FACEFLOW: LEARNING INTRA-CLASS VARIATION VIA FLOW MATCHING IN THE IDENTITY EMBEDDING SPACE FOR SYNTHETIC FACE GENERATION

Anonymous authors
Paper under double-blind review

ABSTRACT

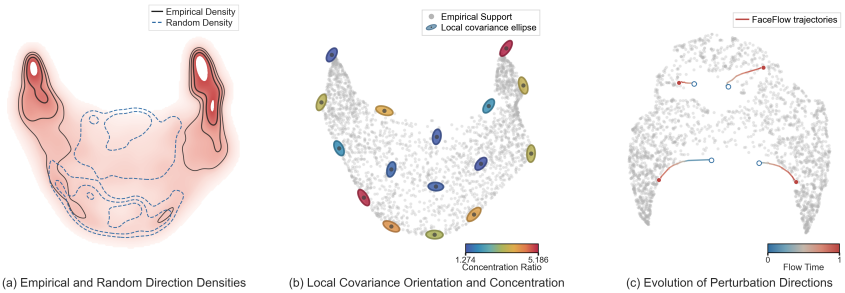
Generating synthetic data for face recognition (FR) training offers a practical alternative to authentic face datasets and supports privacy protection. Existing identity-conditioned generative models can synthesize photorealistic faces with high identity consistency, yet they typically exhibit limited intra-class diversity. Perturbation of the identity condition has been introduced to alleviate this limitation, but existing methods constrain only the perturbation magnitude and sample the perturbation direction uniformly at random in the identity embedding space, without modeling the intra-class variation of real faces. We propose FaceFlow, a flow-based model trained on same-identity pairs to learn an identity-conditioned flow whose endpoint follows the high-dimensional intra-class distribution. FaceFlow guides the perturbation of identity embeddings, which are then decoded by a pre-trained generator to construct synthetic face datasets. FR models trained on the resulting synthetic datasets outperform those trained on datasets constructed using state-of-the-art synthetic-data methods on standard benchmarks and surpass 94% average accuracy at a training scale of 0.5M synthetic images. These results show that FaceFlow enables the reliable generation of diverse intra-class samples while maintaining high identity consistency. The FaceFlow code and generated data will be publicly available at `https://anonymous.4open.science/r/FaceFlow-63EF/`.

1 INTRODUCTION

Face recognition (FR) has achieved remarkable progress in recent years, driven by advances in deep network architectures, by training objectives based on angular margins, and by the availability of large-scale annotated face datasets (Deng et al., 2019; Wang et al., 2018; Boutros et al., 2022a; Yi et al., 2014). Acquiring diverse and legally compliant face data, however, remains a significant challenge. Privacy concerns and regulatory constraints on biometric data have led to the withdrawal of major datasets such as MS-Celeb-1M (Guo et al., 2016) and VGGFace2 (Cao et al., 2018), severely limiting access to authentic training data.

These limitations have spurred growing interest in synthetic data as a scalable and privacy-compliant alternative for training FR models (Qiu et al., 2021; Bae et al., 2023). Recent identity-conditioned diffusion models (DMs) synthesize photorealistic face images that adhere closely to the conditioning identity (Boutros et al., 2023a; Kim et al., 2023; Papantoniou et al., 2024; Lin et al., 2025), and the downstream accuracy of FR models trained on their data has improved steadily. For these models the bottleneck is no longer identity consistency but intra-class diversity: under a fixed identity condition, the generated samples concentrate in a narrow appearance neighborhood and lack the intra-class variation of authentic data. One way to enlarge this diversity without modifying the generator is to perturb the identity embedding within a constrained angular region of the face embedding hypersphere (Boutros et al., 2026), where the angular distance to the reference identity embedding controls the perturbation magnitude. Angular perturbation alone already yields a substantial gain in intra-class diversity. This formulation, however, parameterizes only the perturbation magnitude: the perturbation direction is sampled uniformly at random, without modeling how the identity being perturbed varies. Figure 1 makes this explicit at a fixed angle. In this UMAP (McInnes et al., 2018) projection, the directions of real same-identity pairs appear clustered in a few elongated

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



(a) Empirical and Random Direction Densities (b) Local Covariance Orientation and Concentration (c) Evolution of Perturbation Directions

Figure 1: **UMAP projection of perturbation directions at a fixed angle.** **(a)** Density of the directions of real same-identity pairs (solid) and of random perturbation directions (dashed). **(b)** Local covariance ellipses of the real directions, colored by concentration ratio. **(c)** Trajectories of the learned flow on an unseen identity, from $t{=}0$ to $t{=}1$.

regions, whereas most directions sampled uniformly at random fall outside them. At the same angle, the perturbation is therefore spent on variation that real identities do not exhibit. This clustering further indicates that intra-class variation is multimodal, so in general neither the leading axes of a single intra-class covariance shared by all identities nor random directions capture the distribution of intra-class directions conditioned on the reference identity embedding.

In this work, we propose FaceFlow, which learns an identity-conditioned flow in the identity embedding space from real same-identity pairs. FaceFlow retains the constrained angular region and replaces the uniformly sampled direction with the endpoint of this flow. The flow is driven by a velocity field trained with Riemannian flow matching (Lipman et al., 2022; Chen & Lipman, 2024); it starts from a random direction at $t{=}0$ and, at $t{=}1$, reaches a direction along which the given identity varies (Fig. 1c). The perturbed identity embedding then serves as the identity condition of a frozen pre-trained generator, and Fig. 2 shows faces decoded along the learned flow. Unlike previous methods that rely on auxiliary labels (Xu et al., 2024), learned style modules (Kim et al., 2023), or modifications of the denoising process (Lin et al., 2025), FaceFlow leaves the generator untouched and changes a single degree of freedom of the generation protocol, namely the flow that fixes the perturbation direction. Extensive experiments show that, because this direction comes from a flow conditioned on the identity rather than from a random draw, FaceFlow improves identity consistency and intra-class variation together. As a result, FR models trained on FaceFlow-generated data achieve the highest average verification accuracy among synthetic-data methods at both the 0.5M and 1.0M scales.

The contributions of this work are threefold:

- We propose FaceFlow, a flow-based model that captures the intra-class variation of real faces in the identity embedding space. Trained with same-identity pairs, FaceFlow learns a conditional velocity field that preserves rich intra-class diversity while maintaining high identity separability.
- We demonstrate that FaceFlow enables scalable face synthesis for recognition training, achieving state-of-the-art performance at both the 0.5M and 1.0M scales and surpassing existing methods, particularly those based on random perturbation. Notably, models trained solely on FaceFlow-generated data surpass 94% average accuracy across five benchmarks at the 0.5M scale.
- We provide comprehensive quantitative and qualitative analyses in both image space and the identity embedding space, demonstrating that FaceFlow captures the intra-class variation of real faces beyond simple perturbation amplification. FaceFlow is also robust to its only hyperparameter, the cosine lower bound, and outperforms random perturbation at every tested value.

2 Related Work

Synthetic Data for Face Recognition. Synthetic face datasets offer a privacy-compliant alternative to authentic training data, and their usefulness for training FR models depends on combining identity fidelity with intra-class diversity (Qiu et al., 2021; Bae et al., 2023). Early pipelines based on generative adversarial networks (GANs), including SynFace (Qiu et al., 2021), SFace (Boutros et al.,

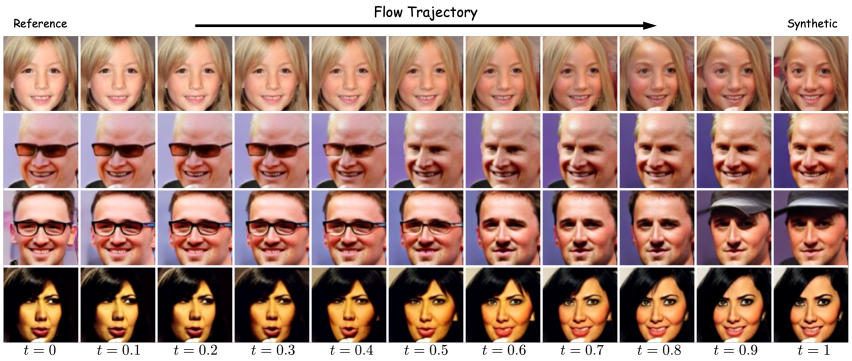


Figure 2: **Faces decoded along the learned flow.** The flow time t runs from 0 to 1: at t=0 the perturbation direction is random, and at t=1 it is the direction produced by the learned flow. Each row follows one identity along the flow trajectory. Pose, expression, and accessories vary, while the reference identity remains visually recognizable.

2022b), USynthFace (Boutros et al., 2023c), and ExFaceGAN (Boutros et al., 2023b), established this direction but leave a large accuracy gap to authentic data. Identity-conditioned DMs such as IDiff-Face (Boutros et al., 2023a), DCFace (Kim et al., 2023), Arc2Face (Papantoniou et al., 2024), CemiFace (Sun et al., 2024), and UIFace (Lin et al., 2025) generate photorealistic samples that adhere closely to the conditioning identity and achieve higher downstream accuracy, which has made pre-trained identity-conditioned DMs the prevailing basis for synthetic FR data. For these models, identity consistency is largely attained, and the remaining bottleneck is the limited intra-class variation of the samples generated under a fixed identity condition.

Intra-Class Diversity in Identity-Conditioned DMs. Approaches that enlarge this variation fall into two groups. Methods of the first group modify the generator, either by introducing additional conditions or by changing the sampling process: a style bank in DCFace (Kim et al., 2023), binary attribute conditions in ID^3 (Xu et al., 2024), two-phase sampling in UIFace (Lin et al., 2025), and contextual dropout in IDiff-Face (Boutros et al., 2023a). They control appearance explicitly, but require auxiliary supervision, architectural changes, or retraining of the generator. Methods of the second group construct identities or intra-class variation in the identity embedding space. Vec2Face (Wu et al., 2025) trains a generator that decodes sampled identity vectors into face images and obtains intra-class samples by perturbing these vectors, and HyperFace (Otroshi Shahreza & Marcel, 2025) optimizes the placement of synthetic identity embeddings on the face embedding hypersphere. IDPerturb (Boutros et al., 2026) keeps a pre-trained generator unchanged and perturbs a reference identity embedding within a constrained angular region. It exposes the degree of deviation from the reference identity as an explicit parameter, and this angular perturbation alone already yields a substantial gain in intra-class diversity. Its formulation, however, parameterizes only the magnitude of the perturbation: the perturbation direction is sampled at random in the identity embedding space and is therefore unrelated to the directions along which images of a real identity vary. As the angular range widens, this mismatch reduces identity consistency and produces samples that fail face detection or identity verification.

Structure of Intra-Class Variation. The assumption that intra-class variation has a shared, transferable structure underlies probabilistic linear discriminant analysis (PLDA) (Prince & Elder, 2007), probabilistic face embeddings (Shi & Jain, 2019), and feature transfer (Yin et al., 2019), all of which describe this structure with a covariance or a Gaussian. Sub-center analysis shows, however, that the samples of one identity form several clusters on the face embedding hypersphere (Deng et al., 2020), which a single covariance or Gaussian cannot represent. Capturing this structure therefore calls for a model that represents the multimodal conditional distribution of perturbation directions on the unit tangent hypersphere and transfers it to identities unseen during training. FaceFlow is designed around this learnable directional structure of intra-class variation in the identity embedding space, and learns it with Riemannian flow matching (Lipman et al., 2022; Chen & Lipman, 2024).

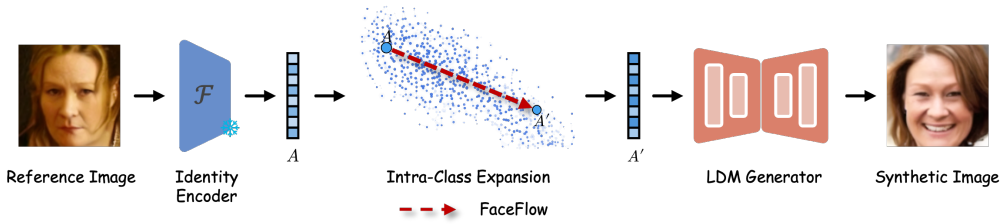


Figure 3: **FaceFlow learns an identity-conditioned flow in the identity embedding space.** The identity encoder $\mathcal{F}$ maps a reference image to $\mathbf{A}$. The learned flow (red dashed arrow) moves a random direction $\mathbf{u}_0$ on $\mathcal{S}_\mathbf{A}$ to $\mathbf{u}_K$, which together with the perturbation angle θ gives the perturbed identity embedding $\mathbf{A}'$. The frozen LDM generator decodes $\mathbf{A}'$ into a synthetic face.

3 Method

3.1 Problem Formulation and Overall Pipeline

FaceFlow represents a perturbed identity embedding by its angular displacement from the reference identity embedding (Boutros et al., 2026): an angle sets how far the embedding moves, and a unit direction sets where it moves. The angle is sampled, and the direction is produced by a learned flow. A frozen encoder $\mathcal{F}$ maps a reference image x_{ref} to a reference identity embedding on the face embedding hypersphere. Displacements that remain on this hypersphere are tangent to it at the reference identity embedding, and the unit vectors in that tangent space form the *unit tangent hypersphere* at the reference identity embedding. The reference identity embedding, its tangent space, and the unit tangent hypersphere are

$$\mathbf{A} = \frac{\mathcal{F}(x_{\text{ref}})}{\|\mathcal{F}(x_{\text{ref}})\|_2} \in \mathbb{S}^{d-1}, \qquad T_\mathbf{A}\mathbb{S}^{d-1} = \{\mathbf{v} \in \mathbb{R}^d : \mathbf{v}^\top \mathbf{A} = 0\},$$
$$\mathcal{S}_\mathbf{A} = \{\mathbf{u} \in T_\mathbf{A}\mathbb{S}^{d-1} : \|\mathbf{u}\|_2 = 1\} \cong \mathbb{S}^{d-2}. \tag{1}$$

An embedding $\mathbf{A}'$ at geodesic distance θ from $\mathbf{A}$, which we call the *perturbation angle*, then decomposes uniquely into a component along the reference identity embedding and a direction $\mathbf{u} \in \mathcal{S}_\mathbf{A}$:

$$\mathbf{A}' = \cos\theta\, \mathbf{A} + \sin\theta\, \mathbf{u}, \qquad \mathbf{u} \in \mathcal{S}_\mathbf{A}, \qquad 0 < \theta < \tfrac{\pi}{2}, \qquad \|\mathbf{A}'\|_2 = 1, \qquad \mathbf{A}^\top \mathbf{A}' = \cos\theta. \tag{2}$$

This decomposition separates geodesic distance from direction without assuming that $\mathbf{A}$ is the class center. FaceFlow learns the conditional distribution of $\mathbf{u}$ from same-identity pairs, so that this direction follows the high-dimensional intra-class distribution. A velocity field on $\mathcal{S}_\mathbf{A}$ moves an initial direction $\mathbf{u}_0 \sim \text{Uniform}(\mathcal{S}_\mathbf{A})$ toward the distribution of intra-class directions, shown as the red dashed curve in Fig. 3. The uniform distribution on the unit tangent hypersphere is obtained by projecting a standard Gaussian sample onto the tangent space and normalizing it:

$$\mathbf{u}_0 = \frac{\boldsymbol{\epsilon} - (\boldsymbol{\epsilon}^\top \mathbf{A})\mathbf{A}}{\left\|\boldsymbol{\epsilon} - (\boldsymbol{\epsilon}^\top \mathbf{A})\mathbf{A}\right\|_2}, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \tag{3}$$

The target distribution $q_\phi(\mathbf{u} \mid \mathbf{A}, \theta)$ depends on the geodesic distance, so distance and direction are coupled through the learned velocity field. A pre-trained generator $\mathcal{G}_\psi$, kept fixed while this field is learned, maps the resulting embedding $\mathbf{A}'$ to a synthetic face.

3.2 Training: Learning the Velocity Field from Same-Identity Pairs

Supervision. FaceFlow learns the velocity field from real same-identity pairs. Let $\mathbf{a}$ and $\mathbf{b}$ be the normalized encoder features of two images of the same identity. Their cosine similarity sets the perturbation angle, and the component of $\mathbf{b}$ orthogonal to $\mathbf{a}$ sets the target direction:

$$s = \mathbf{a}^\top \mathbf{b}, \qquad \theta = \arccos s, \qquad \mathbf{u}_1 = \frac{\mathbf{b} - s\mathbf{a}}{\|\mathbf{b} - s\mathbf{a}\|_2} \in \mathcal{S}_\mathbf{a}. \tag{4}$$

Each pair supplies one training sample: θ as a condition and $\mathbf{u}_1$ as the flow target.

Flow matching on the unit tangent hypersphere. The remaining task is to learn a flow that carries a random initial direction onto the conditional distribution of the observed directions $\mathbf{u}_1$ on $\mathcal{S}_{\mathbf{a}} \cong \mathbb{S}^{d-2}$.

Flow matching regresses the velocity along a prescribed path, and the minimizer of this regression objective generates the target marginal (Lipman et al., 2022). On the unit tangent hypersphere the path admits a closed-form geodesic, so the training target is obtained without simulation or divergence estimation (Chen & Lipman, 2024). The conditional path is the geodesic

$$\mathbf{u}_t = \exp_{\mathbf{u}_1}\big(\kappa(t)\,\log_{\mathbf{u}_1}(\mathbf{u}_0)\big), \qquad \kappa(t) = 1 - t, \tag{5}$$

and the regression target is its velocity $\dot{\mathbf{u}}_t$. With geodesic distance as the premetric, Eq. (5) is spherical geodesic interpolation (Slerp), and $\kappa(t) = 1 - t$ makes the path constant-speed: the norm of $\dot{\mathbf{u}}_t$ equals the geodesic distance Ω at every t. This is the kinetic-energy-minimizing parameterization between the endpoints. With $\mathbf{u}_0$ from Eq. (3), $t \sim \mathcal{U}[0,1]$, and $\Omega = \arccos(\mathbf{u}_0^\top \mathbf{u}_1)$, Slerp and its velocity are

$$\begin{aligned} \mathbf{u}_t &= \frac{\sin((1-t)\Omega)}{\sin\Omega}\,\mathbf{u}_0 + \frac{\sin(t\Omega)}{\sin\Omega}\,\mathbf{u}_1, \\ \dot{\mathbf{u}}_t &= \frac{\Omega}{\sin\Omega}\big[-\cos((1-t)\Omega)\,\mathbf{u}_0 + \cos(t\Omega)\,\mathbf{u}_1\big]. \end{aligned} \tag{6}$$

The geodesic remains in $\mathcal{S}_{\mathbf{a}}$ because both endpoints are orthogonal to $\mathbf{a}$. A projection then restricts the predicted velocity to the tangent space at $\mathbf{u}_t$:

$$\mathbf{P}_{\mathbf{a},\mathbf{u}_t} = \mathbf{I} - \mathbf{a}\mathbf{a}^\top - \mathbf{u}_t\mathbf{u}_t^\top, \qquad \mathbf{v}_\phi = \mathbf{P}_{\mathbf{a},\mathbf{u}_t}\mathbf{f}_\phi(\mathbf{u}_t, t \mid \mathbf{a}, \theta). \tag{7}$$

The training objective is Riemannian flow matching on the unit tangent hypersphere (Chen & Lipman, 2024):

$$\mathcal{L}_{\mathrm{FM}}(\phi) = \mathbb{E}_{(\mathbf{a},\mathbf{b}),\,\mathbf{u}_0,\,t}\big[\|\mathbf{v}_\phi - \dot{\mathbf{u}}_t\|_2^2\big]. \tag{8}$$

Its minimizer is the Bayes-optimal velocity $\mathbf{v}^*{=}\mathbb{E}[\dot{\mathbf{U}}_t \mid t, \mathbf{U}_t, \mathbf{A}, \theta]$; the loss decomposes into approximation error plus a ϕ-independent constant. Training is simulation-free: $\mathbf{u}_t$ and $\dot{\mathbf{u}}_t$ are computed in closed form, without solving an ordinary differential equation (ODE) or calling the generator.

Network. The velocity field is $\mathbf{f}_\phi(\mathbf{u}, t \mid \mathbf{A}, \theta)$, evaluated at $\mathbf{A} = \mathbf{a}$ during training. The reference identity embedding $\mathbf{A}$ determines the identity-specific intra-class geometry, and the perturbation angle θ controls the perturbation magnitude and hence how much identity consistency is preserved. The flow time t and the normalized angle $\theta/\theta_{\max}$ are encoded by Fourier features and added to a linear projection of $\mathbf{A}$. This Fourier encoding is part of the model definition, since it determines how the two scalar conditions enter the velocity field. Zero initialization gives $\mathbf{f}_{\phi_0} \equiv \mathbf{0}$, so $\mathbf{u}_0$ is left unchanged and training learns only an identity-conditioned correction. The generator is held fixed, and the magnitude of θ is set by the generation protocol. Algorithm 1 summarizes the training procedure.

Algorithm 1 FaceFlow training.

Require: Same-identity pairs $\mathcal{D}$, encoder $\mathcal{F}$
Ensure: Parameters ϕ, with an exponential moving average (EMA) $\bar{\phi}$ used at inference
for each update step **do**
  Sample (x_i, x_j) from $\mathcal{D}$
  $\mathbf{a} \leftarrow \mathcal{F}(x_i)/\|\mathcal{F}(x_i)\|$, $\mathbf{b} \leftarrow \mathcal{F}(x_j)/\|\mathcal{F}(x_j)\|$
  $s \leftarrow \mathbf{a}^\top\mathbf{b}$, $\theta \leftarrow \arccos s$, $\mathbf{u}_1 \leftarrow (\mathbf{b} - s\mathbf{a})/\|\mathbf{b} - s\mathbf{a}\|$
  $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, $t \sim \mathcal{U}[0,1]$, $\mathbf{u}_0 \leftarrow \mathrm{Proj}_{\mathcal{S}_{\mathbf{a}}}(\boldsymbol{\epsilon})$
  $\Omega \leftarrow \arccos(\mathbf{u}_0^\top\mathbf{u}_1)$, $(\mathbf{u}_t, \dot{\mathbf{u}}_t)$ via Slerp
  $\mathbf{v}_\phi \leftarrow \mathbf{P}_{\mathbf{a},\mathbf{u}_t}\,\mathbf{f}_\phi(\mathbf{u}_t, t \mid \mathbf{a}, \theta)$
  Update ϕ on $\|\mathbf{v}_\phi - \dot{\mathbf{u}}_t\|_2^2$; update EMA $\bar{\phi}$
end for

3.3 Inference: Integrating the Flow and Generating Faces

At inference, real image pairs are no longer required. Given the reference identity embedding $\mathbf{A}$ of an unseen identity and a perturbation angle θ sampled according to the protocol of Section 4.1, a random initial direction $\mathbf{u}_0$ is drawn as in Eq. (3). The velocity field then moves $\mathbf{u}_0$ on the unit tangent hypersphere toward the learned conditional direction distribution by integrating the ODE that defines the flow

$$\frac{d\mathbf{u}(t)}{dt} = \mathbf{P}_{\mathbf{A},\mathbf{u}(t)}\mathbf{f}_{\bar{\phi}}\big(\mathbf{u}(t), t \mid \mathbf{A}, \theta\big), \qquad \mathbf{u}(0) = \mathbf{u}_0. \tag{9}$$

Because the unit tangent hypersphere is curved, Euler steps leave the sphere. We discretize with the exponential map, using step size $h = 1/K$ and $t_k = kh$:

$$\mathbf{u}_{k+1} = \cos\big(h\|\mathbf{v}_k\|_2\big)\,\mathbf{u}_k + \sin\big(h\|\mathbf{v}_k\|_2\big)\,\frac{\mathbf{v}_k}{\|\mathbf{v}_k\|_2}, \qquad \mathbf{v}_k = \mathbf{P}_{\mathbf{A},\mathbf{u}_k}\mathbf{f}_{\bar{\phi}}(\mathbf{u}_k, t_k \mid \mathbf{A}, \theta). \tag{10}$$

Table 1: Downstream FR verification accuracy (%) compared with representative synthetic-data methods. CASIA is the authentic training set and is listed for reference. Both FaceFlow entries use $\mathrm{lb} = 0.5$. **Bold** and underlined numbers indicate the best and second-best synthetic-data results within each data scale, respectively.

Method	Venue	Generator	#imgs (IDs × imgs/ID)	LFW	AgeDB	CFP-FP	CALFW	CPLFW	Avg
CASIA	(Real)	-	0.49M (∼10.5k×47)	99.55	94.55	95.31	93.78	89.95	94.63
DigiFace	WACV 23	Rendering	0.5M (10k×50)	95.40	76.97	87.40	78.62	78.87	83.45
SFace	IJCB 22	GAN	0.63M (10.5k×60)	91.87	71.68	73.86	77.93	73.20	77.71
SynFace	ICCV 21	GAN	0.5M (10k×50)	91.93	61.63	75.03	74.73	70.43	74.75
USynthFace	FG 23	GAN	0.4M (400k×1)	92.23	71.62	78.56	77.05	72.03	78.30
ExFaceGAN	IJCB 23	GAN	0.5M (10k×50)	93.50	78.92	73.84	82.98	71.60	80.17
IDiff-Face	ICCV 23	Diffusion	0.5M (10k×50)	98.00	86.43	85.47	90.65	80.45	88.20
Arc2Face	ECCV 24	Diffusion	0.5M (10k×50)	98.81	90.18	91.87	92.63	85.16	91.73
DCFace	CVPR 23	Diffusion	0.5M (10k×50)	98.55	89.70	85.33	91.60	82.62	89.56
ID^3	NeurIPS 24	Diffusion	0.5M (10k×50)	97.68	91.00	86.84	90.73	82.77	89.80
CemiFace	NeurIPS 24	Diffusion	0.5M (10k×50)	99.03	91.33	91.06	92.42	87.65	92.30
HyperFace	ICLR 25	Diffusion	0.5M (10k×50)	98.50	86.53	88.83	89.40	84.23	89.50
Vec2Face	ICLR 25	Diffusion	0.5M (10k×50)	98.87	93.12	88.97	**93.57**	85.47	92.00
UIFace	ICLR 25	Diffusion	0.5M (10k×50)	99.27	90.95	94.29	92.25	**89.58**	93.27
IDPerturb	CVPR 26	Diffusion	0.5M (10k×50)	**99.40**	93.20	93.61	93.50	88.37	93.62
FaceFlow	Ours	Diffusion	0.5M (10k×50)	99.15	**93.68**	**94.49**	93.40	89.32	**94.01**
DigiFace	WACV 23	Rendering	1.2M (10k×72 + 100k×5)	96.17	81.10	89.81	82.55	82.23	86.37
DCFace	CVPR 23	Diffusion	1.2M (20k×50 + 40k×5)	98.58	90.97	88.61	92.82	85.07	91.21
CemiFace	NeurIPS 24	Diffusion	1.0M (20k×50)	99.18	91.97	92.75	93.01	88.42	93.07
Arc2Face	ECCV 24	Diffusion	1.2M (20k×50 + 40k×5)	98.92	92.45	94.58	93.33	86.45	93.14
Vec2Face	ICLR 25	Diffusion	1.0M (20k×50)	98.87	93.85	89.87	93.65	86.13	92.47
UIFace	ICLR 25	Diffusion	1.0M (20k×50)	99.22	92.45	95.03	93.18	90.42	94.06
IDPerturb	CVPR 26	Diffusion	1.0M (20k×50)	99.48	94.03	95.01	93.85	90.01	94.48
FaceFlow	Ours	Diffusion	1.0M (20k×50)	**99.50**	**94.12**	**95.71**	**94.23**	**90.55**	**94.82**

The exponential map advances along a geodesic arc and keeps $\mathbf{u}_k$ on the unit tangent hypersphere by construction. When $\|\mathbf{v}_k\|_2 = 0$, we set $\mathbf{u}_{k+1} = \mathbf{u}_k$. After K steps, the perturbation direction $\mathbf{u}_K$ is combined with the perturbation angle θ to form the final identity condition, which the generator then decodes:

$$\mathbf{A}' = \cos\theta\,\mathbf{A} + \sin\theta\,\mathbf{u}_K, \qquad x_{\mathrm{syn}} = \mathcal{G}_\psi(\mathbf{A}', x_T), \tag{11}$$

where x_T is the initial diffusion noise. The separation of perturbation magnitude (θ) and perturbation direction ($\mathbf{u}_K$) persists through inference: θ prescribes how far the condition departs from the reference identity embedding, while FaceFlow determines in which direction.

4 Experiments

4.1 Experimental Setup

Training dataset. FaceFlow is trained on CASIA-WebFace (CASIA) (Yi et al., 2014). The dataset contains 490K face images from 10,575 identities with large intra-class variation. Training pairs are formed from images of the same identity, and the velocity field is trained on their ElasticFace features, that is, in the identity embedding space induced by the ElasticFace encoder. We hold out 5% of the identities, which are never seen during training, for the embedding-space analysis.

Generator. We use a frozen UIFace latent diffusion model (LDM) (Lin et al., 2025) as the generator, and an off-the-shelf ElasticFace-Arc R100 model (Boutros et al., 2022a) with ℓ_2 normalization as the identity encoder. We keep the sampling configuration of UIFace, with the scale of classifier-free guidance set to 1 and the stage-partition threshold of UIFace set to 0.005.

Synthetic data generation. We use the synthetic reference identity embeddings publicly released by UIFace (Lin et al., 2025), extracted from images generated by the unconditional IDiff-Face (Boutros et al., 2023a). The perturbation angle of a reference identity embedding is drawn as

$$c \sim \mathcal{U}[\mathrm{lb}, 1], \qquad \theta = \arccos(c), \tag{12}$$

where lb lower-bounds $\cos\theta$ and therefore fixes the angular range. The main setting is $\mathrm{lb} = 0.5$; the robustness study scans $\mathrm{lb} \in \{0.9, 0.8, 0.7, 0.6, 0.5, 0.4\}$, and the comparisons at the same perturbation angle use $\mathrm{lb} \in \{0.3, 0.5, 0.7\}$. Inference integrates the velocity field with $K = 16$ exponential-map

Table 2: Identity consistency, identity separability, and generation stability at the same perturbation angle. TAR denotes the true acceptance rate of the generated samples against their reference identity embedding, and FDR is the Fisher discriminant ratio between their genuine and impostor scores. All results are from our reproduction. **Bold** marks the better result in each pair. *Zero by definition.

Method	lb	TAR (%) @10^{-3} ↑	TAR (%) @10^{-4} ↑	FID ↓	C_{intra} ↑	FDR ↑	Detection Failure (%) ↓	Total Rejection (%) ↓
CASIA	—	87.47	85.09	0.00*	0.9684	1.217	1.08	13.50
Unperturbed (lb = 1)	—	99.00	97.96	24.12	0.9936	1.083	0.12	1.12
IDPerturb	0.3	62.84	52.04	26.40	0.6720	0.759	0.96	37.76
FaceFlow	0.3	**86.08**	**74.16**	**24.19**	**0.9236**	**0.960**	**0.20**	**14.09**
IDPerturb	0.5	80.44	69.08	25.49	0.8364	1.320	0.68	20.11
FaceFlow	0.5	**96.68**	**90.48**	**24.37**	**0.9772**	**1.646**	**0.28**	**3.59**
IDPerturb	0.7	92.84	87.16	24.67	0.9448	2.652	0.48	7.61
FaceFlow	0.7	**98.44**	**96.20**	**24.17**	**0.9884**	**3.232**	**0.16**	**1.72**

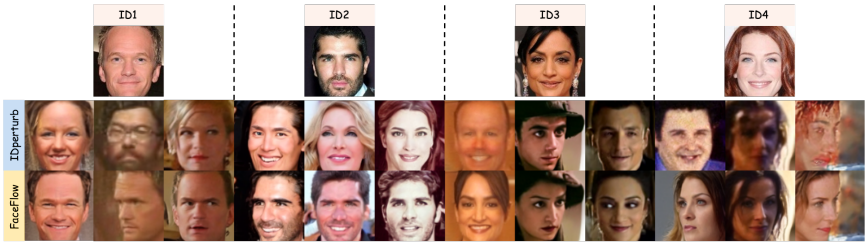


Figure 4: **Qualitative comparison at the same perturbation angle.** Each group corresponds to one identity, with its reference image on top, and all faces are generated at $\cos\theta = 0.6$. **Upper:** IDPerturb. **Lower:** FaceFlow.

steps. For 10k and 20k identities, we decode 50 images at 128^2, giving the 0.5M and 1.0M training sets.

Downstream FR. We train a ResNet-50 (He et al., 2016) with the CosFace loss (Wang et al., 2018) (margin 0.35, scale 64) for 34 epochs using SGD with batch size 512, momentum 0.9, and weight decay 5e−4. The learning rate starts at 0.1 and is divided by 10 at epochs 22, 28, and 32. We evaluate the model on five widely used benchmarks, LFW (Huang et al., 2008), AgeDB-30 (Moschoglou et al., 2017), CFP-FP (Sengupta et al., 2016), CALFW (Zheng et al., 2017), and CPLFW (Zheng & Deng, 2018). Their variation in age, pose, and expression makes them a broad test of generalization.

4.2 Comparison with Synthetic-Data Methods

We compare FaceFlow with state-of-the-art synthetic-data methods based on digital rendering, GANs, and DMs in Table 1. Compared with the methods based on digital rendering or GANs, such as DigiFace, SFace, and ExFaceGAN, FaceFlow achieves higher accuracy on all five benchmarks. Among the methods based on DMs at the 0.5M scale, FaceFlow ranks first on AgeDB-30 and CFP-FP and second on CPLFW, slightly behind UIFace. Averaged over the five benchmarks, FaceFlow reaches 94.01%, the highest among all compared synthetic-data methods. At the 1.0M scale, FaceFlow ranks first on every benchmark, with an average accuracy of 94.82%.

4.3 Effect of the Learned Flow on Identity Consistency and Intra-Class Variation

With the velocity field set to zero, FaceFlow leaves $\mathbf{u}_0$ unchanged and reduces to random perturbation, so the learned flow is the only component that separates the two. We therefore generate both with the same reference identity embeddings, per-sample θ, and initial diffusion noise x_T, and examine what the learned flow changes in identity consistency and intra-class variation.

Identity consistency. Figure 4 shows the comparison at $\cos\theta = 0.6$. At the same perturbation angle, faces generated with the learned flow drift less from the reference identity. Table 2 quantifies this

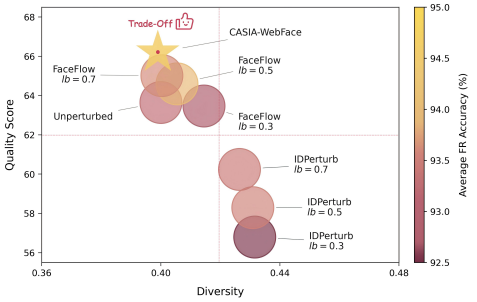


Figure 5: **Intra-class diversity against face image quality.** Each bubble is one synthetic dataset: FaceFlow and IDPerturb at $\mathrm{lb} \in \{0.3, 0.5, 0.7\}$ and the unperturbed baseline; the star marks CASIA. The horizontal axis is D_{intra}, the vertical axis is SDD-FIQA, and color indicates the mean downstream FR accuracy.

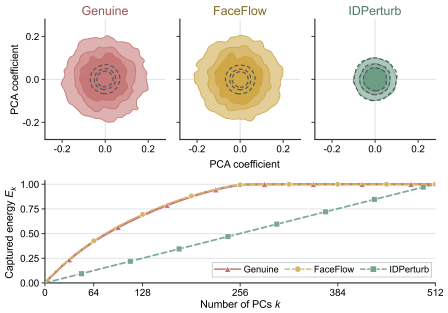


Figure 6: **PCA of intra-class directions on unseen identities.** All directions are taken at the same perturbation angle. **Top:** Coefficient densities on the leading PCA axes for the directions of real same-identity pairs (Genuine), FaceFlow, and IDPerturb. **Bottom:** Energy fraction E_k captured by the leading k axes.

observation. A decoded image is counted as rejected if face detection fails on it or its ElasticFace embedding does not verify against the reference identity embedding at a false acceptance rate (FAR) of 10^{-3}. The detection failure rate and the total rejection rate are the two quantities through which we report generation stability. We further report TAR at two FAR operating points, which measures how reliably a generated sample is recognized as the identity it was conditioned on. The intra-class consistency C_{intra} (Boutros et al., 2026), the fraction of samples whose cosine similarity to their class center is at least 0.3, measures how tightly the samples of one synthetic identity cluster around that identity. The Fréchet inception distance (FID) (Heusel et al., 2017) against CASIA measures how far the overall image distribution departs from authentic data. At the same angle, the faces generated with the learned flow are therefore more consistent with the identity they were conditioned on. The gain is also visible in C_{intra}, which is higher for FaceFlow at every angular range, so the samples of each synthetic identity form a tighter cluster around it. FaceFlow also attains a higher Fisher discriminant ratio (FDR) (Poh & Bengio, 2004) between genuine and impostor scores at every angular range, indicating better identity separability. In terms of FID, their distribution stays close to that of the unperturbed baseline, whereas the images of IDPerturb move further from authentic data.

Intra-class variation. The intra-class variation of a synthetic dataset is what makes it useful for training an FR model, and it has to be obtained without sacrificing face image quality. Figure 5 shows the intra-class diversity D_{intra} (Boutros et al., 2026) against the SDD-FIQA face image quality score (Ou et al., 2021), where larger values indicate greater appearance variation and better face image quality, respectively; CASIA is the reference because it already combines the two. IDPerturb reaches a higher D_{intra}, but its TAR and rejection rates show that part of this excess is a loss of identity consistency rather than genuine intra-class variation. The learned flow gains in both respects, which explains its advantage in downstream accuracy.

4.4 The Learned Flow Recovers Real Intra-Class Structure

The preceding measurements are taken on decoded images and therefore confound the perturbation with the generator. To test the premise of the method itself, we repeat the comparison in the identity embedding space, without involving the generator. On the held-out identities we draw 20,000 same-identity pairs and initialize FaceFlow and IDPerturb with the same initial perturbation directions $\mathbf{u}_0$. Agreement with the real structure is measured by $E_k(\mathbf{u}) = \|\mathbf{Q}_{k,\mathbf{A}}^{\top}\mathbf{u}\|_2^2$, the fraction of the energy of a unit perturbation direction that $\mathbf{Q}_{k,\mathbf{A}}$ captures. Here $\mathbf{Q}_{k,\mathbf{A}}$ is an orthonormal basis of the leading k PCA directions of real intra-class variation, projected onto the tangent space at $\mathbf{A}$.

Figure 6 compares both methods with the directions of the real held-out pairs. For a direction drawn uniformly at random, the expected energy in any k-dimensional subspace is $k/(d-1)$, so the curve of IDPerturb grows linearly in k: a random direction spreads its energy evenly over all axes and favors none of the axes along which identities vary. The curve of FaceFlow instead rises steeply over the leading axes and follows that of the real directions, so the learned flow concentrates the

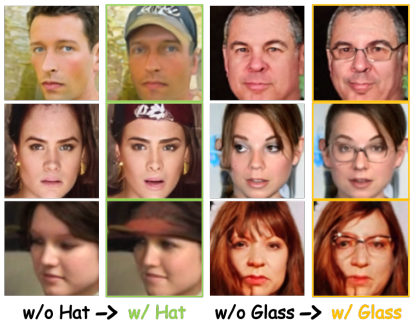


Figure 7: **Attribute-specific flows.** Each pair shows a reference image without the target attribute and the face decoded with the fine-tuned flow at $\theta = 52.1°$, in which the attribute appears. Each row is a different identity. **Left:** Hats. **Right:** Eyeglasses.

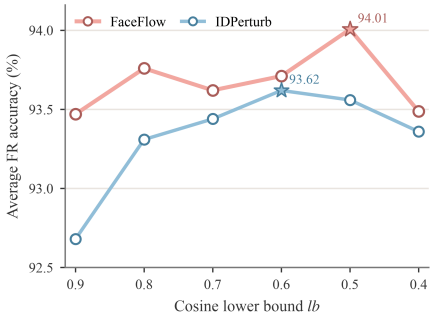


Figure 8: **Downstream accuracy across the angular range.** Mean verification accuracy (%) over the five benchmarks for FR models trained on 0.5M images. The angular range widens from left to right as lb decreases, and stars mark the best lb of each method.

perturbation on the dominant axes of real intra-class variation. Since no generator is involved, this agreement is a property of the flow itself rather than of the images decoded from it.

Attribute-specific flows. The intra-class variation that FaceFlow learns can also be made specific to one attribute. Fine-tuning the velocity field on same-identity pairs ordered from the absence to the presence of an attribute gives an attribute-specific flow, trained separately for hats and eyeglasses with the generator frozen. Hat pairs are drawn from CASIA and CelebA-HQ (Karras et al., 2017), eyeglasses pairs from MeGlass (Guo et al., 2018), and the fine-tuned flow is compared with IDPerturb at the same $\theta = 52.1°$, the mean angle between the two images of a real eyeglasses pair in MeGlass, which we use for both attributes. Figure 7 shows the decoded pairs, and Table 3 reports the attribute occurrence rates measured by the FaRL face-attribute classifier of the FACER toolkit (Zheng et al., 2022). The target attribute appears far more often than under a random perturbation direction, while the face remains visually similar to the reference. These results show that FaceFlow can learn attribute-specific flows by fine-tuning a separate model for each attribute. A single attribute-aware flow that takes the target attribute as an explicit control signal is a natural next step; it would let the synthetic data be tailored to a target domain and could improve the accuracy and robustness of FR models in that domain.

Table 3: Attribute occurrence rates.

Method	Hat		Eyeglasses	
	θ	Rate ↑	θ	Rate ↑
Unperturbed (lb = 1)	0°	4.2%	0°	14.7%
IDPerturb	52.1°	9.6%	52.1°	14.1%
FaceFlow	52.1°	**38.9%**	52.1°	**61.2%**

4.5 Robustness to the Angular Range

To test whether the comparison depends on lb, the only hyperparameter of the generation protocol, we generate one dataset per method at each lb and train an FR model on it, keeping the identity pool, the random seeds, and the FR protocol fixed. Figure 8 reports the mean accuracy over the five benchmarks. FaceFlow is ahead at every tested lb, so its advantage does not depend on the choice of lb. Its accuracy also varies over a narrower range than that of IDPerturb, indicating that FaceFlow is robust to the cosine lower bound.

5 Conclusion

FaceFlow replaces the random perturbation direction of embedding-space perturbation with an identity-conditioned flow, learned from same-identity pairs by Riemannian flow matching on the unit tangent hypersphere of the reference identity embedding, so that the synthesized intra-class variation follows that of real faces more closely. With the generator frozen, this single change raises identity consistency at the same perturbation angle, and FR models trained on the resulting datasets reach 94.01% and 94.82% average accuracy at the 0.5M and 1.0M scales.

AI USE STATEMENT

We used generative AI tools in two ways. First, to aid and polish the writing, including editing the manuscript for readability and assisting with LaTeX formatting. Second, for retrieval and discovery, including searching for related work. All AI-assisted text and retrieved references were checked by the authors, who take full responsibility for the content of this work, including its text, claims, and artifacts.

REFERENCES

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 *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 3515–3524. IEEE, 2023.

Fadi Boutros, Naser Damer, Florian Kirchbuchner, and Arjan Kuijper. ElasticFace: Elastic margin loss for deep face recognition. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1577–1586. IEEE, 2022a.

Fadi Boutros, Marco Huber, Patrick Siebke, Tim Rieber, and Naser Damer. SFace: Privacy-friendly and accurate face recognition using synthetic data. In *2022 IEEE International Joint Conference on Biometrics (IJCB)*, pp. 1–11. IEEE, 2022b.

Fadi Boutros, Jonas Henry Grebe, Arjan Kuijper, and Naser Damer. IDiff-Face: Synthetic-based face recognition through fizzy identity-conditioned diffusion models. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 19593–19604. IEEE, 2023a.

Fadi Boutros, Marcel Klemt, Meiling Fang, Arjan Kuijper, and Naser Damer. ExFaceGAN: Exploring identity directions in GAN's learned latent space for synthetic identity generation. In *2023 IEEE International Joint Conference on Biometrics (IJCB)*, pp. 1–10. IEEE, 2023b.

Fadi Boutros, Marcel Klemt, Meiling Fang, Arjan Kuijper, and Naser Damer. Unsupervised face recognition using unlabeled synthetic data. In *2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG)*, pp. 1–8. IEEE, 2023c.

Fadi Boutros, Eduarda Caldeira, Tahar Chettaoui, and Naser Damer. IDperturb: Enhancing variation in synthetic face generation via angular perturbations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 40119–40129, 2026.

Qiong Cao, Li Shen, Weidi Xie, Omkar M Parkhi, and Andrew Zisserman. VGGFace2: A dataset for recognising faces across pose and age. In *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*, pp. 67–74. IEEE, 2018.

Ricky T. Q. Chen and Yaron Lipman. Flow matching on general geometries. In *International Conference on Learning Representations*, volume 2024, pp. 47922–47945, 2024.

Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. ArcFace: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4690–4699, 2019.

Jiankang Deng, Jia Guo, Tongliang Liu, Mingming Gong, and Stefanos Zafeiriou. Sub-center ArcFace: Boosting face recognition by large-scale noisy web faces. In *European Conference on Computer Vision*, pp. 741–757. Springer, 2020.

Jianzhu Guo, Xiangyu Zhu, Zhen Lei, and Stan Z Li. Face synthesis for eyeglass-robust face recognition. In *Chinese Conference on biometric recognition*, pp. 275–284. Springer, 2018.

Yandong Guo, Lei Zhang, Yuxiao Hu, Xiaodong He, and Jianfeng Gao. MS-Celeb-1M: A dataset and benchmark for large-scale face recognition. In *European conference on computer vision*, pp. 87–102. Springer, 2016.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.

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. *Advances in neural information processing systems*, 30, 2017.

Gary B Huang, Marwan Mattar, Tamara Berg, and Eric Learned-Miller. Labeled Faces in the Wild: A database for studying face recognition in unconstrained environments. In *Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition*, 2008.

Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of GANs for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*, 2017.

Minchul Kim, Feng Liu, Anil Jain, and Xiaoming Liu. DCFace: Synthetic face generation with dual condition diffusion model. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12715–12725. IEEE, 2023.

Xiao Lin, Yuge Huang, Jianqing Xu, Yuxi Mi, Shuigeng Zhou, and Shouhong Ding. UIFace: Unleashing inherent model capabilities to enhance intra-class diversity in synthetic face recognition. In *International Conference on Learning Representations*, volume 2025, pp. 87296–87309, 2025.

Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022.

Leland McInnes, John Healy, and James Melville. UMAP: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.

Stylianos Moschoglou, Athanasios Papaioannou, Christos Sagonas, Jiankang Deng, Irene Kotsia, and Stefanos Zafeiriou. AgeDB: the first manually collected, in-the-wild age database. In *2017 IEEE conference on computer vision and pattern recognition workshops (CVPRW)*, pp. 1997–2005. IEEE, 2017.

Hatef Otroshi Shahreza and Sébastien Marcel. HyperFace: Generating synthetic face recognition datasets by exploring face embedding hypersphere. In *International Conference on Learning Representations*, volume 2025, pp. 38167–38185, 2025.

Fu-Zhao Ou, Xingyu Chen, Ruixin Zhang, Yuge Huang, Shaoxin Li, Jilin Li, Yong Li, Liujuan Cao, and Yuan-Gen Wang. SDD-FIQA: Unsupervised face image quality assessment with similarity distribution distance. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7666–7675. IEEE, 2021.

Foivos Paraperas Papantoniou, Alexandros Lattas, Stylianos Moschoglou, Jiankang Deng, Bernhard Kainz, and Stefanos Zafeiriou. Arc2Face: A foundation model for ID-consistent human faces. In *European Conference on Computer Vision*, pp. 241–261. Springer, 2024.

William Peebles and Saining Xie. Scalable diffusion models with transformers. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4172–4182. IEEE, 2023.

Norman Poh and Samy Bengio. A study of the effects of score normalisation prior to fusion in biometric authentication tasks. Technical report, IDIAP, 2004.

Simon JD Prince and James H Elder. Probabilistic linear discriminant analysis for inferences about identity. In *2007 IEEE 11th international conference on computer vision*, pp. 1–8. IEEE, 2007.

Haibo Qiu, Baosheng Yu, Dihong Gong, Zhifeng Li, Wei Liu, and Dacheng Tao. SynFace: Face recognition with synthetic data. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 10860–10870. IEEE, 2021.

Soumyadip Sengupta, Jun-Cheng Chen, Carlos Castillo, Vishal M Patel, Rama Chellappa, and David W Jacobs. Frontal to profile face verification in the wild. In *2016 IEEE winter conference on applications of computer vision (WACV)*, pp. 1–9. IEEE, 2016.

Yichun Shi and Anil K Jain. Probabilistic face embeddings. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 6902–6911, 2019.

Zhonglin Sun, Siyang Song, Ioannis Patras, and Georgios Tzimiropoulos. CemiFace: Center-based semi-hard synthetic face generation for face recognition. In *Advances in Neural Information Processing Systems*, volume 37, pp. 35612–35638. Curran Associates, Inc., 2024.

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 *2018 IEEE/CVF conference on computer vision and pattern recognition*, pp. 5265–5274. IEEE, 2018.

Haiyu Wu, Jaskirat Singh, Sicong Tian, Liang Zheng, and Kevin Bowyer. Vec2Face: Scaling face dataset generation with loosely constrained vectors. In *International Conference on Learning Representations*, volume 2025, pp. 23490–23503, 2025.

Jianqing Xu, Shen Li, Jiaying Wu, Miao Xiong, Ailin Deng, Jiazhen Ji, Yuge Huang, Guodong Mu, Wenjie Feng, Shouhong Ding, et al. ID^3: Identity-preserving-yet-diversified diffusion models for synthetic face recognition. *Advances in Neural Information Processing Systems*, 37:77777–77798, 2024.

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

Xi Yin, Xiang Yu, Kihyuk Sohn, Xiaoming Liu, and Manmohan Chandraker. Feature transfer learning for face recognition with under-represented data. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5697–5706. IEEE, 2019.

Tianyue Zheng and Weihong Deng. Cross-pose LFW: A database for studying cross-pose face recognition in unconstrained environments. *Beijing University of Posts and Telecommunications, Tech. Rep*, 5(7):5, 2018.

Tianyue Zheng, Weihong Deng, and Jiani Hu. Cross-age LFW: A database for studying cross-age face recognition in unconstrained environments. *arXiv preprint arXiv:1708.08197*, 2017.

Yinglin Zheng, Hao Yang, Ting Zhang, Jianmin Bao, Dongdong Chen, Yangyu Huang, Lu Yuan, Dong Chen, Ming Zeng, and Fang Wen. General facial representation learning in a visual-linguistic manner. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 18676–18688. IEEE, 2022.

A Notation

Table 4 summarizes the symbols used in the main text.

Table 4: Notation used in the main text.

Symbol	Meaning
$\mathcal{F}$, $\mathcal{G}_\psi$	Frozen identity encoder, frozen latent diffusion generator
$\mathbf{A} \in \mathbb{S}^{d-1}$, $\mathbf{A}'$	Reference identity embedding, perturbed identity embedding
$T_{\mathbf{A}}\mathbb{S}^{d-1}$	Tangent space at $\mathbf{A}$: $\{\mathbf{v} \in \mathbb{R}^d : \mathbf{v}^\top \mathbf{A} = 0\}$
$\mathcal{S}_{\mathbf{A}}$	Unit tangent hypersphere at $\mathbf{A}$: $\{\mathbf{u} \in T_{\mathbf{A}}\mathbb{S}^{d-1} : \|\mathbf{u}\| = 1\} \cong \mathbb{S}^{d-2}$
θ, $\mathbf{u}$	Perturbation angle, perturbation direction on $\mathcal{S}_{\mathbf{A}}$
$(\mathbf{a}, \mathbf{b})$, $\mathbf{u}_1$	Normalized encoder features of a real same-identity pair, and the observed direction recovered from it
$\mathbf{u}_0$, $\mathbf{u}_t$, $\dot{\mathbf{u}}_t$	Random initial direction, geodesic interpolant, and its tangent velocity
$\mathbf{f}_\phi$, $\mathbf{v}_\phi$, $\mathbf{P}_{\mathbf{A},\mathbf{u}}$	Velocity field, projected velocity, tangent-space projection
ϕ, β	EMA parameters and their decay rate
K, h, t_k	Number of integration steps, step size $1/K$, time at step k
x_T, x_{syn}	Initial diffusion noise, synthetic face

B Network Architecture and Training Details

Condition encoding. We normalize θ by $\theta_{\max} = \pi/2$ and encode t and θ with Fourier features

$$\gamma(z) = \big[\sin(2\pi\omega_k z), \cos(2\pi\omega_k z)\big]_{k=1}^{m}, \qquad \omega_k = \exp\Big(\frac{k-1}{m-1}\log\omega_{\max}\Big), \tag{13}$$

so that $t \in [0, 1]$ and θ occupy a comparable numerical range. The three conditions are then mapped to a shared width and added:

$$\mathbf{c}_t = E_t(\gamma(t)), \qquad \mathbf{c}_\theta = E_\theta\big(\gamma(\theta/\theta_{\max})\big), \qquad \mathbf{c}_A = E_A(\mathbf{A}), \qquad \mathbf{c} = \mathbf{c}_t + \mathbf{c}_\theta + \mathbf{c}_A. \tag{14}$$

The fused condition is injected into the backbone H_ϕ block by block through adaptive layer normalization with zero initialization (adaLN-Zero) (Peebles & Xie, 2023):

$$\mathbf{f}_\phi(\mathbf{u}, t \mid \mathbf{A}, \theta) = H_\phi\big(\mathbf{u};\, \mathbf{c}_t + \mathbf{c}_\theta + \mathbf{c}_A\big). \tag{15}$$

We use $m = 128$ frequencies with $\omega_{\max} = 1000$ (Table 5). The maps E_t and E_θ are two-layer multilayer perceptrons (MLPs) with SiLU activations, and E_A is a linear layer.

Velocity backbone. Algorithm 2 gives the layer-wise implementation of H_ϕ. Layer normalization (LN) has no learnable affine parameters. The per-block shift, scale, and gate are regressed from the condition vector. The condition projection $W_{\text{ada}}^{(\ell)}$ of each block and the output head W_{out} are zero-initialized, so that every block is initially an identity map and the full network outputs the zero vector.

Algorithm 2 Velocity backbone $H_\phi(\mathbf{u}; \mathbf{c})$.

Require: Current direction $\mathbf{u} \in \mathbb{R}^d$, condition vector $\mathbf{c} \in \mathbb{R}^w$
Ensure: Raw velocity $\mathbf{f} \in \mathbb{R}^d$ (projection is applied outside)
$\mathbf{h} \leftarrow W_{\text{in}}\mathbf{u}$ $\{d \to w\}$
for $\ell = 1, \ldots, L$ **do**
  $(\text{shift}, \text{scale}, \text{gate}) \leftarrow W_{\text{ada}}^{(\ell)}\mathbf{c}$ $\{w \to 3w$, zero-initialized$\}$
  $\mathbf{h} \leftarrow \mathbf{h} + \text{gate} \odot \text{MLP}^{(\ell)}\big(\text{LN}(\mathbf{h}) \odot (1 + \text{scale}) + \text{shift}\big)$ {MLP: $w \to 4w \to w$, GELU}
end for
return $W_{\text{out}}\mathbf{h}$ $\{w \to d$, zero-initialized$\}$

Inference procedure. Algorithm 3 is the inference procedure of Section 3.3: the perturbation direction of an unseen identity is moved along the learned flow on the unit tangent hypersphere, and the resulting perturbed identity embedding is decoded by the frozen generator.

Algorithm 3 FaceFlow inference.

Require: Reference identity embedding $\mathbf{A}$, generator $\mathcal{G}_\psi$, velocity field $\mathbf{f}_{\bar{\phi}}$, perturbation angle θ, step count K, initial diffusion noise x_T
Ensure: Synthetic face x_{syn}
$h \leftarrow 1/K$
$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, $\mathbf{u}_0 \leftarrow \text{Proj}_{\mathcal{S}_\mathbf{A}}(\boldsymbol{\epsilon})$
for $k = 0, \ldots, K-1$ **do**
  $\mathbf{v}_k \leftarrow \mathbf{P}_{\mathbf{A},\mathbf{u}_k}\mathbf{f}_{\bar{\phi}}(\mathbf{u}_k, t_k \mid \mathbf{A}, \theta)$
  if $\|\mathbf{v}_k\| = 0$ **then**
    $\mathbf{u}_{k+1} \leftarrow \mathbf{u}_k$
  else
    $\mathbf{u}_{k+1} \leftarrow \cos(h\|\mathbf{v}_k\|)\,\mathbf{u}_k + \sin(h\|\mathbf{v}_k\|)\,\mathbf{v}_k/\|\mathbf{v}_k\|$
  end if
end for
$\mathbf{A}' \leftarrow \cos\theta\,\mathbf{A} + \sin\theta\,\mathbf{u}_K$
return $x_{\text{syn}} \leftarrow \mathcal{G}_\psi(\mathbf{A}', x_T)$

Exponential moving average. In addition to the online parameters ϕ, we maintain an EMA copy

$$\bar{\phi} \leftarrow \beta\bar{\phi} + (1-\beta)\phi, \qquad \bar{\phi}_0 = \phi_0, \tag{16}$$

and use $\bar{\phi}$ for all flow integration and image generation at inference.

Pair construction. Near-duplicate image pairs are removed using encoder features, because pairs with cosine similarity near 1 do not supply a usable direction. Same-identity pairs are kept only when the cosine similarity satisfies $s \geq 0.10$, which suppresses label noise. Each identity contributes at most 2000 pairs, so that identities with many images do not dominate the training set. Pairs are assigned to 20 equal-width bins in θ, and each mini-batch is balanced across bins. The held-out identities are never used for training, and the flow is always applied to reference identity embeddings of unseen identities.

C Closed-Form Geometry and Invariance Properties

Geodesic interpolation. In Eq. (6), $\mathbf{u}_t$ is a linear combination of $\mathbf{u}_0$ and $\mathbf{u}_1$, so $\mathbf{u}_t^\top \mathbf{a} = 0$. Direct computation gives $\|\mathbf{u}_t\|_2 = 1$, $\mathbf{u}_t^\top \dot{\mathbf{u}}_t = 0$, and $\|\dot{\mathbf{u}}_t\|_2 = \Omega$.

Projection operator. Write $\Pi_\mathbf{A}\mathbf{v} = \mathbf{v} - (\mathbf{v}^\top\mathbf{A})\mathbf{A}$ for projection onto $\mathbf{A}^\perp$. Because $\mathbf{u} \perp \mathbf{A}$, the operator in Eq. (7), evaluated at $\mathbf{A}$, satisfies $\mathbf{P}_{\mathbf{A},\mathbf{u}} = \Pi_\mathbf{u} \circ \Pi_\mathbf{A}$. The two projections commute. The implementation subtracts the two projection components and never forms a $d \times d$ matrix.

Table 5: Architecture and optimization hyperparameters of FaceFlow.

Architecture	
Embedding dimension (d)	512
Backbone width (w)	768
Number of blocks (L)	8
MLP expansion ratio	4
Fourier frequencies (m)	128
Maximum frequency ($\omega_{\max}$)	1000
Angle normalization cap ($\theta_{\max}$)	$\pi/2$
Data	
Minimum pairing cosine	0.10
Maximum pairs per identity	2000
Number of angle bins	20
Held-out identity fraction	5%

Optimization	
Optimizer	AdamW
Learning rate / weight decay	10^{-4} / 0.01
Learning rate schedule	2000 step linear warmup + cosine decay
Total steps / batch size	100k / 4096
Gradient clipping	1.0
Numerical precision	bf16
EMA decay (β)	0.999

Exponential-map discretization. Eq. (10) replaces the Euler update $\mathbf{u}_k + h\mathbf{v}_k$, which leaves the unit tangent hypersphere, with the exponential map: it advances along the geodesic of tangent $\mathbf{v}_k$ by arc length $h\|\mathbf{v}_k\|_2$.

Constraint preservation. Suppose $\mathbf{u}_k^\top \mathbf{A} = 0$, $\|\mathbf{u}_k\|_2 = 1$, and $\mathbf{v}_k$ satisfies $\mathbf{v}_k \perp \mathbf{A}$ and $\mathbf{v}_k \perp \mathbf{u}_k$ (by construction of $\mathbf{P}$). If $\mathbf{v}_k = \mathbf{0}$, then $\mathbf{u}_{k+1} = \mathbf{u}_k$ and both constraints hold trivially; otherwise, let $\alpha = h\|\mathbf{v}_k\|_2$. The update $\mathbf{u}_{k+1} = \cos\alpha\, \mathbf{u}_k + \sin\alpha\, \mathbf{v}_k/\|\mathbf{v}_k\|_2$ is a linear combination of two vectors in $\mathbf{A}^\perp$ that are orthogonal and of unit norm. Hence $\mathbf{u}_{k+1}^\top \mathbf{A} = 0$ and $\|\mathbf{u}_{k+1}\|_2^2 = \cos^2\alpha + \sin^2\alpha = 1$. After K steps, $\mathbf{u}_K \in \mathcal{S}_\mathbf{A}$, and $\mathbf{A}' = \cos\theta\, \mathbf{A} + \sin\theta\, \mathbf{u}_K$ satisfies $\|\mathbf{A}'\|_2 = 1$ and $\mathbf{A}^\top \mathbf{A}' = \cos\theta$. In floating-point arithmetic, round-off may cause small departures. After each exponential-map step we apply $\Pi_\mathbf{A}$ and renormalize; these corrections are $O(\varepsilon_{\mathrm{mach}})$ and do not change the mathematical meaning of the update.

Zero-velocity recovery. When $\mathbf{f}_{\phi_0} \equiv \mathbf{0}$, $\mathbf{v}_k = \mathbf{0}$ for every k, so the update after Eq. (10) sets $\mathbf{u}_{k+1} = \mathbf{u}_k$ and $\mathbf{u}_K = \mathbf{u}_0$. Eq. (11) then gives the same identity condition as random perturbation.

Fixed-angle distance scaling. For a fixed reference identity embedding $\mathbf{A}$ and a fixed perturbation angle $\theta \in (0, \pi/2)$, define $F_{\mathbf{A},\theta} \colon \mathcal{S}_\mathbf{A} \to \mathbb{S}^{d-1}$ by $F_{\mathbf{A},\theta}(\mathbf{u}) = \cos\theta\, \mathbf{A} + \sin\theta\, \mathbf{u}$. For any $\mathbf{u}, \tilde{\mathbf{u}} \in \mathcal{S}_\mathbf{A}$:

$$\|F_{\mathbf{A},\theta}(\mathbf{u}) - F_{\mathbf{A},\theta}(\tilde{\mathbf{u}})\|_2 = \sin\theta\, \|\mathbf{u} - \tilde{\mathbf{u}}\|_2,$$

because the $\cos\theta\, \mathbf{A}$ terms cancel. This extends to distributional distances: under the ambient Euclidean metric on $\mathbb{R}^d$,

$$W_2\big((F_{\mathbf{A},\theta})_\# q,\ (F_{\mathbf{A},\theta})_\# r\big) = \sin\theta\, W_2(q, r)$$

for any distributions q, r on $\mathcal{S}_\mathbf{A}$ with finite second moments, because $F_{\mathbf{A},\theta}$ scales all pairwise distances by the constant $\sin\theta$ and optimal couplings map bijectively.

D Conditional Velocity Decomposition

This appendix derives the loss decomposition stated in Section 3.2.

Fix a reference identity embedding $\mathbf{A}$ and a perturbation angle θ. The random initial direction $\mathbf{u}_0 \sim \mathrm{Uniform}(\mathcal{S}_\mathbf{A})$ and an observed direction $\mathbf{u}_1$ from same-identity pairs define the random Slerp path $\mathbf{U}_t$ and its velocity $\dot{\mathbf{U}}_t$ via Eq. (6). Assume $\dot{\mathbf{U}}_t$ has finite second moment. The Bayes-optimal velocity is

$$\mathbf{v}^*(\mathbf{u}, t \mid \mathbf{A}, \theta) = \mathbb{E}\big[\dot{\mathbf{U}}_t \mid \mathbf{U}_t{=}\mathbf{u},\ t,\ \mathbf{A},\ \theta\big].$$

By the orthogonality property of conditional expectations, the squared-error loss decomposes as

$$\mathbb{E}\big[\|\mathbf{v}_\phi(\mathbf{U}_t, t \mid \mathbf{A}, \theta) - \dot{\mathbf{U}}_t\|_2^2\big] = \mathbb{E}\big[\|\mathbf{v}_\phi(\mathbf{U}_t, t \mid \mathbf{A}, \theta) - \mathbf{v}^*(\mathbf{U}_t, t \mid \mathbf{A}, \theta)\|_2^2\big] + C(\mathbf{A}, \theta), \quad (17)$$

where $C(\mathbf{A}, \theta) = \mathbb{E}[\|\dot{\mathbf{U}}_t - \mathbf{v}^*(\mathbf{U}_t, t \mid \mathbf{A}, \theta)\|_2^2]$ is independent of ϕ. Taking expectation over the training distribution of $(\mathbf{A}, \theta)$ preserves this structure: the training loss in Eq. (8) is the approximation error plus a ϕ-independent constant.