \theta), z(x_j; W_G)\rangle. \tag{16}$$

Because a descent step on L_{reg} pulls the student embedding toward the teacher embedding, which satisfies the alignment condition equation 8, we make the following assumption (again with the convention that a negative inner product means that descent *increases* S):

$$\mathbb{E}_{\tilde{X}^B}\big[\langle \nabla_\theta L_{\mathrm{reg}}, \nabla_\theta S(W_t)\rangle\big] \leq -\kappa, \qquad \kappa > 0. \tag{17}$$

Let the task terms satisfy $\mathbb{E}\big[\langle \nabla_\theta(L_{\mathrm{ce}} + \lambda L_{\mathrm{con}}), \nabla_\theta S(W_t)\rangle\big] \leq \delta_{\mathrm{head}}$ for some $\delta_{\mathrm{head}} > 0$. Then

$$\mathbb{E}\big[\langle \nabla_\theta L_{\mathrm{total}}, \nabla_\theta S(W_t)\rangle\big] \leq \delta_{\mathrm{head}} - \gamma\kappa. \tag{18}$$

Proposition 1 (Manifold Stability Criterion)**.** *Suppose $\|\nabla_\theta L_{\mathrm{total}}\|_2 \leq G$, $\|\nabla^2 S\|_2 \leq M$, and*

$$\gamma \geq \frac{\delta_{\mathrm{head}}}{\kappa}. \tag{19}$$

Then, for plain gradient descent on θ with step size η,

$$\mathbb{E}[S(W_{T_{\mathrm{ad}}})] \geq S(W_0) - \frac{T_{\mathrm{ad}}M\eta^2G^2}{2} \geq 1 - \epsilon - \frac{T_{\mathrm{ad}}M\eta^2G^2}{2}. \tag{20}$$

Proof. By the two-sided Taylor bound, $S(W_{t+1}) \geq S(W_t) - \eta\langle \nabla_\theta L_{\text{total}}, \nabla_\theta S(W_t)\rangle - \frac{M\eta^2 G^2}{2}$. Taking expectations and using equation 18 with equation 19, the inner-product term is ≤ 0, so $\mathbb{E}[S(W_{t+1}) \mid W_t] \geq S(W_t) - \frac{M\eta^2 G^2}{2}$. Summing over t and using $S(W_0) \geq 1 - \epsilon$ yields equation 20. □

Remarks. (i) Condition equation 19 is a *sufficient threshold* on γ expressed through the unmeasured constants δ_{head} and κ; it does not by itself select $\gamma = 50$, and it does not explain the degradation observed for large γ in Fig. 7. (ii) The analysis uses plain gradient descent, whereas Algorithm 1 uses Adam. (iii) The distillation loss is evaluated on Blue-band inputs only; Assumption equation 17 therefore encodes that anchoring on the observed band transfers to the cross-spectral alignment S. We recommend supporting it with an embedding-drift measurement on real NIR images (student vs. teacher). (iv) S also depends on BN running statistics, which are not updated by gradients but change in training mode; the analysis treats them as fixed (BN in `eval()`). Hard freezing fixes parameters only, so the ablation without BN locking violates this premise, consistent with its degradation in Appendix A. In our implementation θ contains the projection layers, so S depends on θ. (v) We therefore describe the result as a *sufficient condition under stated assumptions*, not as an unconditional guarantee.