Non-Colliding Biometric Identities for Digital Entities: Geometry, Capacity, and Million-Scale Virtual Identity Provisioning

Yuyang Ji[1], **Yixuan Shen**[1], **Anil Jain**[2], **Xiaoming Liu**[3], **Feng Liu**[1]
[1] Department of Computer Science, Drexel University
[2] Department of Computer Science and Engineering, Michigan State University
[3] Department of Computer Science, University of North Carolina at Chapel Hill
jain@msu.edu, liuxm@cs.unc.edu, {yj428,ys844,fl397}@drexel.edu

Abstract

Digital entities such as AI agents and humanoid robots increasingly operate alongside real humans, yet their identity infrastructure is based on credentials rather than embodied biometric identity. We introduce Biometric Identity Provisioning (BIP), a new problem and solution framework that addresses: given an enrollment gallery of real human identities, provision virtual identities that are non-colliding with every enrolled identity, maintain sufficient inter-class separability, and are realizable as high-fidelity face images. The key geometric insight is that real face identities occupy a low-dimensional subspace of the embedding hypersphere, leaving no residual subspace for virtual identities. Hence, virtual identities must instead be allocated to unclaimed gaps within the real face manifold itself. BIP is therefore a constrained packing problem: available gaps vastly exceed any foreseeable enrollment scale, and provisioned identities are long-lived and lifecycle-managed, with rare conflicts under continued enrollment resolved by revocation and reassignment. Grounded in this geometry, our repulsion-based allocation is not bounded by any fixed provisioning count; we demonstrate 10M non-colliding virtual identity embeddings against a gallery of 360K real identities. Realizing these embeddings as face images requires a generator that operates outside the training distribution of real face images; we introduce GapGen, a gap-aware generator trained with a curriculum that progressively extends synthesis into non-colliding regions, validated at 1M photorealistic virtual face images. We further construct v-LFW, a virtual counterpart to the LFW face dataset, with protocols for virtual face verification, cross-reality matching, real-vs-virtual detection, and unified recognition and detection. Project

1 Introduction

Identity systems are being reshaped by a new class of actors: AI agents autonomously coordinate enterprise workflows [43], humanoid robots interact face-to-face with people in service environments, and digital co-workers operate under zero-trust policies alongside human employees. This is not a projected future: a recent industry survey reports that 40% of organizations already have AI agents in production, with another 31% running pilots or tests [5]. Microsoft has launched dedicated infrastructure to govern these agents as first-class identities [33]. Yet this infrastructure assigns credential identity (tokens, certificates, access scopes), not *embodied biometric identity*. As digital entities operate in physical spaces alongside real humans, a question arises: *how do we provision digital entities with long-lived, unique biometric identities that do not collide with any enrolled real human identity?* We term this the **Biometric Identity Provisioning (BIP)** problem (see Fig. 1).

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

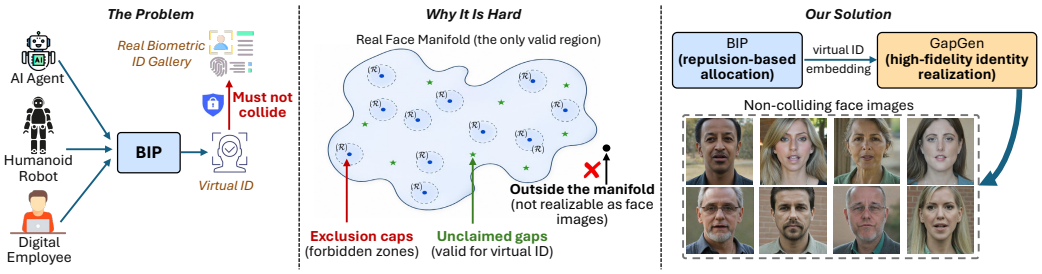

Figure 1: Should a digital entity have a distinctive face? As AI agents, humanoid robots, and digital employees operate alongside real humans, they require long-lived biometric identities that must not be mistaken for any enrolled real person (left). This is geometrically non-trivial: real face identities occupy a low-dimensional manifold with no residual subspace, forcing virtual identities into the narrow unclaimed gaps between exclusion caps; outside the manifold, embeddings cannot be realized as natural face images (middle). BIP provisions non-colliding virtual identity embeddings via repulsion-based allocation, and GapGen realizes them as high-fidelity 1024×1024 portrait faces (right).

Face is the natural entry point for biometric identity provisioning for virtual entities: it is the modality most directly involved in human-entity interaction, the most mature in generative modeling, and the one for which large-scale enrollment galleries already exist [23]; gait [61, 44], body shape [28, 30] and whole-body traits [26, 27, 46, 20, 47] are natural extensions. A substantial body of work has studied synthetic face generation, but with a fundamentally different objective: producing training data for face recognition models. Methods such as Arc2Face [36] and Vec2Face/Vec2Face+ [54, 53] condition face generation on real human identity embeddings, generating faces that correspond to existing real people in embedding space; they are *identity cloners*, not *identity creators*. Methods without this conditioning, such as DCFace [18], optimize for separability among synthetic faces, but whether generated identities collide with real enrolled humans is simply not a design concern; empirically, we show such collisions do occur at non-trivial rates. Thus, a synthetic identity useful for training a classifier is not necessarily a biometric identity suitable to assign to a digital entity. Existing methods neither define nor jointly satisfy the requirements of high-fidelity realization, inter-class separability, and guaranteed **non-collision** with any enrolled real humans.

Hence, we propose a new formulation, termed BIP: synthetic identities shall be treated not merely as generated images, but as allocatable positions in a shared biometric identity space. We define BIP as:

> *Given a mapping from face images to d-dimensional L2-normalized embeddings on $\mathbb{S}^{d-1}$ and an enrollment gallery $\mathcal{R} \subset \mathbb{S}^{d-1}$, provision a set of virtual embeddings $\mathcal{V} \subset \mathbb{S}^{d-1}$ satisfying non-collision with $\mathcal{R}$ at threshold τ, inter-class separability within $\mathcal{V}$, and high-fidelity realizability.*

A natural but naïve geometric view treats BIP as a **capacity** problem: the number of separable positions on $\mathbb{S}^{d-1}$ at threshold τ is arbitrarily large even within the effective face subspace, suggesting ample room for virtual identities. This view is misleading: ***(1) No residual subspace exists.*** Virtual identities satisfying BIP constraints do not occupy arbitrary low-energy directions of the hypersphere; their PCA energy distribution closely follows that of real identities, with both concentrating over 95% of their variance within the same $k{=}269$ leading principal components out of $d{=}512$. Valid virtual identities must occupy *unclaimed gaps within the real face manifold* rather than a geometrically separate region. ***(2) Capacity does not equal realizability.*** Embeddings satisfying the non-collision constraint are not automatically realizable as high-fidelity face images: the further a virtual identity is pushed from real identity territories, the harder it becomes for a generative model to generate it as a natural face. BIP is, therefore, a constrained identity-allocation problem on the real identity manifold, not an unconstrained sampling problem on $\mathbb{S}^{d-1}$.

Grounded in this geometric view, we address BIP through an identity-first pipeline that separates allocation in embedding space from realization in image space:

(1) Formalization. We introduce the first formal definition of BIP and three evaluation metrics operationalizing its requirements: non-collision, inter-class separability, and perceptual image quality.
(2) Geometry and allocation. We characterize the geometry of the real identity manifold and show that safe virtual identities must occupy directional gaps within it. We derive the **repulsion direction**, z^* as a heuristic that moves candidates away from real identity clusters, enriched with PCA-aware

noise to maintain face-manifold compatibility. Combined with exact hard checks, the allocation strategy is not bounded by any fixed provisioning count: the face manifold contains vastly more unclaimed gaps than any foreseeable enrollment gallery, and we demonstrate 10M non-colliding virtual identity embeddings against a gallery of 360K real identities with no observed collision. ***(3) Realization via GapGen.*** We introduce a gap-aware generator trained with a progressive curriculum that extends face synthesis into the non-collision region. ***(4) Benchmark.*** We construct **v-LFW**, a virtual counterpart to the LFW face dataset with 5,749 provisioned identities and 13,233 images, and introduce protocols spanning virtual face verification, cross-reality matching, real-vs-virtual detection, and unified recognition and detection. To support these protocols, we provide IAPCT (Identity-Anchored Patch Consistency Transformer) as a lightweight diagnostic evaluation tool.

2 Related Work

Synthetic Face Generation for Recognition. Synthetic face data has been widely explored as a privacy-preserving alternative to web-collected face datasets. Early work such as SynFace [39] and SFace [4] uses class-conditional generative models, while DigiFace-1M [2] adopts graphics-based rendering. More recent methods improve identity diversity via diffusion: IDiff-Face [3] introduces identity-conditioned diffusion, and DCFace [18] disentangles identity and style via dual-condition diffusion. Arc2Face [36] and Vec2Face and Vec2Face+ [54, 53] further generate identity- and attribute-controllable face datasets from recognition-based features, with Vec2Face+ explicitly controlling inter-class separability, intra-class variation, and identity consistency. VIGFace [21] is especially related because it pre-assigns virtual identities in feature space before synthesis. However, these methods are designed primarily for privacy-friendly or high-performing face recognition training data [29]. They do not formulate identity generation as gallery-conditioned provisioning: whether a generated identity collides with an enrolled real human identity is not a primary constraint, and long-lived assignment to digital entities is outside their problem setting.

Hyperspherical Identity Allocation. Modern face recognition systems [17, 19] map faces to normalized embeddings on a high-dimensional hypersphere, where identity verification uses cosine similarity or angular distance. SphereFace [32], CosFace [50], ArcFace [8], and AdaFace [16] introduce angular margin losses that structure this hypersphere to improve inter-class separation and intra-class compactness, with AdaFace further adapting the margin to image quality via feature norms. The most closely related work to BIP is HyperFace [34], which formulates synthetic face data generation as a packing problem on the recognition hypersphere and optimizes identity embeddings via gradient descent, while regularizing embeddings on the face manifold. However, HyperFace maximizes inter-class separation *among synthetics only*; non-collision with real humans is not a design objective. BIP differs fundamentally: the enrollment gallery $\mathcal{R}$ is a hard constraint, non-collision with every identity in $\mathcal{R}$ at threshold τ is the primary requirement, and our geometric analysis shows that valid virtual identities must reside within the real identity manifold rather than a geometrically separate space.

Identity-Aware Forensics. Deepfake and face forgery detection has been formulated as binary classification between real and manipulated images [7]. Many detectors exploit low-level artifacts, frequency-domain cues, blending boundaries, or generator-specific traces [40, 41, 10, 38], but they often generalize poorly across manipulation types and generation models, and benchmark accuracy may overstate forensic ability [35]. Identity-aware forensics treats identity consistency as a forensic signal: ID-Reveal [6] learns person-specific motion patterns and detects manipulations as deviations from expected behavior, while other works exploit identity inconsistency or face recognition features to detect face swaps [55, 22]. These methods assume that real identities already exist and ask whether a given image is manipulated. BIP addresses a complementary question: how to provision new biometric identities guaranteed to lie outside all enrolled real identity territories. The resulting non-overlapping partition of the embedding space between real and virtual identities is a direct benefit of the BIP constraints, not a product of forensic training.

3 Methodology

3.1 Formalization

Let $\phi : \mathcal{X} \to \mathbb{S}^{d-1}$ denote a face recognition encoder that maps a face image x to a d-dimensional L2-normalized embedding on the unit hypersphere $\mathbb{S}^{d-1}$.

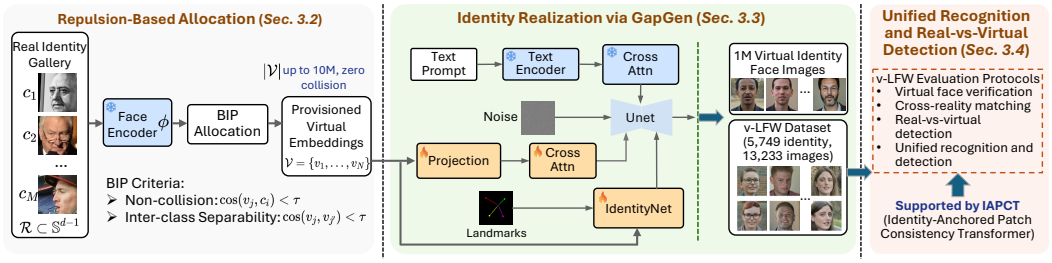

Figure 2: BIP pipeline. Left: Repulsion-based allocation provisions $\mathcal{V}=\{v_1,\ldots,v_N\}$ satisfying $\cos(v_j,c_i)<\tau$ and $\cos(v_j,v_{j'})<\tau$, scaling to $|\mathcal{V}|$=10M with zero observed collision. Middle: GapGen renders each $s\in\mathcal{V}$ into a 1024×1024 face image $\tilde{x}=G(s)$, producing 1M virtual identity images and the v-LFW benchmark. Right: v-LFW supports four protocols spanning virtual face verification, cross-reality matching, real-vs-virtual detection, and unified recognition and detection, supported by IAPCT as a lightweight diagnostic tool.

Assumption 1 (Angular-margin Encoder)**.** *The encoder ϕ is trained with an angular margin loss (*e.g.*, ArcFace [8] or AdaFace [16]). Embeddings with $\cos(\phi(x),\phi(x'))\geq\tau$ are accepted as the same identity, where $\tau\in(0,1)$ is set at the operating point of the recognition system (see Appx. B for details on τ). Equivalently, τ defines an angular boundary on $\mathbb{S}^{d-1}$: any two embeddings within angular distance $\arccos(\tau)$ are recognized as the same identity.*

Each real identity $i\in\{1,\ldots,M\}$ with images $\mathcal{X}_i=\{x_i^{(1)},\ldots,x_i^{(n_i)}\}$ is represented by the dominant direction of its embedding cluster:

$$c_i=\text{normalize}\left(\sum_{k=1}^{n_i}\phi\left(x_i^{(k)}\right)\right), \tag{1}$$

with $c_i=\phi(x_i^{(1)})$ when $n_i=1$. The enrollment gallery is $\mathcal{R}=\{c_1,\ldots,c_M\}\subset\mathbb{S}^{d-1}$.

Definition 1 (Biometric Identity Provisioning)**.** *Given an encoder ϕ satisfying Assumption 1, an enrollment gallery $\mathcal{R}\subset\mathbb{S}^{d-1}$ of M real identity centroids, a verification threshold τ, and a provisioning count N, the **Biometric Identity Provisioning (BIP)** problem is to generate $\mathcal{V}=\{v_1,\ldots,v_N\}\subset\mathbb{S}^{d-1}$ satisfying:*

$$\cos(v_j,c_i)<\tau,\quad\forall i\in\{1,\ldots,M\},\ \forall j\in\{1,\ldots,N\} \tag{2}$$

$$\cos(v_j,v_{j'})<\tau,\quad\forall j\neq j'. \tag{3}$$

The realizability requirement, that each v_j corresponds to a high-fidelity face image whose re-encoded embedding satisfies Eqs. (2)–(3), is addressed by the gap-aware generator in Sec. 3.3. Fig. 2 summarizes the full pipeline of allocation, realization, and evaluation.

Remark 1. *The constraints in Definition 1 are* encoder-specific *and* gallery-relative*: they are stated for a given ϕ and τ, and are guaranteed w.r.t. $\mathcal{R}$ at provisioning time. BIP identities are therefore long-lived and lifecycle-managed rather than permanent. When new real identities $\Delta\mathcal{R}$ are subsequently enrolled, any v_j with $\cos(v_j,r')\geq\tau$ for some $r'\in\Delta\mathcal{R}$ must be revoked and reassigned against the updated gallery $\mathcal{R}\cup\Delta\mathcal{R}$; the enrolled real identity is never modified, and the rest of $\mathcal{V}$ is never re-provisioned. This requires $O(|\mathcal{V}|\times|\Delta\mathcal{R}|)$ cosine operations in the exact case, reducible with approximate nearest-neighbor indexing [14]. Provisioning with a safety buffer $\tau_{safe}=\tau-\Delta$ enlarges the margin to the enrolled gallery and adds a monitoring zone $[\tau_{safe},\tau)$ (Proposition 2 and the lifecycle policy in Appx. C.6); open-world behavior is studied in Sec. 4.*

BIP Criteria. Three metrics directly operationalize Definition 1. ***(i) Non-collision:*** the percentage of provisioned virtual identities $v_j\in\mathcal{V}$ satisfying Eq. (2) against all M enrolled real centroids in $\mathcal{R}$; a score below 100% means some virtual identities are indistinguishable from real humans in the recognition system, posing a direct privacy risk. ***(ii) Inter-class separability (Inter-Sep):*** the percentage of virtual identities v_j satisfying Eq. (3) against every other identity in $\mathcal{V}$; a score below 100% means some virtual entities share an identity, making them unreliable as distinct assignable identities. ***(iii) FID*** [11] (against FFHQ [15]) measures perceptual image quality of generated virtual identities against portrait photographs. Prior synthetic face methods produce low-resolution 112×112 training crops, whereas digital entities require portrait-quality images displayable and recognizable in physical environments.

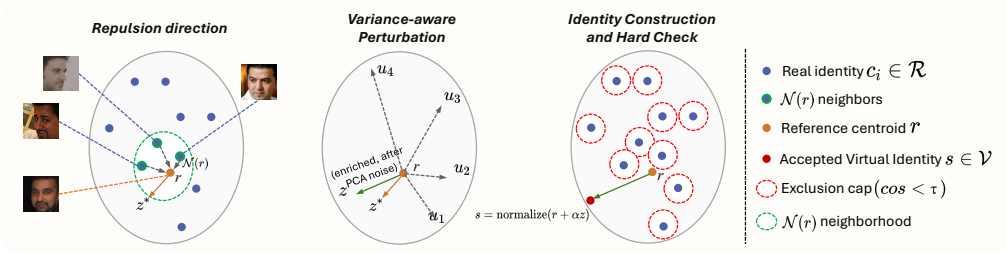

Figure 3: Repulsion-based virtual identity allocation. Left: $z^*{=}-m/\|m\|_2$ points away from the weighted centroid of nearest real neighbors $\mathcal{N}(r)$. Middle: z^* is enriched with PCA-aware noise along $\{u_k\}$ (weighted by σ_k) to obtain z. Right: Candidate $s{=}\text{normalize}(r{+}\alpha z)$ is accepted only if $\cos(s, c_i){<}\tau$ for all $c_i \in \mathcal{R}$ and $\cos(s, v_j){<}\tau$ for all $v_j \in \mathcal{V}_t$; failed candidates are resampled.

3.2 Geometry of the Real Identity Manifold and Repulsion-Based Allocation

Geometry of the Real Identity Manifold. ArcFace ResNet-100 [8] trained on Glint360K [1] is used as ϕ ($d{=}512$), and all $M{=}360{,}232$ identity centroids form $\mathcal{R}$. To characterize where virtual identities can be validly placed and to guide provisioning diversity, we perform PCA on these centroids in the ambient Euclidean space, yielding eigenvectors $U = [u_1, \ldots, u_d]$ with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_d \geq 0$, where $k \in \{1, \ldots, d\}$ denotes rank in explained variance, and per-direction standard deviations $\sigma_k = \sqrt{\lambda_k}$; both $\{u_k\}$ and $\{\sigma_k\}$ are reused in the allocation strategy below.

Observation 1 (Low-dimensional Real Identity Manifold)**.** *Let the* principal energy *for dimension k be $E(k) = \left(\sum_{k'=1}^{k} \lambda_{k'}\right) / \left(\sum_{k'=1}^{d} \lambda_{k'}\right)$, where $k' \in \{1, \ldots, d\}$ is a summation index. The real identity centroids in $\mathcal{R}$ concentrate over 95% of their total variance within the top $k{=}269$ principal components (Fig. 7 in Appx. C.1), despite residing in $\mathbb{S}^{511}$. Since ϕ is trained on real face images, realizable face embeddings are expected to remain concentrated near the dominant real-face subspace. Low-variance residual directions may exist in the ambient embedding space, but they are unlikely to provide a reliable region for high-fidelity face realization. Valid virtual identities should therefore occupy* unclaimed gaps within the real face manifold*, rather than a geometrically separate region.*

Observation 1 rules out orthogonal or free-subspace constructions [21] and motivates the gap-allocation strategy below.

Capacity. Modeling the real face manifold as $\mathbb{S}^{\tilde{d}-1}$ with $\tilde{d}{=}269$ (Observation 1), the Gilbert–Varshamov bound gives $A_{\text{pack}}(\tilde{d}, \tau) \geq A_{\text{GV}}(\tilde{d}, \tau) = 1/\mu(\tau, \tilde{d})$, where A_{pack} is the largest number of points on this sphere with pairwise cosine below τ and $\mu(\tau, \tilde{d})$ is the normalized volume of a spherical cap of cosine radius τ (Proposition 1, Appx. C). At $\tau{=}0.391$, $A_{\text{GV}} = 7.41 \times 10^{10} \approx 2^{36.1}$, about 7,400 times the 10M identities provisioned in Sec. 4. This is a model-based headroom for allocation in embedding space; separability after realization is measured in Sec. 4.

Repulsion Direction. Given a reference centroid $r \in \mathcal{R}$, let $\mathcal{N}(r) = \{c_{n_1}, \ldots, c_{n_K}\}$ be its K nearest neighbors in $\mathcal{R}$, with softmax weights:

$$w_{n_k} = \frac{\exp(-d(r,\, c_{n_k})/t)}{\sum_{k'=1}^{K} \exp(-d(r,\, c_{n_{k'}})/t)}, \tag{4}$$

where $d(r, c) = 1 - \cos(r, c)$ and t denotes temperature.

Remark 2 (Repulsion Direction)**.** *Define z^* as the negation of the weighted neighborhood centroid:*

$$z^* = -\frac{\sum_{k=1}^{K} w_{n_k}\, c_{n_k}}{\left\|\sum_{k=1}^{K} w_{n_k}\, c_{n_k}\right\|_2}. \tag{5}$$

z^* serves as a repulsive heuristic*: it moves candidates away from the weighted neighborhood centroid of r, increasing the probability that a candidate satisfies the constraints of Definition 1. It does not guarantee them; non-collision is enforced by the exact checks applied to every candidate before acceptance (see Appx. C.5 for an analysis of z^* after spherical normalization).*

To improve diversity while keeping provisioned identities on the real face manifold, we enrich z^* with *PCA-aware noise*, random perturbations aligned to the principal directions $\{u_k\}$:

$$z = \text{normalize}\left(z^* + \kappa \sum_{k=1}^{d} \eta_k \sigma_k u_k\right), \quad \eta_k \sim \mathcal{N}(0,1), \tag{6}$$

where $\kappa \geq 0$ balances direction fidelity against sample diversity. Weighting by σ_k ensures perturbations are larger along high-variance principal directions and smaller along low-variance ones, keeping provisioned embeddings aligned with the natural directions of variation in $\mathcal{R}$. The relationship between κ, α, and the theoretical face manifold capacity is analyzed in Appx. C (Sec. C.8).

Identity Construction and Hard Check. A candidate virtual identity is constructed as:

$$s = \text{normalize}(r + \alpha z), \tag{7}$$

where $r \in \mathcal{R}$ is the reference centroid, z is from Eq. (6), and $\alpha > 0$ controls perturbation strength.

Lemma 1 (Effect of Perturbation Strength)**.** *For any $r, z \in \mathbb{S}^{d-1}$, the cosine similarity between $s = \text{normalize}(r + \alpha z)$ and r satisfies:*

$$\cos(s,\, r) = \frac{1 + \alpha\,(r \cdot z)}{\sqrt{1 + 2\alpha\,(r \cdot z) + \alpha^2}}. \tag{8}$$

In the special case κ=0, $z = z^*$; as r is positively correlated with its nearest neighbors $\mathcal{N}(r)$ on $\mathbb{S}^{d-1}$, the weighted centroid $\sum_k w_{n_k} c_{n_k}$ is positively aligned with r, giving $r \cdot z^* < 0$, and $\cos(s, r)$ is strictly decreasing in α. In the orthogonal case $r \perp z$, Eq. (8) simplifies to $\cos(s, r) = 1/\sqrt{1+\alpha^2}$.

The parameter α controls the fundamental trade-off in BIP, as quantified in Tab. 1: larger α increases angular displacement from real identity territories improving non-collision, while smaller α keeps provisioned identities closer to the face manifold improving realizability. Each candidate s is accepted only when it satisfies the BIP constraints of Definition 1 with respect to the current state of the gallery:

$$\cos(s,\, c_i) < \tau \quad \forall\, i \in \{1, \ldots, M\} \quad \text{(non-collision with } \mathcal{R}\text{)}, \tag{9}$$

$$\cos(s,\, v_j) < \tau \quad \forall\, v_j \in \mathcal{V}_t \quad \text{(separability within } \mathcal{V}_t\text{)}, \tag{10}$$

where $\mathcal{V}_t$ denotes the set of accepted virtual identities at step t; failed candidates are resampled with fresh η. By induction, if every candidate passes these checks before acceptance, the final $\mathcal{V}$ satisfies Eqs. (2)–(3) by construction. Appx. C (Corollary 1) establishes that $\alpha > \sqrt{1-\tau^2}/\tau$ is sufficient for non-collision in the repulsive case; the full α ablation is in Tab. 1. We demonstrate 10M non-colliding virtual identity embeddings against $|\mathcal{R}|$=360K with no observed collision at $\tau = 0.391$, well within the capacity headroom above. Fig. 3 shows the three components: repulsion direction, PCA-aware perturbation, and hard non-collision check.

3.3 Identity Realization via GapGen

A provisioned embedding $s \in \mathcal{V}$ defines a target location in the biometric space; **GapGen** realizes it as a face image $\tilde{x}$ such that $\phi(\tilde{x})$ remains faithful to s. The challenge is that, since s is intentionally displaced from the real identity distribution, a pretrained generator G, optimized on real embeddings, exhibits reduced fidelity when conditioned on such out-of-distribution inputs.

Identity-conditioned Generator. We build on InstantID [51] as the identity-conditioned generator G, which accepts an ArcFace embedding as its condition and generates 1024×1024 face images, denoted as $\tilde{x} = G(s)$. Since BIP uses the same ArcFace embedding space for both allocation and generation, provisioned embeddings $s \in \mathcal{V}$ can be passed directly as conditions without additional projection. Full implementation details are in Appx. D.

GapGen Fine-tuning. We fine-tune G with a curriculum that interleaves *real steps* and *virtual steps* at mixing ratio $\beta \in [0, 1]$. Both step types share the round-trip identity loss:

$$\mathcal{L}_{\text{RT}}(\tilde{x},\, e^*) = 1 - \cos(\phi(\tilde{x}),\, e^*)\,, \tag{11}$$

where e^* is the target identity embedding and $\rho_{\text{RT}} = \cos(\phi(\tilde{x}), e^*)$ is the *round-trip similarity* quantifying rendering fidelity.

Real steps train G on real images x from FFHQ [15], chosen for its 1024×1024 resolution consistent with our generation target, with target $e^* = \phi(x)$:

$$\mathcal{L}_{\text{real}} = \mathcal{L}_{\text{denoise}} + \lambda_{\text{id}}\,\mathcal{L}_{\text{RT}}(\tilde{x},\,\phi(x)) + \lambda_{\text{perc}}\,\mathcal{L}_{\text{perc}}, \tag{12}$$

where $\mathcal{L}_{\text{denoise}} = \mathbb{E}_{t,\epsilon}[\|\epsilon - \epsilon_\theta(x_t, t, \phi(x))\|^2]$ is the standard noise prediction loss [12], and $\mathcal{L}_{\text{perc}}$ is a perceptual similarity loss [60]. Real steps adapt G to the high-resolution portrait domain and preserve identity fidelity on in-distribution embeddings.

Virtual steps address the core challenge of BIP realization: provisioned embeddings $s \in \mathcal{V}$ have no paired ground-truth images, as virtual identities have no physical counterpart. At each virtual step, we sample a reference centroid $r \in \mathcal{R}$, draw $\alpha \sim \text{Uniform}(\alpha_{\min}{=}2, \alpha_{\max}{=}5)$, construct $s = \text{normalize}(r + \alpha z)$ on-the-fly, generate $\tilde{x} = G(s)$ via truncated DDIM sampling [42], and optimize $\mathcal{L}_{\text{RT}}(\tilde{x}, s)$ (Eq. (11) with $e^*{=}s$). This unsupervised objective drives G to extend its generative capability into the non-collision region of the embedding space, *without using paired training data*.

Intra-class Variation. Multiple images per virtual identity are generated by varying facial pose, lighting conditions, photographic style, and background through different pose configurations and text style prompts, producing within-identity variation comparable to unconstrained portrait photography.

3.4 Unified Recognition and Real-vs-Virtual Detection

The BIP criteria of Sec. 3.1 evaluate virtual identities: whether they are non-colliding, separable, and realizable. Deploying digital entities alongside real humans raises a complementary set of questions that no existing benchmark addresses: can virtual identities be verified against themselves? How do they interact with real identities under standard recognition protocols? Can a deployed system reliably distinguish real from virtual? **v-LFW** is built to answer these questions.

Benchmark Construction. We construct **v-LFW** to mirror LFW [13] in both identity count (5,749) and image count (13,233), supporting controlled evaluation under established face recognition protocols. Each virtual identity is provisioned against $\mathcal{R}$ augmented with all LFW identities, ensuring non-collision with the benchmark. Images are rendered at 1024×1024 with diverse pose and style variation, then filtered for face quality. We define four protocols: virtual face verification, cross-reality matching, real-vs-virtual detection, and unified recognition/detection.

IAPCT: A Diagnostic Evaluation Tool. To support the real-vs-virtual detection and unified recognition protocols of v-LFW, we provide **IAPCT** (Identity-Anchored Patch Consistency Transformer) as a lightweight diagnostic tool built on the frozen encoder ϕ; it is not a primary contribution of this work. Identity-conditioned generators match the target embedding at the output level but impose no constraint on intermediate feature maps, creating detectable inconsistencies between local patch statistics and the global identity embedding $e = \phi(x)$ that IAPCT exploits. Let $e = \phi(x) \in \mathbb{S}^{d-1}$ denote the identity embedding from the frozen backbone. Multi-scale spatial tokens $\{t_{l,p}\}$ are extracted from four intermediate feature maps $\{F_l\}_{l=1}^4$ and projected to $d_{\text{model}}{=}256$ via per-stage linear layers. The projected identity anchor $e_{\text{proj}} = W_{\text{id}}\,e$ serves as the cross-attention key in a 4-layer transformer encoder [49, 9]; each layer applies self-attention among spatial tokens followed by cross-attention to e_{proj}, producing patch-identity consistency scores:

$$\gamma_{l,p} = \sigma\left(\frac{(W_Q\,t_{l,p})\cdot(W_K\,e_{\text{proj}})}{\sqrt{d_k}}\right) \in (0,1), \tag{13}$$

where σ is the sigmoid function and $\gamma_{l,p}$ measures the consistency of patch p at stage l with the claimed identity e. Real faces yield uniformly high $\gamma_{l,p}$ since every spatial region originates from the same physical person; virtual identity images yield heterogeneous $\gamma_{l,p}$ due to unconstrained intermediate features. A `[CLS]` token produces $\hat{y}$ via an MLP head with loss $\mathcal{L}{=}\text{BCE}(\hat{y}, y){+}\lambda_c(H(\gamma^{\text{virtual}}){-}H(\gamma^{\text{real}}))$, where $H(\gamma^{(x)}){=}{-}\sum_{l,p}\bar{\gamma}_{l,p}^{(x)}\log\bar{\gamma}_{l,p}^{(x)}$ is the entropy of the normalized attention distribution and ϕ remains frozen. The identity embedding $e{=}\phi(x)$ simultaneously serves face verification at no additional backbone cost. Full details are in Appx. E.

4 Experiments

Baselines. We compare BIP with representative synthetic face generation methods: *DCFace* [18], a dual-condition diffusion model optimized for inter-class separability among synthetics; *Arc2Face* [36],

Table 1: BIP Non-Collision / Inter-Class Sep (%) across identity scale $|\mathcal{V}|$, perturbation strength α, and threshold τ. At the operating point τ=0.391, $\alpha \geq 4$ achieves 100% Non-Collision at every scale, with Inter-Sep of 100% up to 1M and 99.98% at 10M.

$\lvert\mathcal{V}\rvert$	α	$\tau = 0.448$	$\tau =$ **0.391**	$\tau = 0.360$	$\tau = 0.341$	$\tau = 0.330$	$\tau = 0.319$
100K	2	100.00/100.00	85.63/100.00	64.57/99.99	48.69/99.93	39.05/99.81	30.55/99.54
	3	100.00/100.00	99.98/100.00	99.80/99.99	98.96/99.93	97.46/99.81	94.28/99.55
	4	100.00/100.00	**100.00 / 100.00**	99.91/99.99	99.39/99.94	98.19/99.82	95.55/99.54
	5	100.00/100.00	100.00/100.00	99.91/99.99	99.40/99.94	98.23/99.81	95.62/99.53
1M	2	100.00/100.00	85.61/100.00	64.48/99.92	48.51/99.42	39.01/98.24	30.42/95.48
	3	100.00/100.00	99.98/100.00	99.78/99.93	98.94/99.43	97.42/98.28	94.20/95.54
	4	100.00/100.00	**100.00 / 100.00**	99.91/99.92	99.38/99.43	98.17/98.27	95.52/95.53
	5	100.00/100.00	100.00/100.00	99.91/99.92	99.38/99.43	98.22/98.27	95.53/95.54
10M	2	100.00/100.00	85.10/99.97	64.74/99.22	48.06/94.51	39.00/85.55	30.16/66.62
	3	100.00/100.00	99.98/99.98	99.77/99.21	98.99/94.67	97.39/85.50	94.16/67.06
	4	100.00/100.00	**100.00 / 99.98**	99.91/99.24	99.39/94.66	98.27/85.57	95.36/66.99
	5	100.00/100.00	100.00/99.98	99.91/99.23	99.40/94.64	98.28/85.53	95.38/66.94

Table 2: Synthetic face generation under BIP criteria. All methods are re-encoded by the same frozen ArcFace, tested against $\mathcal{R}$ (M=360K) at τ=0.391. FID measures perceptual image quality against FFHQ. RT pass rate reports the fraction of generated images whose re-encoded embedding satisfies ρ_{RT}>0.6, quantifying how faithfully the generator realizes the target identity embedding. CASIA-WebFace is shown as a real-data reference. Vec2Face* renders identities allocated by BIP in its own embedding space.

Method	#Identity	Non-Collision ↑	Inter-Sep ↑	FID ↓	RT pass rate ↑	Resolution
CAS. (Real) [58]	10.6K	0.32	90.08	131.70	-	250×250
DCFace [18]	10K	72.38	49.20	139.70	-	112×112
Arc2Face [36]	10K	93.90	88.67	74.11	0.00	512×512
Vec2Face [54]	10K	53.92	65.02	159.30	62.78	112×112
Vec2Face [54]	100K	52.95	45.44	159.18	61.43	112×112
Vec2Face [54]	500K	52.93	36.22	159.29	62.01	112×112
Vec2Face* [54]	500K	98.45	75.61	-	-	112×112
BIP+GapGen (ours)	10K	**98.91**	**99.57**	56.20	87.94	1024×1024
BIP+GapGen (ours)	100K	98.67	97.84	**54.63**	**89.42**	1024×1024
BIP+GapGen (ours)	500K	98.62	86.30	55.62	88.31	1024×1024
BIP+GapGen (ours)	1M	98.61	78.36	56.47	88.75	1024×1024

which conditions generation on real ArcFace embeddings; and *Vec2Face* [54], which synthesizes identity-consistent faces from target embeddings with explicit separability control. All baselines are designed for generating face recognition training data, not biometric identity provisioning.

BIP Configuration. We use ArcFace ResNet-100 [8] trained on Glint360K [1] as ϕ (d=512, M=360,232). Unless otherwise stated, we set τ=0.391 (FAR$\approx 2 \times 10^{-5}$ on IJB-B; Appx. B), α=4.0, K=10 nearest neighbors, temperature t=0.1, and PCA noise weight κ=1.0. GapGen is fine-tuned from InstantID [51] with mixing ratio β=0.2, λ_{id}=0.1, λ_{perc}=0.1, generating faces of 1024×1024 pixels. The large-scale setting provisions ***10 million*** virtual identities; image realization is validated at ***1 million*** identities. Additional details are in the corresponding sections of the Appx.

Geometry and Scale: Non-Collision Holds at Ten Million. Tab. 1 jointly ablates perturbation strength α, provisioning scale $|\mathcal{V}|$, and verification threshold τ across the full IJB-B operating range (see Fig. 6 in Appx. B). At the primary threshold τ=0.391 and $\alpha \geq 4$, Non-Collision reaches **100% at 100K, 1M, and 10M scale**, with Inter-Sep at 100% up to 1M and 99.98% at 10M: ten million virtual identities are provisioned against 360K real identities with zero observed collision, consistent with the theoretical sufficient bound $\alpha^*(0, 0.391) \approx 2.35$ (Corollary 1). α=2 falls below this threshold and yields only 85.10–85.63% Non-Collision, while $\alpha \geq 3$ already recovers near-perfect performance, confirming the theoretical prediction. Scaling from 100K to 10M incurs negligible cost: Non-Collision remains at 100% throughout, and Inter-Sep stays at or above 99.98% at τ=0.391, demonstrating that BIP allocation is not capacity-limited at any scale tested. At stricter thresholds (smaller τ), Inter-Sep degrades more noticeably at 10M (*e.g.*, 66.99% at τ=0.319), consistent with the tighter capacity headroom at extreme operating points (Appx. C); τ=0.391 remains robust across all scales. All sets in Tab. 1 are provisioned with the hard checks of Eqs. (9)–(10) at τ_{check}=0.435 and evaluated at each τ; for $\tau < \tau_{\text{check}}$, the scores thus reflect how well α alone keeps candidates clear of $\mathcal{R}$.

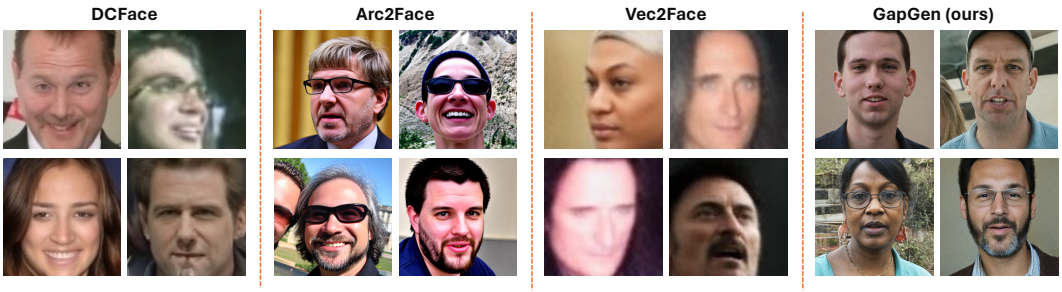

Figure 4: Qualitative comparison with synthetic face generators. Four randomly sampled identities per method. GapGen produces photorealistic 1024×1024 portrait faces with natural texture and coherent lighting, free from blurring, distortion, and uncontrolled backgrounds of the baselines.

Realization Quality: GapGen Outperforms Baselines across All BIP Criteria. Tab. 2 compares BIP+GapGen against baselines under the three BIP criteria at τ=0.391. While BIP allocation scales to 10M embeddings, image realization is validated up to 1M virtual images due to the computational cost of 1024×1024 generation; this reflects a deliberate fidelity-scale trade-off rather than an allocation bottleneck. Existing methods fail on at least one criterion: DCFace [18] reaches only 49.20% Inter-Sep; Vec2Face [54] (reproduced from released code) achieves only 53.92% Non-Collision at 10K with Inter-Sep collapsing to 36.22% at 500K; Arc2Face [36] achieves 93.90% Non-Collision but with an RT pass rate of 0.00, indicating complete failure of embedding fidelity after re-encoding. BIP+GapGen achieves 98.91% Non-Collision, 99.57% Inter-Sep, FID 56.20, and round-trip pass rate 87.94% at 10K, the best across all criteria at up to $9\times$ higher resolution than baselines. As scale increases, Non-Collision remains stable (98.61% at 1M); Inter-Sep decreases to 78.36%, reflecting the realization gap between embedding-space allocation (Tab. 1, Inter-Sep $\geq$ 99.98%) and image-level realization at scale. Allocation capacity and realized capacity therefore differ. Under the same encoder, threshold, and protocol, re-encoded Inter-Sep stays above 99% up to 29,830 identities and above 90% up to 300,132, against 12,777 and 141,248 for real Glint360K identities (Appx. F.3). Fig. 4 confirms the qualitative advantage; additional visual results are in Appx. E.1 and F.2.

Deepfake Detectability of Virtual Faces. A concern with high-fidelity virtual face generation is whether provisioned identities could be misused as deepfakes or synthetic personas. We evaluate this by applying five SoTA deepfake detectors to our generated virtual faces zero-shot, without fine-tuning on virtual faces; all detectors are trained on AIFaceFairnessBench [25]. As in Tab. 3a, all detectors achieve high accuracy on BIP-generated faces, indicating that existing detectors already flag provisioned virtual identities without modification. The non-collision guarantee and deepfake detectability are complementary: the former ensures virtual identities are biometrically distinct from all real humans, while the latter confirms they remain within the reach of existing forensic systems.

Open-World Non-Collision as Real Galleries Grow. The hard checks guarantee non-collision against $\mathcal{R}$ by construction, but provide no guarantee against real identities enrolled after provisioning. To stress-test this, we fix $|\mathcal{V}|$ at one million virtual identities and progressively enroll 180K held-out identities from WebFace4M [62] (disjoint from $\mathcal{R}$), measuring the non-collision rate as $|\mathcal{R}_{\text{test}}|/|\mathcal{R}|$ grows from 0.1 to 0.5 ($|\mathcal{R}|$=360K). In embedding space, Fig. 5 shows flat curves across all α values at τ=0.391, indicating that the collision probability does not grow with test-gallery scale (Appx. C.7); at our default α=4.0, collisions stay near zero while the gallery grows by half. After realization the trend is the same at a lower level: for the re-encoded $\phi(G(s))$, identity-level Non-Collision against the full updated gallery moves from 98.61% to 97.79%, so 0.83% of deployed identities require reassignment while $\mathcal{R}$ grows by half. Provisioning at $\tau_{\text{safe}}=\tau-\Delta$ leaves every identity a cosine margin of at least Δ to the enrolled gallery (Proposition 2), which absorbs any realization or capture shift that raises a gallery similarity by less than Δ. The margin costs capacity: at τ_{safe}=0.360 the bound falls from 7.41×10^{10} to 1.81×10^{9}.

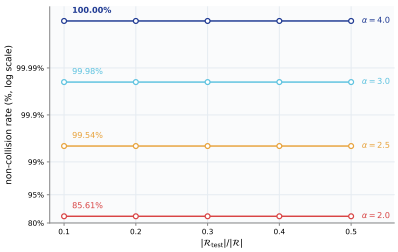

Figure 5: Open-world non-collision. At $\alpha\geq$3.0, collision is near zero in all scales.

Transfer to Other Encoders. The guarantee is encoder-specific (Remark 1). On the same 10K realized identities and the full gallery, two AdaFace IR-101 [16] models never used in provisioning,

Table 3: (a) Five SoTA deepfake detectors [25] applied zero-shot to BIP-generated virtual faces. **(b)** v-LFW evaluation on the combined LFW+v-LFW set. R-V pairs are exclusively impostor pairs by BIP construction. Unified denotes joint recognition and detection on the combined set. "–": task not supported by the method.

Detector	ACC ↑
SPSL [31]	87.76
UCF [56]	93.14
Xception [40]	95.13
EfficientB4 [48]	97.13
PG-FDD [24]	**99.57**

(a)

Metric	DCFace [18]+AdaFace	Vec2Face [54]+ArcFace	BIP+IAPCT (Ours)
R-R Acc ↑	98.58	98.87	**99.80**
V-V Acc ↑	98.85	99.04	**99.18**
R-V FAR ↓	0.05	0.02	**0.01**
Det. AUC ↑	–	–	**98.13**
Unified ↑	–	–	**99.20**

(b)

each calibrated on its own IJB-B ROC at FAR$\approx 2 \times 10^{-5}$, give 99.59% (WebFace4M) and 97.46% (WebFace12M) Non-Collision, against 98.91% for the allocation encoder (Tab. 6).

Cross-Reality Recognition on v-LFW Dataset. A deployed system serving both real humans and digital entities must simultaneously answer two questions for any face presented to it: *who is this?* and *is this a real or virtual identity?* v-LFW is the first benchmark to answer both questions jointly, mirroring LFW in identity count (5,749) and image count (13,233) and enabling four protocols, evaluated alongside the real-image R-R baseline, on the combined LFW+v-LFW set. **R-R** and **V-V** follow the standard LFW verification protocol; **R-V** pairs are exclusively impostor pairs by BIP construction; **Detection** distinguishes real from virtual without identity labels; **Unified** simultaneously produces a verified identity and a real-vs-virtual decision. As shown in Tab. 3b, BIP + IAPCT is the only method supporting all of them. R-R Acc 99.80% confirms that IAPCT leaves recognition performance fully intact. R-V FAR of 0.01% shows that non-collision carries over to image pairs on this benchmark, and Detection AUC 98.13% confirms real and virtual populations are reliably separable without modifying the frozen backbone.

5 Conclusion

We introduced Biometric Identity Provisioning (BIP), the first formal framework for allocating non-colliding biometric identities to digital entities at scale. The key geometric insight is that valid virtual identities must occupy unclaimed gaps within the real face manifold rather than a separate free subspace. This finding shapes repulsion-based allocation, gap-aware realization via GapGen, and the capacity analysis, which leaves about $7{,}400\times$ headroom at the 10M scale under the submanifold model. Against a gallery of 360K real identities, 10M non-colliding virtual identity embeddings are provisioned with no observed collision, and one million are realized as high-fidelity face images, of which 98.61% remain non-colliding after re-encoding. Our v-LFW dataset benchmarks real-virtual coexistence: R-V FAR of 0.01% shows that non-collision carries over to image pairs, and existing deepfake detectors flag the generated faces zero-shot with 87.76–99.57% accuracy.

Limitations. BIP guarantees non-collision against $\mathcal{R}$ at provisioning time; under continued enrollment the identity is long-lived and lifecycle-managed rather than permanent, and the safety buffer (Appx. C.6) trades capacity for margin. The guarantee is exact in embedding space; after realization, separability at scale is bounded by the round-trip fidelity of the generator, which degrades at large α, and applying the hard checks to $\phi(G(s))$ is a direct next step. Capacity bounds rest on a spherical submanifold approximation rather than a precise characterisation of the face manifold (Appx. C). Allocation is specific to the ArcFace ResNet-100 space: two AdaFace encoders retain 97.46–99.59% Non-Collision, and joint multi-encoder provisioning is a natural extension. Robustness to physical display and re-capture remains to be studied, and the demographic composition of provisioned identities is not controlled and may differ from that of the gallery.

Acknowledgments and Disclosure of Funding

This work was supported in part by the National Science Foundation under Grant FMitF-2319242 and by the NVIDIA Academic Grant Program through an award to Feng Liu. The authors declare no competing interests.

References

[1] Xiang An, Xuhan Zhu, Yuan Gao, Yang Xiao, Yongle Zhao, Ziyong Feng, Lan Wu, Bin Qin, Ming Zhang, Debing Zhang, and Ying Fu. Partial FC: Training 10 million identities on a single machine. In *ICCVW*, 2021.

[2] Gwangbin Bae, Martin de La Gorce, Tadas Baltrušaitis, Charlie Hewitt, Dong Chen, Julien Valentin, Roberto Cipolla, and Jingjing Shen. DigiFace-1M: 1 million digital face images for face recognition. In *WACV*, 2023.

[3] Fadi Boutros, Jonas Henry Grebe, Arjan Kuijper, and Naser Damer. IDiff-Face: Synthetic-based face recognition through fizzy identity-conditioned diffusion model. In *ICCV*, 2023.

[4] Fadi Boutros, Marco Huber, Patrick Siebke, Tim Rieber, and Naser Damer. SFace: Privacy-friendly and accurate face recognition using synthetic data. In *IJCB*, 2022.

[5] Cloud Security Alliance. Securing autonomous AI agents. https://cloudsecurityalliance.org/artifacts/securing-autonomous-ai-agents, 2026. Accessed November 5, 2026.

[6] Davide Cozzolino, Andreas Rössler, Justus Thies, Matthias Nießner, and Luisa Verdoliva. ID-Reveal: Identity-aware deepfake video detection. In *ICCV*, 2021.

[7] Hao Dang, Feng Liu, Joel Stehouwer, Xiaoming Liu, and Anil K. Jain. On the detection of digital face manipulation. In *CVPR*, 2020.

[8] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. ArcFace: Additive angular margin loss for deep face recognition. In *CVPR*, 2019.

[9] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *ICLR*, 2021.

[10] Zhihao Gu, Taiping Yao, Yang Chen, Ran Yi, Shouhong Ding, and Lizhuang Ma. Region-aware temporal inconsistency learning for deepfake video detection. In *IJCAI*, 2022.

[11] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. GANs trained by a two time-scale update rule converge to a local Nash equilibrium. In *NeurIPS*, 2017.

[12] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *NeurIPS*, 2020.

[13] Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, 2007.

[14] Jeff Johnson, Matthijs Douze, and Hervé Jégou. Billion-scale similarity search with GPUs. *IEEE Transactions on Big Data*, 7(3):535–547, 2021.

[15] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *CVPR*, 2019.

[16] Minchul Kim, Anil K. Jain, and Xiaoming Liu. AdaFace: Quality adaptive margin for face recognition. In *CVPR*, 2022.

[17] Minchul Kim, Feng Liu, Anil K. Jain, and Xiaoming Liu. Cluster and aggregate: Face recognition with large probe set. In *NeurIPS*, 2022.

[18] Minchul Kim, Feng Liu, Anil K. Jain, and Xiaoming Liu. DCFace: Synthetic face generation with dual condition diffusion model. In *CVPR*, 2023.

[19] Minchul Kim, Yiyang Su, Feng Liu, Anil K. Jain, and Xiaoming Liu. KeyPoint relative position encoding for face recognition. In *CVPR*, 2024.

[20] Minchul Kim, Dingqiang Ye, Yiyang Su, Feng Liu, and Xiaoming Liu. SapiensID: Foundation for human recognition. In *CVPR*, 2025.

[21] Minsoo Kim, Min-Cheol Sagong, Gi Pyo Nam, Junghyun Cho, and Ig-Jae Kim. VIGFace: Virtual identity generation for privacy-free face recognition dataset. In *ICCV*, 2025.

[22] Younghun Kim, Minsuk Jang, Myung-Joon Kwon, Wonjun Lee, and Changick Kim. SELFI: Selective fusion of identity for generalizable deepfake detection. *arXiv preprint arXiv:2506.17592*, 2025.

[23] Stan Z. Li and Anil K. Jain, editors. *Handbook of face recognition*. Springer, London, 2nd edition, 2011.

[24] Li Lin, Xinan He, Yan Ju, Xin Wang, Feng Ding, and Shu Hu. Preserving fairness generalization in deepfake detection. In *CVPR*, 2024.

[25] Li Lin, Santosh, Mingyang Wu, Xin Wang, and Shu Hu. AI-Face: A million-scale demographically annotated AI-generated face dataset and fairness benchmark. In *CVPR*, 2025.

[26] Feng Liu, Ryan Ashbaugh, Nicholas Chimitt, Najmul Hassan, Ali Hassani, Ajay Jaiswal, Minchul Kim, Zhiyuan Mao, Christopher Perry, Zhiyuan Ren, Yiyang Su, Pegah Varghaei, Kai Wang, Xingguang Zhang, Stanley H. Chan, Arun Ross, Humphrey Shi, Zhangyang Wang, Anil K. Jain, and Xiaoming Liu. FarSight: A physics-driven whole-body biometric system at large distance and altitude. In *WACV*, 2024.

[27] Feng Liu, Nicholas Chimitt, Lanqing Guo, Jitesh Jain, Aditya Kane, Minchul Kim, Wes Robbins, Yiyang Su, Dingqiang Ye, Xingguang Zhang, Jie Zhu, Siddharth Satyakam, Christopher Perry, Stanley H. Chan, Arun Ross, Humphrey Shi, Zhangyang Wang, Anil K. Jain, and Xiaoming Liu. Person recognition at altitude and range: Fusion of face, body shape and gait. *arXiv preprint arXiv:2505.04616*, 2025.

[28] Feng Liu, Minchul Kim, ZiAng Gu, Anil K. Jain, and Xiaoming Liu. Learning clothing and pose invariant 3D shape representation for long-term person re-identification. In *ICCV*, 2023.

[29] Feng Liu, Minchul Kim, Anil K. Jain, and Xiaoming Liu. Controllable and guided face synthesis for unconstrained face recognition. In *ECCV*, 2022.

[30] Feng Liu, Minchul Kim, Zhiyuan Ren, and Xiaoming Liu. Distilling CLIP with dual guidance for learning discriminative human body shape representation. In *CVPR*, 2024.

[31] Honggu Liu, Xiaodan Li, Wenbo Zhou, Yuefeng Chen, Yuan He, Hui Xue, Weiming Zhang, and Nenghai Yu. Spatial-phase shallow learning: Rethinking face forgery detection in frequency domain. In *CVPR*, 2021.

[32] Weiyang Liu, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song. SphereFace: Deep hypersphere embedding for face recognition. In *CVPR*, 2017.

[33] Microsoft. What is Microsoft Entra Agent ID? https://learn.microsoft.com/en-us/entra/agent-id/what-is-microsoft-entra-agent-id, 2026. Accessed November 5, 2026.

[34] Hatef Otroshi Shahreza and Sébastien Marcel. HyperFace: Generating synthetic face recognition datasets by exploring face embedding hypersphere. In *ICLR*, 2025.

[35] Samuel Pagon, Yixuan Shen, Vishal Asnani, and Feng Liu. What do deepfake benchmarks measure? An audit using frozen self-supervised representations. *arXiv preprint arXiv:2606.26384*, 2026.

[36] Foivos Paraperas Papantoniou, Alexandros Lattas, Stylianos Moschoglou, Jiankang Deng, Bernhard Kainz, and Stefanos Zafeiriou. Arc2Face: A foundation model for ID-consistent human faces. In *ECCV*, 2024.

[37] Dustin Podell, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Müller, Joe Penna, and Robin Rombach. SDXL: Improving latent diffusion models for high-resolution image synthesis. In *ICLR*, 2024.

[38] Yuyang Qian, Guojun Yin, Lu Sheng, Zixuan Chen, and Jing Shao. Thinking in frequency: Face forgery detection by mining frequency-aware clues. In *ECCV*, 2020.

[39] Haibo Qiu, Baosheng Yu, Dihong Gong, Zhifeng Li, Wei Liu, and Dacheng Tao. SynFace: Face recognition with synthetic data. In *ICCV*, 2021.

[40] Andreas Rössler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, and Matthias Nießner. FaceForensics++: Learning to detect manipulated facial images. In *ICCV*, 2019.

[41] Kaede Shiohara and Toshihiko Yamasaki. Detecting deepfakes with self-blended images. In *CVPR*, 2022.

[42] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *ICLR*, 2021.

[43] Tobin South, Subramanya Nagabhushanaradhya, Ayesha Dissanayaka, Sarah Cecchetti, George Fletcher, Victor Lu, Aldo Pietropaolo, Dean H. Saxe, Jeff Lombardo, Abhishek Maligehalli Shivalingaiah, Stan Bounev, Alex Keisner, Andor Kesselman, Zack Proser, Ginny Fahs, Andrew Bunyea, Ben Moskowitz, Atul Tulshibagwale, Dazza Greenwood, Jiaxin Pei, and Alex Pentland. Identity management for agentic AI: The new frontier of authorization, authentication, and security for an AI agent world. *arXiv preprint arXiv:2510.25819*, 2025.

[44] Yiyang Su, Minchul Kim, Feng Liu, Anil K. Jain, and Xiaoming Liu. Open-set biometrics: Beyond good closed-set models. In *ECCV*, 2024.

[45] Yiyang Su, Minchul Kim, Jie Zhu, Christopher Perry, Feng Liu, Anil K. Jain, and Xiaoming Liu. LocalScore: Local density-aware similarity scoring for biometrics. *arXiv preprint arXiv:2602.01012*, 2026.

[46] Yiyang Su, Yunping Shi, Feng Liu, and Xiaoming Liu. HAMoBE: Hierarchical and adaptive mixture of biometric experts for video-based person ReID. In *ICCV*, 2025.

[47] Yiyang Su, Jie Zhu, Feng Liu, Anil K. Jain, and Xiaoming Liu. SapiensID 2.0: Aligning human recognition foundation models with human perception. In *NeurIPS*, 2026.

[48] Mingxing Tan and Quoc V. Le. EfficientNet: Rethinking model scaling for convolutional neural networks. In *ICML*, 2019.

[49] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *NeurIPS*, 2017.

[50] Hao Wang, Yitong Wang, Zheng Zhou, Xing Ji, Dihong Gong, Jingchao Zhou, Zhifeng Li, and Wei Liu. CosFace: Large margin cosine loss for deep face recognition. In *CVPR*, 2018.

[51] Qixun Wang, Xu Bai, Haofan Wang, Zekui Qin, Anthony Chen, Huaxia Li, Xu Tang, and Yao Hu. InstantID: Zero-shot identity-preserving generation in seconds. *arXiv preprint arXiv:2401.07519*, 2024.

[52] Cameron Whitelam, Emma Taborsky, Austin Blanton, Brianna Maze, Jocelyn Adams, Tim Miller, Nathan Kalka, Anil K. Jain, James A. Duncan, Kristen Allen, Jordan Cheney, and Patrick Grother. IARPA Janus Benchmark-B face dataset. In *CVPRW*, 2017.

[53] Haiyu Wu, Jaskirat Singh, Sicong Tian, Liang Zheng, and Kevin W. Bowyer. Vec2Face+ for face dataset generation. *arXiv preprint arXiv:2507.17192*, 2025.

[54] Haiyu Wu, Jaskirat Singh, Sicong Tian, Liang Zheng, and Kevin W. Bowyer. Vec2Face: Scaling face dataset generation with loosely constrained vectors. In *ICLR*, 2025.

[55] Junhao Xu, Jingjing Chen, Xue Song, Feng Han, Haijun Shan, and Yu-Gang Jiang. Identity-driven multimedia forgery detection via reference assistance. In *ACM MM*, 2024.

[56] Zhiyuan Yan, Yong Zhang, Yanbo Fan, and Baoyuan Wu. UCF: Uncovering common features for generalizable deepfake detection. In *ICCV*, 2023.

[57] Hu Ye, Jun Zhang, Sibo Liu, Xiao Han, and Wei Yang. IP-Adapter: Text compatible image prompt adapter for text-to-image diffusion models. *arXiv preprint arXiv:2308.06721*, 2023.

[58] Dong Yi, Zhen Lei, Shengcai Liao, and Stan Z. Li. Learning face representation from scratch. *arXiv preprint arXiv:1411.7923*, 2014.

[59] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *ICCV*, 2023.

[60] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018.

[61] Ziyuan Zhang, Luan Tran, Feng Liu, and Xiaoming Liu. On learning disentangled representations for gait recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(1):345–360, 2022.

[62] Zheng Zhu, Guan Huang, Jiankang Deng, Yun Ye, Junjie Huang, Xinze Chen, Jiagang Zhu, Tian Yang, Jiwen Lu, Dalong Du, and Jie Zhou. WebFace260M: A benchmark unveiling the power of million-scale deep face recognition. In *CVPR*, 2021.

Appendix

This appendix provides full details supporting the main paper.

- Sec. A: Broader Impact and Ethics.
- Sec. B: Verification threshold selection and IJB-B operating points.
- Sec. C: Capacity bound derivations.
  - Sec. C.1: PCA analysis and submanifold model.
  - Sec. C.2: Spherical geometry preliminaries and GV lower bound.
  - Sec. C.3: Gilbert–Varshamov lower bound.
  - Sec. C.4: Effective capacity, minimum separation threshold (Corollary 1), and sensitivity to τ.
  - Sec. C.5: Repulsion direction analysis.
  - Sec. C.6: Safety buffer (Proposition 2) and lifecycle policy.
  - Sec. C.7: Empirical capacity estimator.
  - Sec. C.8: Capacity, α, and PCA noise.
- Sec. D: Additional implementation Details of G (GapGen).
  - Base pipeline, conditioning, and sampling.
  - Round-trip re-encoding.
  - Gap-aware fine-tuning.
- Sec. E: Verification, Visual Examples, and IAPCT.
  - Sec. E.1: v-LFW protocol and visual examples.
  - Sec. E.2: IAPCT architecture, tokenisation, and training details.
- Sec. F: Additional results.
  - Sec. F.1: t-SNE of $\mathcal{R}$ vs. $\mathcal{V}$.
  - Sec. F.2: Additional image grids.
  - Sec. F.3: Realized capacity.
  - Sec. F.4: Encoder transfer.

A Broader Impact and Ethics

BIP provisions biometric identities for digital entities, not synthetic faces of real people. The non-collision constraint ensures provisioned identities lie outside all enrolled real identity territories, and Sec. 4 confirms they are detectable by existing deepfake detectors without modification. Potential misuse to evade biometric access control is mitigated by the gallery-relative design: an adversary requires access to the full enrollment gallery. All training data are used under their respective licenses, and v-LFW contains no real person's likeness.

B Verification Threshold Selection

The BIP non-collision constraint $\cos(v_j, c_i) < \tau$ and all capacity bounds in Appendix C are stated in terms of a verification threshold τ that must be grounded in the operating point of the deployed recognition system rather than set arbitrarily.

Protocol. We evaluate ArcFace ResNet-100 [8] (Glint360K, `antelopev2/glintr100`, ONNX) on IJB-B [52] under the 1:1 verification protocol across a sweep of FAR operating points from 10^{-5} to 10^{-4}. The ROC curve and operating points are shown in Fig. 6.

Threshold Choice and BIP Implications. The choice of τ directly governs the stringency of the BIP non-collision constraint and the available face manifold capacity:

- **Larger τ (*e.g.*, FAR $= 10^{-5}$, $\tau = 0.448$):** fewer impostor pairs are falsely accepted. The acceptance cap $\{x : \cos(x, c_i) \geq \tau\}$ around each real identity is *smaller*, and the BIP non-collision constraint $\cos(v_j, c_i) < \tau$ is looser.

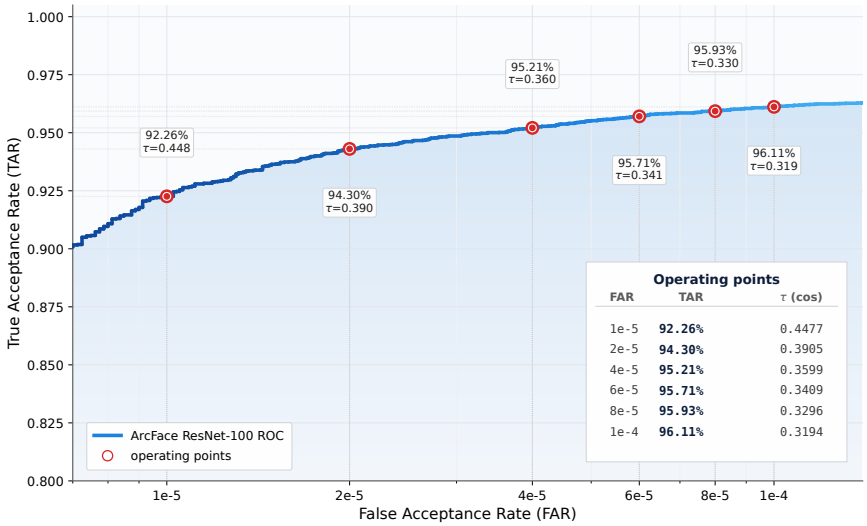

Figure 6: ROC curve for ArcFace ResNet-100 on IJB-B 1:1 verification (8.01M pairs; 10,270 genuine / 8.00M impostor; AUC $= 99.49\%$). Red circles mark the six operating points used as BIP threshold candidates; τ denotes the cosine similarity threshold achieving the stated FAR.

- **Smaller τ (*e.g.*, FAR $= 10^{-4}$, $\tau = 0.319$):** more impostor pairs are accepted. The acceptance cap around each real identity is *larger*, and the BIP non-collision constraint is stricter, requiring more perturbation.

We use $\tau = 0.391$ as the primary operating point in all main experiments, corresponding to FAR $\approx 2 \times 10^{-5}$ and TAR $= 94.30\%$ on IJB-B. At this operating point, one in 5×10^4 impostor pairs is falsely accepted, a conservative threshold appropriate for high-security deployments. We ablate BIP metrics across all six operating points in Tab. 1 of the main paper.

C Capacity Bound Derivation

This section provides full derivations supporting the BIP capacity analysis in Sec. 3.2. Throughout, $\mathbb{S}^{d-1}$ denotes the unit hypersphere in $\mathbb{R}^d$ with $d = 512$, $\tau \in (0, 1)$ is the verification threshold of Assumption 1, and $\tilde{d} = 269$ is the effective manifold dimensionality established in Sec. C.1 below. All reported numerical bounds use the *exact* regularized incomplete beta function; the Gaussian Q-function is used for intuition only.

Notation. We distinguish three related quantities used throughout:

- $A_{\text{pack}}(\tilde{d}, \tau)$: the (unknown) maximum size of a τ-code on the model sphere $\mathbb{S}^{\tilde{d}-1}$ that stands in for $\mathcal{M}_{\text{face}}$ (Sec. C.1).
- $A_{\text{GV}}(\tilde{d}, \tau) := 1/\mu(\tau, \tilde{d})$: the computable Gilbert–Varshamov lower bound, satisfying $A_{\text{pack}} \geq A_{\text{GV}}$.
- $A_{\text{eff}}(\tau) := 1/p_{\text{coll}}(\tau)$: the empirical effective capacity, estimated from observed collision counts between $\mathcal{V}$ and unseen real identities.

Since $\mathcal{M}_{\text{face}}$ is approximated by $\mathbb{S}^{\tilde{d}-1}$ rather than equal to it, A_{GV} is a model-based capacity scale for the face manifold and not a rigorous bound on its packing number. The operational capacity statements of the paper are empirical: allocation at 10M identities (Tab. 1) and realized capacity (Appx. F.3).

C.1 Real Face Manifold: PCA and Submanifold Model

Fig. 7 visualizes the PCA structure of Glint360K identity centroids, confirming Observation 1: 95% of variance concentrates within $\tilde{d}$=269 components out of d=512. We therefore model $\mathcal{M}_{\text{face}}$ as

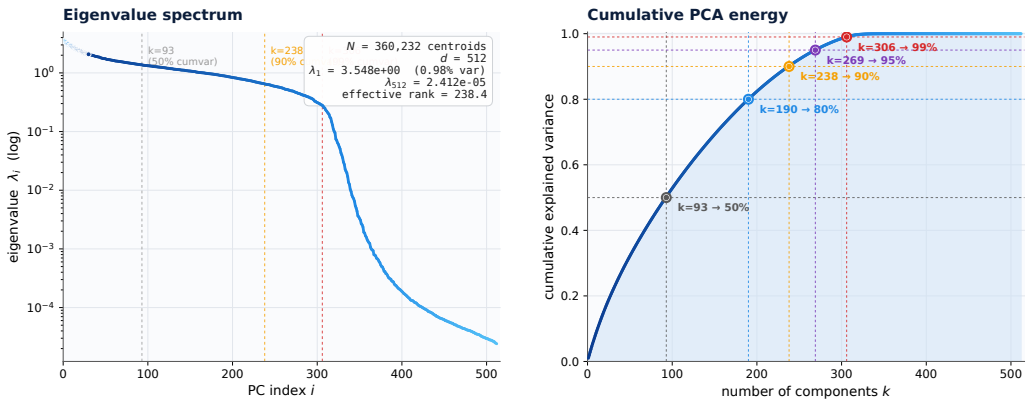

Figure 7: PCA of Glint360K identity centroids (N=360,232, ArcFace ResNet-100, d=512). **Left (eigenvalue spectrum):** λ_k decays sharply after $k\approx300$, with $\lambda_{512} = 2.4\times10^{-5}$ five orders of magnitude below $\lambda_1 = 3.55$; effective rank = 238.4. **Right (cumulative explained variance):** 50% of variance is captured within k=93 components, 90% within k=238, 95% within k=269, and 99% within k=306.

locally approximated by $\mathbb{S}^{\tilde{d}-1}$ and apply all packing bounds within this reduced sphere. This is a *capacity-scale approximation*: the cap angle $\arccos(0.391)\approx67°$ is not a small local neighborhood, so results should be interpreted as headroom estimates rather than precise geometric statements.

C.2 Spherical Geometry Preliminaries

Spherical Cap Volume. The *spherical cap* of cosine radius τ centered at $p \in \mathbb{S}^{d-1}$ is:

$$\mathrm{Cap}(p,\tau) = \big\{x \in \mathbb{S}^{d-1} : \cos(x,p) \geq \tau\big\}. \tag{14}$$

Its normalized volume is independent of p by rotational symmetry:

$$\mu(\tau,d) = \frac{\mathrm{Vol}(\mathrm{Cap}(p,\tau))}{\mathrm{Vol}(\mathbb{S}^{d-1})} = \frac{1}{2}\,I_{1-\tau^2}\!\left(\frac{d-1}{2},\,\frac{1}{2}\right), \tag{15}$$

where $I_x(a,b)$ is the regularized incomplete beta function.

Verification of Eq. (15). For $\mathbb{S}^1$ (circle in $\mathbb{R}^2$) with $\tau = 0.5$: $\mu = \frac{1}{2}I_{0.75}(\frac{1}{2},\frac{1}{2}) = \frac{1}{\pi}\arcsin(\sqrt{0.75}) = \frac{1}{3}$, matching the $120°/360°$ arc fraction for a 60° half-angle cap. ✓

Monotonicity. $\mu(\tau,d)$ is *strictly decreasing* in τ: larger τ defines a smaller cap. $\mu(\tau,d)$ is also strictly decreasing in d by concentration of measure: for $\tilde{d} < d$, $\mu(\tau,\tilde{d}) \geq \mu(\tau,d)$.

Gaussian Approximation (Intuition Only). $\mu(\tau,d) \approx Q(\tau\sqrt{d})$ by the CLT, but in the far-tail regime this substantially overestimates the exact value. At $\tau = 0.391$, $d = 269$: the Gaussian gives $Q(0.391\sqrt{269}) \approx Q(6.41) \approx 7.2\times10^{-11}$ versus the exact $\mu = 1.35\times10^{-11}$ (a 5.3× overestimate). All reported bounds use Eq. (15) directly.

Packing Number. A *spherical τ-code* is a set $\mathcal{S} \subset \mathbb{S}^{d-1}$ with $\cos(s_i,s_j) < \tau$ for all $i \neq j$. Its maximum cardinality is the packing number:

$$A(d,\tau) = \max\big\{|\mathcal{S}| : \mathcal{S} \subset \mathbb{S}^{d-1},\ \cos(s_i,s_j) < \tau\ \forall i \neq j\big\}, \tag{16}$$

which equals the BIP capacity of Definition 1 when $\mathcal{R} = \emptyset$.

C.3 Gilbert–Varshamov Lower Bound

Theorem 1 (Gilbert–Varshamov Bound)**.**

$$A(d,\tau) \;\geq\; \frac{1}{\mu(\tau,d)} = A_{\mathrm{GV}}(d,\tau). \tag{17}$$

Proof. Let $\mathcal{S}^* \subset \mathbb{S}^{d-1}$ be a maximal τ-code, *i.e.*, no additional point can be appended without violating the separation constraint. Maximality implies the caps $\{\mathrm{Cap}(s,\tau)\}_{s\in\mathcal{S}^*}$ cover all of $\mathbb{S}^{d-1}$;

otherwise an uncovered point could be appended, contradicting maximality. Comparing total cap volume to sphere volume:

$$|\mathcal{S}^*| \cdot \mu(\tau, d) \geq 1 \implies |\mathcal{S}^*| \geq \frac{1}{\mu(\tau, d)}. \tag{18}$$

Since $\mathcal{S}^*$ is itself a valid τ-code and $A(d, \tau)$ is the maximum cardinality over all such codes: $A(d, \tau) \geq |\mathcal{S}^*| \geq \frac{1}{\mu(\tau,d)}$. □

Numerical Evaluation at $\tau = 0.391$, $d = 512$.

$$\mu(0.391, 512) \approx 1.74 \times 10^{-20}, \quad A_{\mathrm{GV}}(512, 0.391) \approx 5.75 \times 10^{19} \approx 2^{65.6}. \tag{19}$$

This full-hypersphere bound is not directly useful for BIP since real face identities occupy only a $\tilde{d}$-dimensional submanifold of $\mathbb{S}^{511}$.

C.4 Effective Capacity and GV Lower Bound

Applying Theorem 1 to the $\mathbb{S}^{\tilde{d}-1}$ submanifold model of Sec. C.1 yields the following capacity lower bound.

Proposition 1 (BIP Capacity — GV Lower Bound)**.** *Under the $\mathbb{S}^{\tilde{d}-1}$ submanifold model:*

$$A_{\mathrm{pack}}(\tilde{d}, \tau) \;\geq\; A_{\mathrm{GV}}(\tilde{d}, \tau) = \frac{1}{\mu(\tau, \tilde{d})} = \frac{2}{I_{1-\tau^2}\left(\frac{\tilde{d}-1}{2}, \frac{1}{2}\right)}. \tag{20}$$

For $\tilde{d} = 269$ and $\tau = 0.391$, $\mu(0.391, 269) = 1.35 \times 10^{-11}$, giving:

$$A_{\mathrm{GV}}(269, 0.391) = 7.41 \times 10^{10} \approx 2^{36.1}. \tag{21}$$

Proof. Apply Theorem 1 to dimension $\tilde{d}$ in place of d. □

Corollary 1 (Minimum Separation Threshold)**.** *For $r, z \in \mathbb{S}^{d-1}$ with $p = r \cdot z$ and $|p| < \tau$, let $s = \mathrm{normalize}(r + \alpha z)$. The unique $\alpha^* > 0$ at which $\cos(s, r) = \tau$ exactly is:*

$$\alpha^*(p, \tau) = \frac{\sqrt{1-\tau^2}\left[p\sqrt{1-\tau^2} + \tau\sqrt{1-p^2}\right]}{\tau^2 - p^2}. \tag{22}$$

For $\alpha > 0$ and $|p| < 1$, $\cos(s, r)$ is strictly decreasing in α:

$$\frac{d}{d\alpha}\frac{1+\alpha p}{\sqrt{1+2\alpha p+\alpha^2}} = \frac{\alpha(p^2-1)}{(1+2\alpha p+\alpha^2)^{3/2}} < 0, \tag{23}$$

so $\cos(s, r) < \tau$ requires $\alpha > \alpha^*(p, \tau)$ strictly.

In the orthogonal case $p = 0$, Eq. (22) simplifies to:

$$\alpha^*(0, \tau) = \frac{\sqrt{1-\tau^2}}{\tau}. \tag{24}$$

For the operating thresholds used in this work ($\tau \leq 0.5 < 1/\sqrt{2}$), $\alpha^*(0, \tau)$ is the maximum of $\alpha^*(p, \tau)$ over the repulsive range $p \in [-\tau, 0]$. Therefore, $\alpha > \sqrt{1-\tau^2}/\tau$ is sufficient for $\cos(s, r) < \tau$ over the repulsive range $p \in [-\tau, 0]$.

***Note on positive** p: With PCA-aware noise (Eq. (6)), $p = r \cdot z$ can be positive. For $p > 0$, α^* grows rapidly: $\alpha^*(0.3, 0.391) \approx 9.5$, far exceeding the orthogonal bound 2.35. The hard checks of Eqs. (9)–(10) enforce BIP constraints irrespective of the sign of $r \cdot z$.*

Proof. **Deriving $\alpha^*(p, \tau)$.** From Lemma 1:

$$\cos(s, r) = \frac{1+\alpha p}{\sqrt{1+2\alpha p+\alpha^2}}. \tag{25}$$

Setting $\cos(s, r) = \tau > 0$: the right-hand side equals $\tau > 0$, so the numerator $1 + \alpha p$ must be positive at any solution (since the denominator is always positive). Squaring the equality at the solution is therefore valid:

$$\tau^2(1 + 2\alpha p + \alpha^2) = (1 + \alpha p)^2, \tag{26}$$

which rearranges to the quadratic:

$$\alpha^2(\tau^2 - p^2) + 2\alpha p(\tau^2 - 1) + (\tau^2 - 1) = 0. \tag{27}$$

Since $|p| < \tau$, we have $\tau^2 - p^2 > 0$, so Eq. (27) is a well-posed quadratic. Its discriminant is:

$$\begin{aligned}\Delta &= 4p^2(\tau^2 - 1)^2 - 4(\tau^2 - p^2)(\tau^2 - 1) = 4(\tau^2 - 1)\left[p^2(\tau^2 - 1) - (\tau^2 - p^2)\right] \\ &= 4(\tau^2 - 1)\tau^2(p^2 - 1) = 4\tau^2(1 - \tau^2)(1 - p^2) > 0,\end{aligned} \tag{28}$$

giving $\sqrt{\Delta} = 2\tau\sqrt{(1 - \tau^2)(1 - p^2)}$. The positive root of (27) is:

$$\alpha^*(p, \tau) = \frac{p(1 - \tau^2) + \tau\sqrt{(1 - \tau^2)(1 - p^2)}}{\tau^2 - p^2} = \frac{\sqrt{1 - \tau^2}\left[p\sqrt{1 - \tau^2} + \tau\sqrt{1 - p^2}\right]}{\tau^2 - p^2}, \tag{29}$$

establishing Eq. (22). At $p = 0$: $\alpha^*(0, \tau) = \tau\sqrt{1 - \tau^2}/\tau^2 = \sqrt{1 - \tau^2}/\tau$, confirming Eq. (24).

Strict inequality Eq. (23). Direct differentiation of $f(\alpha) = (1 + \alpha p)/\sqrt{1 + 2\alpha p + \alpha^2}$:

$$f'(\alpha) = \frac{p\sqrt{1 + 2\alpha p + \alpha^2} - (1 + \alpha p) \cdot \dfrac{p + \alpha}{\sqrt{1 + 2\alpha p + \alpha^2}}}{1 + 2\alpha p + \alpha^2} = \frac{\alpha(p^2 - 1)}{(1 + 2\alpha p + \alpha^2)^{3/2}}. \tag{30}$$

Since $\alpha > 0$ and $p^2 < 1$ (as $|p| < \tau < 1$), $f'(\alpha) < 0$, establishing Eq. (23). Since $f(0) = 1 > \tau$ and $\lim_{\alpha\to\infty} f(\alpha) = p < \tau$ (as $|p| < \tau$), strict monotonicity of f implies the solution $\alpha^*(p, \tau)$ is unique.

Upper bound at $\tau \le 1/\sqrt{2}$. By the implicit function theorem applied to $f(\alpha^*(p), p) = \tau$, with $f_\alpha < 0$ (from Eq. (23)) and $f_p = \alpha^2(\alpha + p)/(1 + 2\alpha p + \alpha^2)^{3/2}$:

$$\frac{d\alpha^*}{dp} = -\frac{f_p}{f_\alpha} = \frac{\alpha^*(\alpha^* + p)}{1 - p^2}. \tag{31}$$

The sign of (31) depends on $\alpha^*(p, \tau) + p$. Taking the limit $p \to -\tau^+$ using Eq. (22):

$$\lim_{p\to-\tau^+}(\alpha^*(p, \tau) + p) = \frac{1}{2\tau} - \tau = \frac{1 - 2\tau^2}{2\tau} \ge 0 \quad \iff \quad \tau \le \frac{1}{\sqrt{2}}. \tag{32}$$

At $p = 0$: $\alpha^*(0, \tau) + 0 = \sqrt{1 - \tau^2}/\tau > 0$. Since $\alpha^* > 0$ and $1 - p^2 > 0$, the sign of $d\alpha^*/dp$ is determined by $\alpha^* + p$. Using Eq. (22):

$$\alpha^*(p, \tau) + p = \frac{\sqrt{1 - p^2}\left[p\sqrt{1 - p^2} + \tau\sqrt{1 - \tau^2}\right]}{\tau^2 - p^2}. \tag{33}$$

The denominator $\tau^2 - p^2 > 0$ for $|p| < \tau$, and $\sqrt{1 - p^2} > 0$, so the sign reduces to $p\sqrt{1 - p^2} + \tau\sqrt{1 - \tau^2}$. Define $h(p) = p\sqrt{1 - p^2}$; then $h'(p) = (1 - 2p^2)/\sqrt{1 - p^2} \ge 0$ for $p^2 \le \tau^2 \le 1/2$ (since $\tau \le 1/\sqrt{2}$), so h is non-decreasing on $[-\tau, 0]$ and $h(p) \ge h(-\tau) = -\tau\sqrt{1 - \tau^2}$. Therefore $p\sqrt{1 - p^2} + \tau\sqrt{1 - \tau^2} \ge 0$, giving $\alpha^*(p, \tau) + p \ge 0$ and $d\alpha^*/dp \ge 0$ throughout $p \in [-\tau, 0]$. Hence $\alpha^*(0, \tau) = \sqrt{1 - \tau^2}/\tau$ is the maximum over the repulsive range. □

Sensitivity to τ. Table 4 reports exact GV bounds and the sufficient perturbation threshold. Since $\mu(\tau, \tilde{d})$ is strictly decreasing in τ, larger τ corresponds to a less restrictive non-collision constraint (smaller excluded cap per identity), yielding a larger A_{GV}. The safety buffer requires provisioning at $\tau_{\text{safe}} < \tau$ (stricter constraint), which *reduces* A_{GV} and *increases* the required perturbation; see Proposition 2.

Model-Based Headroom. At $\tau = 0.391$ and $|\mathcal{V}| = 1\text{M}$:

$$\frac{A_{\text{GV}}(269,\ 0.391)}{|\mathcal{V}|} = \frac{7.41 \times 10^{10}}{10^6} \approx 7.41 \times 10^4 \approx 2^{16.2}. \tag{34}$$

Table 4: GV lower bound $A_{\mathrm{GV}} = 1/\mu(\tau, 269)$ at the six IJB-B operating points (Fig. 6), computed via exact regularized incomplete beta. $\alpha^*(0,\tau) = \sqrt{1-\tau^2}/\tau$ is the orthogonal worst-case sufficient perturbation threshold over repulsive directions $p \in [-\tau, 0]$, valid for $\tau \leq 1/\sqrt{2}$. **Bold**: primary operating point ($\tau = 0.391$, FAR$\approx 2 \times 10^{-5}$).

τ	$\mu(\tau, 269)$	$\log_2 A_{\mathrm{GV}}$	$A_{\mathrm{GV}} = 1/\mu$	$\alpha^*(0,\tau)$
0.319	4.21×10^{-8}	24.50	2.38×10^{7}	2.97
0.330	1.40×10^{-8}	26.09	7.15×10^{7}	2.86
0.341	4.45×10^{-9}	27.74	2.25×10^{8}	2.76
0.360	5.52×10^{-10}	30.75	1.81×10^{9}	2.59
0.391	$\mathbf{1.35 \times 10^{-11}}$	**36.11**	$\mathbf{7.41 \times 10^{10}}$	**2.35**
0.448	4.92×10^{-15}	47.53	2.03×10^{14}	2.00

The GV lower bound exceeds one million provisioned identities by a factor of $\approx 74{,}000$. For the 10M embedding-allocation experiment:

$$\frac{A_{\mathrm{GV}}(269,\, 0.391)}{10^7} = \frac{7.41 \times 10^{10}}{10^7} \approx 7.41 \times 10^3. \tag{35}$$

Even the GV lower bound leaves $\approx 7{,}400\times$ headroom at the largest tested allocation scale, indicating that BIP provisioning remains well within the model-based face manifold capacity throughout all reported experiments.

C.5 Repulsion Direction: Heuristic Motivation and Formal Guarantee

Remark 2 in the main text introduces z^* as a repulsion heuristic. We provide the precise mathematical analysis here.

What is proven. With $m = \sum_{k=1}^{K} w_{n_k} c_{n_k}$ and $z^* = -m/\|m\|_2$:

$$z^* \cdot c_{n_k} < 0 \quad \text{for any } c_{n_k} \text{ satisfying } c_{n_k}^\top m > 0. \tag{36}$$

Neighbors positively aligned with the weighted centroid m satisfy this condition; whether *all* neighbors do depends on the local geometry and is not guaranteed solely by the nearest-neighbour construction.

What is not proven after normalization. The relevant quantity for BIP is the rate of change of cosine similarity to a neighbor under the spherically-normalized perturbation $s(\alpha) = \mathrm{normalize}(r + \alpha z^*)$. By the chain rule:

$$\frac{d}{d\alpha} \cos(s(\alpha), c_{n_k})\bigg|_{\alpha=0} = z^* \cdot c_{n_k} - (r \cdot z^*)(r \cdot c_{n_k}). \tag{37}$$

Even when $z^* \cdot c_{n_k} < 0$, the second term $(r \cdot z^*)(r \cdot c_{n_k})$ can dominate: if $r \cdot z^* < 0$ (typical for a repulsion direction) and $r \cdot c_{n_k} > 0$ (neighbor close to r), their product is negative and contributes positively to the derivative. Whether Eq. (37) is negative, *i.e.*, whether moving along z^* genuinely increases cosine distance to c_{n_k} after renormalization — depends on the local geometry and is not guaranteed by the construction of z^* alone.

Correct status. z^* is a *repulsive heuristic*: it improves the probability that a candidate passes the BIP hard checks, but does not formally guarantee it. The formal non-collision guarantee comes exclusively from Eqs. (9)–(10), which accept a candidate s only after exact cosine verification against all of $\mathcal{R}$ and $\mathcal{V}_t$.

C.6 Safety Buffer for Open-World Robustness

Proposition 2 (Safety Buffer)**.** *Let τ be the operating verification threshold (collision when $\cos \geq \tau$) and let $\tau_{\mathrm{safe}} = \tau - \Delta$ with $\Delta > 0$. If $\mathcal{V}$ is provisioned such that $\cos(v_j, c_i) < \tau_{\mathrm{safe}}$ for all $v_j \in \mathcal{V}$, $c_i \in \mathcal{R}$, then:*

(i) Every provisioned identity has a cosine margin of at least Δ below the operating collision boundary τ with respect to all enrolled gallery identities at provisioning time.

(ii) For a subsequently enrolled real identity r', any v_j satisfying $\cos(v_j, r') \in [\tau_{\text{safe}}, \tau)$ does not *cause an operational collision (since $\cos(v_j, r') < \tau$), but lies within the monitoring zone and may trigger reassessment under the provisioning policy. Only $\cos(v_j, r') \geq \tau$ constitutes a collision requiring revocation.*

(iii) The safety buffer requires more *perturbation than provisioning at τ: since $\tau_{\text{safe}} < \tau$ and $d\alpha^*(0,\tau)/d\tau = -1/(\tau^2\sqrt{1-\tau^2}) < 0$, it follows that $\alpha^*(0, \tau_{\text{safe}}) > \alpha^*(0,\tau)$. Concretely, $\tau_{\text{safe}} = 0.360$, $\tau = 0.391$: $\alpha^* \approx 2.59$ versus 2.35.*

Proof. Part (i). The provisioning constraint gives $\cos(v_j, c_i) < \tau_{\text{safe}} = \tau - \Delta < \tau$.

Part (ii). A collision under the operating system occurs iff $\cos(v_j, r') \geq \tau$; if $\cos(v_j, r') < \tau$, no collision occurs regardless of whether $\cos(v_j, r') \geq \tau_{\text{safe}}$.

Part (iii). Monotonicity follows from $d\alpha^*(0,\tau)/d\tau = -1/(\tau^2\sqrt{1-\tau^2}) < 0$, so $\tau_{\text{safe}} < \tau$ implies $\alpha^*(0, \tau_{\text{safe}}) > \alpha^*(0,\tau)$. □

Remark 3. *The safety buffer reduces the available GV capacity bound. Since $\mu(\cdot, \tilde{d})$ is* strictly decreasing *in τ:*

$$\tau_{safe} < \tau \implies \mu(\tau_{safe}, \tilde{d}) \geq \mu(\tau, \tilde{d}) \implies \frac{1}{\mu(\tau_{safe}, \tilde{d})} \leq \frac{1}{\mu(\tau, \tilde{d})}, \tag{38}$$

so $A_{GV}(\tau_{safe}) \leq A_{GV}(\tau)$. The capacity cost depends strongly on the choice of Δ: at τ_{safe}=0.360, τ=0.391 (Δ=0.031, one operating step): $A_{GV} \geq 1.81 \times 10^9 \approx 2^{30.8}$, giving headroom $\approx 1{,}810\times$ over $|\mathcal{V}|$=10^6, a comfortable margin. By contrast, at τ_{safe}=0.319 (Δ=0.072, two operating steps below): $A_{GV} \geq 2.38 \times 10^7 \approx 2^{24.5}$, leaving only $\approx 24\times$ headroom over one million identities, a tight constraint that may limit provisioning at scale. Practitioners should choose Δ based on the expected enrollment growth rate.

Lifecycle Policy. When a collision occurs, the newly enrolled real identity is never modified: r' enters $\mathcal{R}$ unconditionally and correction falls entirely on the virtual side. Each enrollment batch $\Delta\mathcal{R}$ triggers one screening pass over $\mathcal{V}$, which costs $O(|\mathcal{V}| \times |\Delta\mathcal{R}|)$ cosine operations in the exact case and is reducible with approximate nearest-neighbor indexing [14], with three outcomes. Identities whose similarity to every $r' \in \Delta\mathcal{R}$ stays below τ_{safe} are left untouched; those in the monitoring zone $[\tau_{\text{safe}}, \tau)$ of Proposition 2 are queued for reassessment and reassigned where needed; only those reaching τ require immediate revocation. Only the affected identities are reconsidered: the gallery and the remaining virtual population are never re-provisioned. A collision between two real identities admits no such remedy, since neither person can be re-provisioned.

C.7 Empirical Effective Capacity Estimator

Distinction from Theoretical Capacity. $A_{\text{GV}}(\tilde{d}, \tau)$ is a theoretical packing lower bound under the spherical submanifold approximation. The empirical estimator below measures a distinct quantity:

$$A_{\text{eff}}(\tau) := \frac{1}{p_{\text{coll}}(\tau)}, \qquad p_{\text{coll}}(\tau) = \Pr\big[\cos(v_j, r') \geq \tau\big], \tag{39}$$

the inverse empirical collision probability for a real identity r' drawn from $\mathcal{M}_{\text{face}}$. Under the idealized uniform $\mathbb{S}^{\tilde{d}-1}$ model, $p_{\text{coll}} = \mu(\tau, \tilde{d})$, so $A_{\text{eff}} = A_{\text{GV}}$ numerically, but the two are different mathematical objects: A_{GV} lower-bounds the packing number A_{pack} of the model sphere, while A_{eff} characterises the empirical collision density of real identities relative to the provisioned set.

Statistical Model. Let $v_j \in \mathcal{V}$ ($j = 1, \ldots, N$) be provisioned virtual identities and let $r'_l \in \mathcal{R}_{\text{test}}$ ($l = 1, \ldots, L$) be unseen real identities drawn independently from $\mathcal{M}_{\text{face}}$. Define the collision indicator $X_{jl} = \mathbf{1}[\cos(v_j, r'_l) \geq \tau]$, with $\Pr[X_{jl} = 1] = 1/A_{\text{eff}}$ under the uniform distribution assumption. The total collision count $C = \sum_{j=1}^{N} \sum_{l=1}^{L} X_{jl}$ satisfies:

$$C \sim \text{Poisson}\left(\frac{NL}{A_{\text{eff}}}\right), \tag{40}$$

by the Poisson limit theorem for sums of rare independent events, where the individual event probability $p_{\text{coll}} = 1/A_{\text{eff}}$ is small; the expected count $\lambda = NL/A_{\text{eff}}$ need not be much smaller than one.

Maximum Likelihood Estimator. The log-likelihood of model (40) is:

$$\ell(A) = -\frac{NL}{A} + C \log \frac{NL}{A} - \log C! \,. \tag{41}$$

Setting $d\ell/dA = 0$ yields:

$$\hat{A}_{\text{eff}}^{\text{MLE}} = \frac{NL}{C} = \frac{|\mathcal{V}| \times |\mathcal{R}_{\text{test}}|}{C}. \tag{42}$$

Confidence Interval. An exact $(1 - \alpha_{\text{CI}})$ confidence interval for A_{eff} is obtained by inverting the exact Poisson confidence interval for C. With $\chi^2_{\nu,q}$ denoting the q-th quantile of the chi-squared distribution with ν degrees of freedom:

$$\left[\frac{NL}{\frac{1}{2}\chi^2_{2(C+1),\, 1-\alpha_{\text{CI}}/2}}, \; \frac{NL}{\frac{1}{2}\chi^2_{2C,\, \alpha_{\text{CI}}/2}} \right]. \tag{43}$$

This interval maintains the stated coverage probability under the Poisson model regardless of sample size.

Zero-Collision Lower Bound. When $C = 0$, the MLE (42) is undefined. From $P(C = 0) = e^{-NL/A_{\text{eff}}} \geq 0.05$:

$$\hat{A}_{\text{eff}} > \frac{NL}{\ln 20} \approx \frac{NL}{2.996}. \tag{44}$$

For $|\mathcal{V}| = 1\text{M}$, $|\mathcal{R}_{\text{test}}| = 180\text{K}$:

$$NL = 1.8 \times 10^{11}, \qquad \hat{A}_{\text{eff}} > 6.0 \times 10^{10} \approx 2^{35.8}. \tag{45}$$

In this setting, the empirical lower bound lies below $A_{\text{GV}}(269, 0.391) \approx 2^{36.1}$, consistent with the interpretation of Eq. (44) as a confidence bound conditioned on $C = 0$, while A_{GV} is a model-based packing lower bound.

Poisson Approximation Validity. The Poisson approximation requires individual collision probabilities $\mu(\tau, \tilde{d}) \ll 1$, not necessarily $\lambda = NL \cdot \mu \ll 1$. Across all six operating points, the cap probability remains small: the largest value occurs at τ=0.319, where $\mu = 4.21 \times 10^{-8}$; at the primary operating point τ=0.391, $\mu = 1.35 \times 10^{-11}$; and at τ=0.448, $\mu = 4.92 \times 10^{-15}$. These small individual collision probabilities support the Poisson approximation under the rare-event model. The expected total collision count λ varies substantially across operating points. For $|\mathcal{V}|$=1M and $|\mathcal{R}_{\text{test}}|$=180K, giving $NL = 1.8 \times 10^{11}$:

$$\lambda(0.319) = 1.8 \times 10^{11} \times 4.21 \times 10^{-8} \approx 7.58 \times 10^{3}, \tag{46}$$

$$\lambda(0.391) = 1.8 \times 10^{11} \times 1.35 \times 10^{-11} \approx 2.43, \tag{47}$$

$$\lambda(0.448) = 1.8 \times 10^{11} \times 4.92 \times 10^{-15} \approx 8.9 \times 10^{-4}. \tag{48}$$

At τ=0.319, $\lambda \gg 1$, so zero-collision bounds are not meaningful and the MLE Eq. (42) should be used directly. At τ=0.391, $\lambda \approx 2.43$: C=0 with probability $\approx 9\%$, C=1 with $\approx 21\%$, and $C \geq 2$ with $\approx 70\%$; the MLE applies in the majority of cases. At τ=0.448, $\lambda \ll 1$ and C=0 is near-certain ($> 99.9\%$), so the zero-collision bound Eq. (44) will typically apply. We verify stability of $\hat{A}_{\text{eff}}$ as $|\mathcal{R}_{\text{test}}|$ grows from 36K to 180K (Sec. 4); flat curves confirm that estimates are not dominated by local demographic density effects.

Assumption Audit. The estimator (42) rests on two assumptions:

(i) Uniform distribution on $\mathcal{M}_{\text{face}}$. Real identities cluster by demographic factors (ethnicity, age, gender), so r'_l is not strictly uniform over $\mathcal{M}_{\text{face}}$ [45]. If $\mathcal{V}$ resides in a demographically sparse region, $\hat{A}_{\text{eff}}$ is biased upward. The stability check described above detects this.

(ii) Independence of collision events. X_{jl} and $X_{j'l}$ for different j, j' at fixed l are not exactly independent when v_j and $v_{j'}$ are close. Under the Poisson limit theorem, the approximation in (40) holds under weak dependence when $p_{\text{coll}} \ll 1$, which is satisfied here.

C.8 Relationship between Capacity, α, and PCA Noise

The GV bound A_{GV} and the model capacity A_{pack} are determined by $(\tau, \tilde{d})$ alone and are independent of the proposal parameters α and κ. These parameters control how efficiently the algorithm proposes candidates that pass the hard BIP checks.

Define the valid region at step t:

$$\mathcal{V}_t^* = \left\{ v \in \mathbb{S}^{d-1} : \cos(v, c_i) < \tau \ \forall c_i \in \mathcal{R},\ \cos(v, v_j) < \tau \ \forall v_j \in \mathcal{V}_t \right\}, \tag{49}$$

and the acceptance probability at step t:

$$P_{\text{accept}}^{(t)}(\alpha, \kappa) = \Pr\Big[\text{normalize}(r + \alpha z) \in \mathcal{V}_t^*\Big]. \tag{50}$$

Then:

$$\mathbb{E}[|\mathcal{V}|] = \sum_{t=1}^{N_{\text{attempts}}} P_{\text{accept}}^{(t)}(\alpha, \kappa) \ \approx\ \bar{P}_{\text{accept}}(\alpha, \kappa) \cdot N_{\text{attempts}}, \qquad \mathbb{E}[|\mathcal{V}|] \le A_{\text{pack}}, \tag{51}$$

where $\bar{P}_{\text{accept}}$ is the average acceptance probability over the provisioning process.

$\kappa = 0$ **(pure repulsion):** $z = z^*$ provides directional bias away from the neighborhood centroid but yields low diversity across repeated candidates sampled from the same reference r.

$\kappa \to \infty$ **(pure PCA noise):** Under a crude disjoint-cap and uniform-density approximation:

$$P_{\text{accept}}^{(t)} \approx 1 - (M + |\mathcal{V}_t|)\, \mu(\tau, \tilde{d}) = 1 - \frac{M + |\mathcal{V}_t|}{A_{\text{GV}}}. \tag{52}$$

At $\tau = 0.391$ and $|\mathcal{V}| = 10^6$:

$$1 - \frac{360{,}232 + 10^6}{7.41 \times 10^{10}} \approx 99.998\%. \tag{53}$$

For the 10M embedding-allocation experiment:

$$1 - \frac{360{,}232 + 10^7}{7.41 \times 10^{10}} \approx 99.986\%. \tag{54}$$

The acceptance probability remains near unity throughout provisioning at both scales.

PCA-aware weighting by σ_k *biases* perturbations toward high-variance identity directions and suppresses low-variance directions, keeping candidates close to the empirically dominant face-identity subspace. The sum in Eq. (6) runs over all $d = 512$ components, so perturbations are not strictly confined to the top-$\tilde{d}$ subspace, but the contribution from low-variance directions is substantially down-weighted by σ_k, although the residual variance beyond the top-$\tilde{d}$ components is not exactly zero.

D Additional Implementation Details of G (GapGen)

Fig. 8 summarizes the GapGen training pipeline. Below we describe the base pipeline, conditioning, sampling, round-trip re-encoding, and the gap-aware fine-tuning curriculum in turn.

Base pipeline. The SDXL [37] UNet and text encoders are frozen; the two identity-aware components, the IP-Adapter [57] cross-attention projections and the IdentityNet, are jointly fine-tuned.

Conditioning. Because BIP allocates and renders identities in the same ArcFace embedding space, a provisioned $s \in \mathcal{V} \subset \mathbb{R}^{512}$ is fed directly through the pipeline's `image_embeds` input *without* any face image, face detector, or projection MLP on the condition side. We use a fixed canonical 5-point keypoint layout (face scale 0.28 of the canvas, y-offset -5%) as the IdentityNet condition for every identity.

Sampling. All images are generated at 1024×1024 with steps=30, guidance_scale=3.0, controlnet_scale=0.8, ip_adapter_scale=0.8, and the default SDXL scheduler. For virtual step, we use a single fixed prompt (*"candid color portrait photo of a person, natural lighting"*) and negative prompt suppressing CGI/cartoon styles, watermarks, oversmooth skin, and low-resolution artifacts.

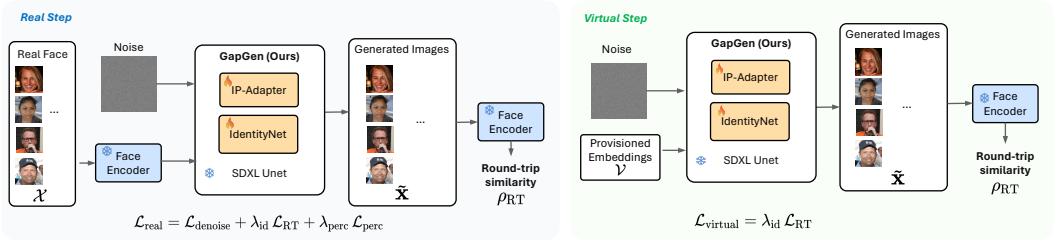

Figure 8: GapGen training pipeline. Real and virtual steps are interleaved at mixing ratio β during fine-tuning. **Real step** (left): real face $x \in \mathcal{X}$ is encoded by the frozen face encoder ϕ to produce target $e^* = \phi(x)$; GapGen renders $\tilde{x} = G(e^*)$, supervised by $\mathcal{L}_{\text{denoise}} + \lambda_{\text{id}}\mathcal{L}_{\text{RT}} + \lambda_{\text{perc}}\mathcal{L}_{\text{perc}}$. **Virtual step** (right): a provisioned embedding $s \in \mathcal{V}$ is fed directly as condition with no paired ground-truth image; supervision relies solely on the round-trip identity loss $\mathcal{L}_{\text{RT}} = 1 - \cos(\phi(\tilde{x}), s)$. IP-Adapter cross-attention projections and IdentityNet are jointly fine-tuned; SDXL UNet and the face encoder remain frozen throughout.

Round-trip re-encoding. We re-extract embeddings from $\tilde{x} = G(s)$ with the same frozen ArcFace pipeline used during allocation: `antelopev2` face detection followed by `glintr100` (IResNet-100). The re-encoded $\tilde{e} = \phi(\tilde{x})$ is directly comparable to the target s via cosine similarity since both lie on the same 512-d unit sphere.

Fine-tuning. We initialize the IP-Adapter from the pretrained InstantID checkpoint and IdentityNet from the pretrained ControlNet [59], then jointly fine-tune both on a gap-aware curriculum (Sec. 3.3). SDXL UNet and text encoders remain frozen throughout.

E Verification, Visual Examples, and IAPCT

E.1 v-LFW Protocol and Visual Examples

Construction. We construct **v-LFW** as a structural counterpart to LFW: it inherits the identity count and the official 10-fold $\times$ 600-pair split, but populates every identity slot with a freshly allocated virtual identity $s_n \in \mathcal{V}$ (distinct from any real LFW person by BIP non-collision), rendered as $\tilde{x}_n = G(s_n)$ at 1024×1024 following the conditioning in App. D.

Protocols. The shared pair list is then reused under the standard real-image baseline (R-R) and the four v-LFW protocols:

- **R-R**: standard LFW verification on original real images.
- **V-V**: same protocol as R-R but on v-LFW images at the matched slot positions, so the same/different structure is preserved over virtual identities.
- **R-V**: one image taken from LFW, the other from v-LFW at the same slot index. Every R-V pair crosses both identity and reality by construction, so its score distribution is a non-mate distribution. We report FAR at the threshold calibrated on R-R same-identity scores at TAR = 95%.
- **Detection**: per-image binary classification of real vs. virtual without identity labels, evaluated on the union of LFW and v-LFW images. We report AUC of the IAPCT forensics score $\hat{p}(x)$ (Sec. E.2).
- **Unified**: joint recognition and detection on the combined LFW + v-LFW pair set. A pair is unified-correct iff verification and the per-image real-vs-virtual decisions are all correct; the formal indicator is given in Eq. (57) below.

Visual examples.

Fig. 9 shows multiple images rendered from the same virtual identity $s \in \mathcal{V}$ under varying pose configurations, lighting conditions, and text style prompts. Each row corresponds to one virtual identity; columns show three independent generations sharing the same conditioning embedding s but differing in keypoint layout, prompt, and sampling seed. Identity is preserved across columns within each row, while pose, expression, lighting, and background vary, supporting the within-identity variation required for v-LFW (Tab. 3b).



Figure 9: v-LFW visual examples. Each row shows three images rendered from a single virtual identity $s \in \mathcal{V}$ at α=4, varying pose, lighting, and style while preserving identity.

E.2 IAPCT: Identity-Anchored Patch Consistency Transformer

Design. IAPCT augments the frozen ArcFace backbone with a transformer head that interrogates whether local patch statistics at each intermediate layer are consistent with the global identity embedding $e = \phi(x)$. The recognition path is preserved verbatim; the forensics path is additive and introduces no overhead at inference beyond the transformer forward pass.

Backbone and intermediate features. We use IResNet-100 with pretrained ArcFace weights. For an aligned input $x \in \mathbb{R}^{3\times112\times112}$, spatial feature maps are tapped at four intermediate stages: $\{F_l\}_{l=1}^{4}$, with spatial resolutions 56×56, 28×28, 14×14, 7×7 and channel widths 64, 128, 256, 512 respectively. $e_5 = \phi(x) \in \mathbb{R}^{512}$ is the standard L2-normalized ArcFace identity output, left unchanged for downstream recognition.

Tokenisation. For each stage l, all spatial positions p of F_l are extracted as patch tokens and projected to d_{model}=256 via a per-stage linear layer:

$$t_{l,p} = W_l\, F_l[:, p] \in \mathbb{R}^{256}, \quad p = 1, \ldots, H_l W_l. \tag{55}$$

The identity anchor is projected as: $e_{\text{proj}} = W_{\text{id}}\, e \in \mathbb{R}^{256}$.

Transformer encoder. A 4-layer transformer encoder processes the patch tokens $\{t_{l,p}\}$ from all four stages jointly. Each layer applies self-attention among spatial tokens followed by cross-attention to the identity anchor e_{proj}, producing patch-identity consistency scores:

$$\gamma_{l,p} = \sigma\left(\frac{(W_Q\, t_{l,p}) \cdot (W_K\, e_{\text{proj}})}{\sqrt{d_k}}\right) \in (0, 1), \tag{56}$$

where σ is the sigmoid function. Real faces yield uniformly high $\gamma_{l,p}$ since every spatial region originates from the same physical person; virtual faces yield heterogeneous $\gamma_{l,p}$ due to unconstrained intermediate features.

Head and loss. A `[CLS]` token produces a forensics logit through:

$$\text{LN}(256) \to \text{Linear}(256{\to}128) \to \text{GELU} \to \text{Dropout}(0.1) \to \text{Linear}(128{\to}1).$$

Training loss: $\mathcal{L} = \text{BCE}(\hat{y}, y) + \lambda_c(H(\gamma^{\text{virtual}}) - H(\gamma^{\text{real}}))$, where $H(\gamma^{(x)}) = -\sum_{l,p} \bar{\gamma}_{l,p}^{(x)} \log \bar{\gamma}_{l,p}^{(x)}$ is the entropy of the normalized attention distribution, y=0 for real and y=1 for virtual, and ϕ remains frozen throughout.

Training. Real samples: Glint360K 112×112 aligned crops. Virtual samples: 100K images at α=4 from GapGen. Batch size 128 with 30% virtual fraction; weighted random sampling. AdamW, lr 10^{-4}, weight decay 10^{-2}, gradient clip 5.

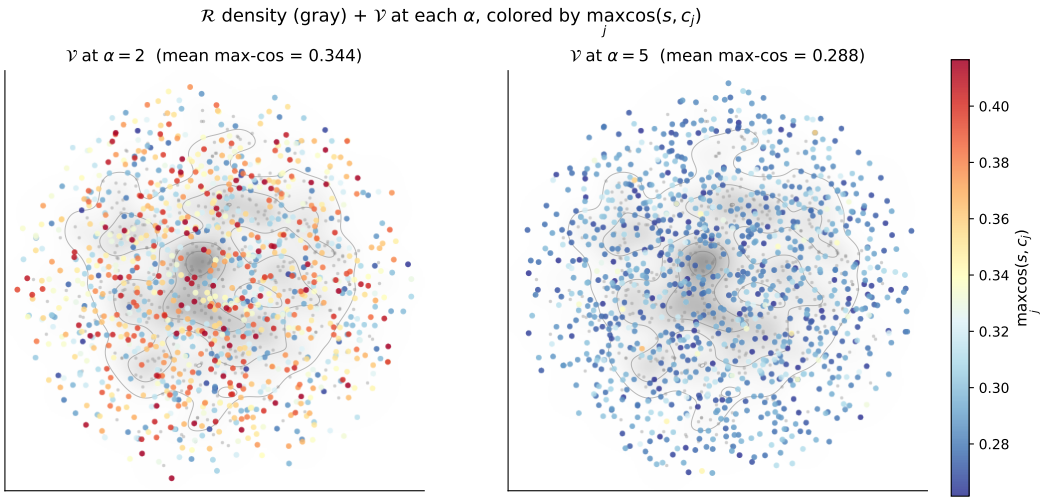

Figure 10: t-SNE of real and virtual identities. Real centroids $\mathcal{R}$ are shown as a gray density background. Each virtual identity in $\mathcal{V}$ is coloured by $\max_{c_j \in \mathcal{R}} \cos(s, c_j)$ (red: close to a real cluster; blue: located in a low-density gap). Increasing α shifts virtual identities from regions dense with real centroids (α=2, mean max-cos 0.344) toward low-density gaps (α=5, mean max-cos 0.288). No internal collapse is observed within $\mathcal{V}$.

Unified metric. A pair (x_a, x_b) is unified-correct iff verification and the two per-image real-vs-virtual decisions are all correct:

$$\mathbf{1}_{\text{uni}} = \mathbf{1}[\hat{y}_{\text{ver}}{=}y_{\text{ver}}] \cdot \mathbf{1}[\hat{y}^a_{\text{rv}}{=}y^a_{\text{rv}}] \cdot \mathbf{1}[\hat{y}^b_{\text{rv}}{=}y^b_{\text{rv}}], \tag{57}$$

with $\hat{y}_{\text{ver}} = \mathbf{1}[\cos(e_5(x_a), e_5(x_b)) \geq \tau_{\text{ver}}]$ at the per-fold LFW threshold and $\hat{y}^\bullet_{\text{rv}} = \mathbf{1}[\hat{p}(x_\bullet) \geq 0.5]$. Unified accuracy averages $\mathbf{1}_{\text{uni}}$ over the union of R-R, V-V, and R-V pairs. The joint indicator (rather than an average over verification and detection accuracies) matches the deployment scenario where a system must simultaneously verify the identity *and* flag the face as real or virtual.

F Additional Results

F.1 t-SNE of $\mathcal{R}$ vs. $\mathcal{V}$

Fig. 10 shows a t-SNE projection of real identity centroids $\mathcal{R}$ (rendered as a gray density background) together with virtual identities $\mathcal{V}$ at $\alpha \in \{2, 5\}$ (1K samples per panel). Each point in $\mathcal{V}$ is coloured by $\max_{c_j \in \mathcal{R}} \cos(s, c_j)$, i.e., its cosine similarity to the nearest real centroid: red indicates virtual identities that lie close to some real cluster, while blue indicates identities that fall into low-density gaps between clusters. Two observations are consistent with Obs. 1. (i) $\mathcal{V}$ interleaves with $\mathcal{R}$ rather than forming a separate cloud: virtual identities appear throughout the manifold, including regions dense with real centroids. (ii) Increasing α from 2 to 5 systematically pushes $\mathcal{V}$ away from real clusters and into the gaps: the mean max-cos decreases from 0.344 (α=2) to 0.288 (α=5), consistent with the non-collision improvement reported in Tab. 1.

F.2 Additional Image Grids

Fig. 11 shows 40 randomly selected faces generated by G from virtual identities $s \in \mathcal{V}$ at α=4. The samples vary visibly in apparent age, hairstyle, and accessories (e.g. glasses) and show no obvious identity duplication across images. This is qualitative evidence that the perturbation-based provisioning scheme, $s = \text{normalize}(r + \alpha z)$, yields distinct identities without collapse; it is not a measurement of demographic composition.

Growing the gallery by half costs 0.82 points, with no evident acceleration. The offset from 100% at 0% growth is a realization effect that is present before any new enrollment (Tab. 2), not an open-world one. Relative to the identities that are valid at deployment, 0.83% require reassignment after the 180K additions.

Table 5: Largest number of identities at which re-encoded Inter-Sep stays above a target (τ=0.391).

Inter-Sep target	BIP+GapGen, $\phi(G(s))$	Glint360K (real)	Ratio
99%	29,830	12,777	2.33×
90%	300,132	141,248	2.12×

Table 6: Encoder transfer on 10K realized identities. τ is the threshold of each encoder at FAR $\approx 2 \times 10^{-5}$ on IJB-B.

Encoder	Training set	τ	Non-Collision (%)	Inter-Sep (%)
ArcFace R100 (allocation encoder)	Glint360K	0.3910	98.91	99.57
AdaFace IR-101 (unseen)	WebFace4M	0.3870	99.59	99.68
AdaFace IR-101 (unseen)	WebFace12M	0.3537	97.46	98.39

F.3 Realized Capacity

Inter-Sep is computed per identity, so it decreases with the number of identities n even at constant pairwise quality. Tab. 5 reports the largest scale at which re-encoded Inter-Sep stays above a target, with real Glint360K identities as a reference under the same encoder, threshold, and protocol.

Under a rare-pair model, $S_n \approx (1-p)^{n-1}$ for a pairwise collision probability p. Inverting Tab. 2 gives $p \approx 2$–4×10^{-7} across a $100\times$ range in n, with no systematic increase: per-pair separability at 1M matches that at 10K, and the decline of Inter-Sep from 99.57% to 78.36% is combinatorial, since each identity must clear $100\times$ more neighbors.

F.4 Encoder Transfer

The guarantee is encoder-specific: Definition 1 conditions BIP on a given ϕ and τ, and a different encoder induces a different geometry. Tab. 6 re-encodes the same 10K realized identities and the full 360K gallery with two AdaFace IR-101 [16] models, neither trained on Glint360K. Each encoder is calibrated on its own IJB-B ROC at the same operating point (FAR $\approx 2 \times 10^{-5}$, Appx. B), since cosine thresholds are not comparable across encoders.

Both AdaFace models match or exceed ArcFace in verification accuracy at this operating point (TAR of 94.37% and 95.06% against 94.30%), so the comparison is not confounded by a weaker recognizer. Non-Collision ranges from 97.46% to 99.59% and Inter-Sep from 98.39% to 99.68%, on either side of the values of the allocation encoder. The worst case is 1.45 points below ArcFace, at the model whose calibrated threshold is lowest, which makes the non-collision constraint the most demanding of the three in its own geometry.



Figure 11: Sample virtual identities from BIP. Forty randomly sampled virtual identities rendered by GapGen at 1024×1024.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction state the four contributions (formalization, geometry and allocation, GapGen realization, v-LFW benchmark) and their scope. Theoretical claims are supported by proofs in Appendix C; empirical claims are supported by Tabs. 1–3, Figs. 4–5, and Appendix F.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Limitations are discussed in the Limitations paragraph of Sec. 5: the gallery-relative guarantee and revocation under continued enrollment, the gap between allocation and realization, the spherical submanifold approximation, allocation in a single encoder space with transfer measured on two other encoders, robustness to physical capture as an open question, and the uncontrolled demographic composition of provisioned identities.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: All assumptions are stated explicitly (Assumption 1). Theorem 1, Proposition 1, Corollary 1, and Proposition 2 include complete proofs in Appendix C. Lemma 1 follows by direct computation from the definition of s.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: Full implementation details are provided in Sec. 3, Sec. 4 and the Appendix, including encoder, gallery, threshold, hyperparameters (α, κ, β, K, t), and generation settings. The analyses after realization (realized capacity, encoder transfer) are specified in Appendix F. Vec2Face baselines are reproduced from the authors' released code.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [No]

   Justification: The GapGen model, the BIP allocation and evaluation code, the provisioned virtual identity embeddings, the generated virtual-face data, and the v-LFW dataset will be released through the project page (https://vilab-group.com/project/bip). All training data (Glint360K [1], FFHQ [15]) are publicly available. The base generator builds on InstantID [51], which is publicly released.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: Encoder, gallery, threshold, and all hyperparameters are specified in Sec. 3, Sec. 4 and the Appendix. GapGen loss weights and mixing ratio are in Sec. 4, its curriculum in Sec. 3.3, and its conditioning and sampling settings in Appendix D. IAPCT architecture and training are in Appendix E.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [No]

Justification: Non-Collision and Inter-Sep are deterministic given the provisioned set and threshold; there is no randomness across runs for these metrics. FID and the RT pass rate are computed over the full provisioned set. Error bars are not reported as the provisioning process is deterministic conditioned on the random seed, and running multiple full-scale (1M-identity) experiments is computationally prohibitive.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: All experiments use NVIDIA A100 GPUs. Embedding-space allocation of 1M identities requires approximately 1 GPU-hour. GapGen fine-tuning requires approximately 48 GPU-hours.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics https://neurips.cc/public/EthicsGuidelines?

   Answer: [Yes]

   Justification: The paper has been reviewed against the NeurIPS Code of Ethics. All training data are used under their respective licenses. v-LFW contains no real person's likeness. Broader impact and potential misuse are discussed in Appendix A.

   Guidelines:

- The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: This work can support safe biometric identity infrastructure for digital entities. A potential risk is misuse for identity spoofing, which is mitigated by our gallery-relative provisioning design and by the deepfake detectability of virtual faces reported in the paragraph *Deepfake Detectability of Virtual Faces* in Sec. 4 and further discussed in Appendix A.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [Yes]

    Justification: The paragraph *Deepfake Detectability of Virtual Faces* in Sec. 4 demonstrates that provisioned virtual faces are detectable by existing SoTA deepfake detectors zero-shot. The gallery-relative design requires access to the full enrollment gallery to provision non-colliding identities, limiting misuse. v-LFW will be released with usage restrictions.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.

- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: All assets are cited. Glint360K [1] is used under its research license. FFHQ [15] is used under CC-BY-NC-SA 4.0. InstantID [51] is used under Apache 2.0. ArcFace [8] model weights are from the publicly released `antelopev2` package. The other assets (LFW, IJB-B, WebFace4M, CASIA-WebFace, SDXL, the AdaFace models, the baseline generators DCFace, Arc2Face and Vec2Face, and the deepfake detectors) are cited where they are used and were obtained under the terms set by their providers.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: v-LFW is introduced as a new benchmark. Construction details are in Appendix E.1. v-LFW contains only BIP-provisioned virtual identities with no real person's likeness. The dataset will be released with documentation and usage guidelines upon publication.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

Justification: The paper does not involve crowdsourcing or research with human subjects. All evaluation is performed on existing datasets or algorithmically generated virtual identities.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A]

Justification: No human subjects are involved. All face images used for training are from publicly available datasets with appropriate research licenses.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: LLMs are not used as a component of the core methodology. LLMs were used only for writing assistance and do not affect the scientific contributions, methodology, or results.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.