Ridges Beyond the Window: Structure-First Completion of Contactless Fingerprints from the Capture Sweep

Anonymous authors
Paper under double-blind review

Abstract

A contactless fingerprint captured in the wild is both *degraded* and *partial*, because a phone camera sees a blurred, noisy patch of the fingertip rather than a clean rolled impression. We present SweepFill, a single network that restores what such a capture shows and generates ridges of the correct flow where it shows nothing. Each capture is a short video of the turning finger, which we call a *sweep*, so a part hidden in one frame often appears in another, and no published completion method uses this. Because the exact position of ridges (their *phase*) beyond the visible window is not determined by the observation, while their direction (*flow*) and spacing (*period*) are, SweepFill predicts structure before pixels. A head on the bottleneck predicts ridge flow and frequency for the whole print, a phase-free spectral loss lets the decoder draw ridges without being punished for phase it cannot know, and joint attention over other frames of the same sweep turns much of the generation into registration, all learned by regression alone. We also show that inserting the mask token after the positional encoding makes every hidden token of a masked-bottleneck transformer identical, a failure not specific to fingerprints. We evaluate with two fingerprint matchers under a cross-session protocol. When 30% of the print is visible, SweepFill reduces hidden-region orientation error from 20.07 to 11.01 degrees and raises ridge contrast from 0.15 to 0.67 relative to a DiT-UNet restorer trained on more samples, and it lowers the deep matcher's EER from 26.0% to 11.6% on held-out MMFV subjects. Completing a poor frame before fusing scores over the other frames raises fused rank-1 by +8.3 and +11.5 points when 50% and 70% of the print is visible (not significantly at 30%), and an in-domain fine-tune improves both matchers on a second device. The generated flow is within 10.5 degrees RMSE of the clean frame, but its phase is uncorrelated with it, and only 21% of true minutiae are recovered, so the completions are plausible rather than reconstructed.

1 Introduction

Fingerprint recognition is moving off the sensor, and smartphone applications now sit beside certified contact scanners in government interoperability tests (Libert et al., 2020). A phone camera photographs the fingertip instead of pressing it against glass, which trades the hygiene and deformation problems of contact capture (U.S. Department of Homeland Security et al., 2024; Libert et al., 2020) for photographic ones such as motion blur, noise, compression and uneven illumination, and above all for *partial coverage*, since a finger held at an angle shows only part of its print. Given a degraded photograph showing a fraction of the print, can we produce the whole clean print, and is what we produce worth anything to a fingerprint matcher? Figure 1 shows the task and one result. Contactless capture is not a snapshot but a short *video*, in which the finger is rolled, pitched and yawed over a couple of seconds, as in the MMFV database (Malhotra et al., 2023). We call such a video a *sweep*, and because the finger turns during it, the region hidden in one frame is usually visible in another. To our knowledge, no published completion method uses this signal (Section 2). A print is a pattern of parallel ridges, described by its *flow* (the local direction of the ridges), its *period* (the spacing between them) and its *phase* (where exactly each ridge line lies). Phase more than a few ridges beyond the visible window is not determined by what is visible, so a pixel loss drives any model toward the

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

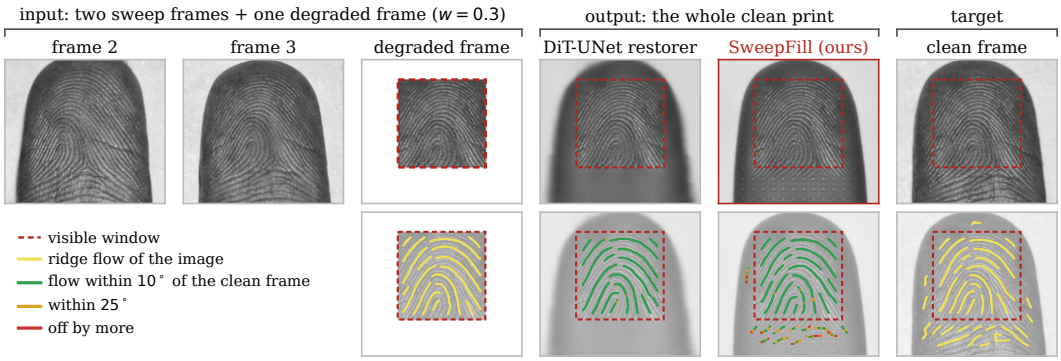

Figure 1: **The task and a result.** The input is one degraded frame of which only a window covering 30% of the crop is visible (w=0.3, dashed), together with two other frames of the same sweep as references, and the output is the whole clean print, restored inside the window and generated outside it. Hatched blocks of the scored hidden region (dotted) have ridge flow more than 15° from the clean frame's. Outside the window the restorer draws flat skin, whereas SweepFill draws ridges whose flow matches the truth, though not the true ridges (Section 5.4). The probe is the held-out identity-protocol probe at the median of SweepFill's per-probe error, and the faint dot lattice is analyzed in Appendix F.4.

average of all possible phases, which is flat, ridge-less skin, the perception-distortion tradeoff (Blau & Michaeli, 2018). Flow and period behave differently, because they are smooth fields that a partial print already constrains and that the sweep constrains further. We present SweepFill, a single network that restores the visible part of a degraded contactless print and generates the rest from the sweep it came from, predicting structure first and pixels second. **(i)** A *structure head*, a small predictor on the network's bottleneck, estimates the flow and period of the whole print, so the bottleneck features of regions the camera never saw must encode ridge structure, and the estimate is fed to every decoder stage. **(ii)** The hidden region is trained with a *phase-free spectral loss*, which compares the frequency content of image patches rather than their pixels, so a correctly flowing pattern half a period out of place is no longer penalized like flat skin. **(iii)** The bottleneck also attends to features of two *reference frames* taken from elsewhere in the same sweep, which turns much of the generation problem into registration. SweepFill is trained by regression alone. None of the three helps if the bottleneck cannot represent the hidden region at all. The bottleneck is a transformer over 16×16-px patches, and each patch with no visible pixel is replaced by a learned *mask token*. Section 3.3 shows that adding the positional embedding before this replacement, rather than after it, makes every hidden token the same vector, so the bottleneck can only output a constant over the hidden region, while the mask-first order of masked autoencoders (He et al., 2022) removes the failure at no cost. Because image metrics cannot say whether a matcher can use the result (Engelsma et al., 2023; Grosz & Jain, 2025; Pan et al., 2026), we evaluate with two open fingerprint matchers (Pandey et al., 2025; Cappelli, 2026b; NIST, 2010) under a protocol grounded in identity (Section 4).

Research Questions.
RQ1. When most of the print never reaches the camera, what can be recovered and what would be invention?
RQ2. What stops a model from representing the very region it was built to fill in?
RQ3. Convincing is not the same as useful. Does our completion help a matcher, and keep helping?

Contributions. (1) SweepFill (RQ1) is, to our knowledge, the first completion model to read the other frames of the capture sweep. When 30% of the print is visible it lowers the orientation error of the generated ridges by 9.1° relative to the strongest restorer we could train, which saw *more* training samples, and the 2.3° its references add is specific to the finger, since another subject's sweep in their place recovers none of it (Table 14, Appendix E). **(2)** We identify the token-ordering requirement above (RQ2), which holds beyond fingerprints, and show that masked autoencoders (He et al., 2022) already satisfy it and that position injected inside attention also avoids the collapse (Appendix C.1). **(3)** We propose an identity-grounded protocol for partial-print completion (RQ3) with two datasets, two matchers, gallery and probe from different sessions, identity leakage, fidelity of *minutiae* (the

ridge endings and splits that matchers compare, Hussein et al., 2026), mosaicking and multi-frame fusion. Under it, completing a 30%-visible probe lowers the equal error rate (EER), the operating point where false accepts and false rejects are equally frequent, from 26.0% to 11.6%.

2 Related work

Contactless Fingerprints. Photographing rather than pressing the finger brings pose distortion, low ridge contrast and background clutter. NIST's interoperability assessment ranks contact capture first, stationary contactless devices close behind and smartphone applications last, names scale control as their persistent weakness, and finds that fusing several fingers closes much of the gap (Libert et al., 2020). We fuse over frames of one finger instead. Like most work, we apply the C2CL preprocessing (Grosz et al., 2022) (segmentation, enhancement, 500 ppi rescaling, unwarping) before every matcher. 3D reconstruction from one finger photo (Cui et al., 2023) or sparse views (Jia et al., 2025) recovers *geometry*, not ridge content that was never imaged. Benchmarks include RidgeBase (Jawade et al., 2022), the finger-selfie databases Malhotra et al. (2020) we use for external validity, and MMFV (Malhotra et al., 2023), whose roll, pitch and yaw video sweeps supply this paper's central signal, although its own baselines score a single frame per video.

Completing Partial Prints. Classical methods model the orientation field globally so that a partial observation constrains the whole field (Wang & Hu, 2011), repair it with orientation dictionaries (Feng et al., 2013; Duan et al., 2021) and grow ridges by contextual Gabor filtering (Hong et al., 1998), a pipeline we implement as a baseline. Learned methods recover missing areas with CNNs and transformers and almost always supervise orientation or minutiae as an intermediate. FingerRT (Jia et al., 2024) works in two stages, Sun et al. (2024) use orientation and texture streams, and Dabouei et al. (2018) emit orientation and frequency maps alongside ridges to avoid inventing false minutiae. Others inpaint clean contact prints (Blais et al., 2020), apply diffusion to tiny partial prints (Hsu et al., 2025), or accumulate partial touches without a generative step (Guan et al., 2026). The established alternative is to acquire the missing area by mosaicking or fusing several impressions (Guan et al., 2026; Jia et al., 2025), and Section 5.3 compares against both. None of these methods uses auxiliary views or has a public contactless implementation, so we compare against a learned inpainter fine-tuned on our data and the classical pipeline.

Evaluating Generated Prints. Synthesis work pairs matcher utility with identity-leakage analyses, as in PrintsGAN (Engelsma et al., 2023), GenPrint (Grosz & Jain, 2025) and IMPOSE (Pan et al., 2026), and we adopt that practice. FingerGAN (Zhu et al., 2023) moves the adversarial signal onto the skeleton and orientation field, the instinct behind our spectral loss, and Hussein et al. (2026) measure how generated prints add and remove minutiae, which we use inside the hidden region.

Inpainting, Periodic Structure, References and Masking. Ridges are a quasi-periodic texture, which convolutional inpainting handles poorly. LaMa (Suvorov et al., 2022) addresses this with Fast Fourier Convolutions (FFC, Chi et al., 2020), convolutions that also act on the feature map's Fourier transform, and we use FFC blocks and LaMa as a baseline. MAT (Li et al., 2022) makes attention mask-aware, ZITS (Dong et al., 2022) grafts predicted structure onto a texture network by zero-initialized addition, our template for structure injection, and Jiang et al. (2021) penalize missing spectral content. Diffusion and flow restoration (Lugmayr et al., 2022; Saharia et al., 2022; Xia et al., 2023; Ohayon et al., 2025) trade distortion for perceptual quality by sampling, whereas we keep a deterministic regression model. Among reference-guided inpainters, LeftRefill (Cao et al., 2024a) finds shared self-attention over stitched views sufficient, TransRef (Liu et al., 2025) gains most from explicit patch alignment on uncanonical views and MVInpainter (Cao et al., 2024b) finds that dense flow guidance can hurt, and because our crops are geometrically normalized we rely on joint attention alone. Masked autoencoders (He et al., 2022) add positional embeddings *after* inserting mask tokens so that masked positions stay distinguishable, and Section 3.3 shows why a generator must do the same and measures the cost of reversing the order.

3 Method

SweepFill is one encoder-decoder network that restores the visible part of a degraded contactless print and generates the rest. Sections 3.1 and 3.2 define the task and the backbone. Section 3.3 answers RQ2 by showing that the bottleneck must keep hidden tokens distinct before anything can be

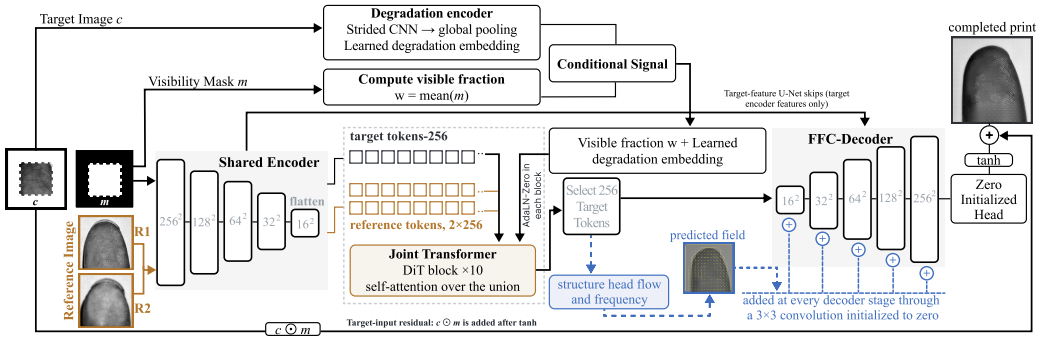

Figure 2: **SweepFill.** Each slab is one resolution stage, labeled by its size, black arrows are the main dataflow and the two secondary pathways are colored. In *orange*, reference frames share the target's encoder and their tokens join its own, so every DiT block attends over all 768 and only the target's return to the grid (Section 3.6). In *dashed blue*, the structure head predicts ridge flow and frequency for the whole print, which is added at every decoder stage through a zero-initialized 3×3 convolution (Section 3.4).

generated there, and Sections 3.4 to 3.6 answer RQ1 with the three constraints the hidden region does admit, namely the print's own ridge structure, an objective that ignores unknowable phase, and the other frames of the sweep.

3.1 Task and data

Restoration and generation are one problem here, parameterized by how much of the print the observation covers. Let $x \in [-1,1]^{1\times H\times W}$ be a clean aligned fingertip crop and $m \in \{0,1\}^{1\times H\times W}$ an axis-aligned rectangular *visibility mask*, the *window*, covering a fraction w of the image. A degradation operator D applies noise, Gaussian or motion blur, downsampling, JPEG compression, or a composition of two or three of these at a per-example severity s, giving the observation

$$c \;=\; D_s(x)\odot m \;+\; \mathbf{1}\odot(1-m), \tag{1}$$

which is degraded inside the window and constant outside (parameters in Table 4, Appendix B). Given (c, m) and optionally *reference* crops of the same finger, the model estimates the clean *whole* print $\hat{x}$. We score the *visible* region m (restoration) and the *hidden* region $1-m$ (generation) separately, and $w{=}1$ is pure restoration. We use MMFV (Malhotra et al., 2023), where each finger is recorded in roll, pitch and yaw sweeps, giving 240,796 upright 256×256 crops over 1,872 sequences and 83 subjects. We split by *subject* (73 for training with 213,655 crops, 10 held out with a fixed 1,727-crop evaluation subsample), so no evaluation identity is seen in training (Appendices A and B).

3.2 Architecture

SweepFill is a U-Net (Ronneberger et al., 2015), an encoder-decoder whose skip connections pass the encoder's features to the decoder at every resolution, with a transformer bottleneck (Figure 2, widths in Appendix A). The bottleneck has 10 diffusion-transformer (DiT) blocks (Peebles & Xie, 2023), used here without diffusion, which are conditioned on the visible fraction and on a learned embedding of the degradation by adaLN-Zero (Peebles & Xie, 2023), an adaptive layer normalization (Perez et al., 2018) whose modulation starts at zero. A zero-initialized head predicts a bounded residual, $\hat{x} = c\odot m + \tanh(\mathrm{head}(\cdot))$, so the model starts as the identity inside the window. The skips matter because a 16×16 latent cannot represent a 6 to 10 px ridge period, so visible ridges are copied and cleaned at full resolution. The same network with the conventional mask-token order and none of the components of Sections 3.3 to 3.6 is the *DiT-UNet restorer*, a strong comparator we train under every protocol.

3.3 Hidden tokens must stay positionally distinct

The bottleneck works on *tokens*, one feature vector per 16×16-px cell. A token whose cell contains no visible pixel carries no evidence, so the model replaces it with a learned mask token μ, and where that token enters matters (Figure 4, Appendix C). SweepFill inserts it before adding the positional embedding p_i, which leaves $t_i = \mu + p_i$ for every fully hidden cell $i \in \mathcal{H}$, the mask-first order of masked autoencoders (He et al., 2022). Inserting it after leaves $t_i = \mu$, so every hidden token is *the same vector*, and that survives the whole transformer.

Proposition. *For any layer f of multi-head self-attention, position-wise MLPs and normalization with position-shared parameters (a standard transformer block, including position-independent adaLN), $t_i = t_j \Rightarrow f(T)_i = f(T)_j$* (Appendix C). Hence, if all hidden tokens start at μ, the bottleneck emits a *constant* over the hidden region for any input. We measure this on two equal-budget models that differ only in this order (Figure 5, Appendix C). With position first, all 156 hidden tokens have pairwise cosine similarity 1.0 after ten blocks, and the structure head reads ridge orientation to within $4.6°$ where the print is visible but only $35.9°$ far inside the hidden region, against $45°$ for chance, while with the mask token first it reaches $5.9°$ at identical cost. The effect is not specific to this model, since a small masked-bottleneck transformer on natural textures collapses the same way (Appendix C.2), and rotary or relative-position attention, which inject position in every layer, also avoid it (Appendix C.1), so the requirement is on the order of operations, not on one embedding.

3.4 Structure as an explicit intermediate

Ridge flow and period are smooth, low-dimensional fields that a partial observation strongly constrains (Wang & Hu, 2011; Feng et al., 2013). A three-layer head on the bottleneck tokens therefore predicts, for every 16×16 block of the *whole* print, the ridge orientation θ as $u = (\cos 2\theta, \sin 2\theta)$ (doubling the angle makes θ and $\theta{+}180°$ the same), the log ridge frequency, and a coherence logit, where *coherence* measures how clearly one orientation dominates the block. The ground truth is computed on the fly from the clean target, orientation and coherence by a structure tensor (the local average of image-gradient products) and frequency by a windowed FFT. Let ρ be the target's coherence and h the hidden fraction of a block. The loss weights each block by $\rho' = \rho(1 + \gamma h)$ with $\gamma{=}2$, so coherent blocks count more and hidden ones more still, and its frequency term is gated by the spectral-peak confidence κ so that ridge-free skin teaches nothing,

$$\mathcal{L}_{\text{str}} = \big\langle \|\hat{u} - u\|^2 \big\rangle_{\rho'} + \big\langle |\log \hat{f} - \log f| \big\rangle_{\rho' \cdot \mathbb{1}[\kappa > \tau]} + \tfrac{1}{2}\,\mathrm{BCE}(\hat{\rho}, \rho). \tag{2}$$

The field enters every decoder stage through a zero-initialized convolution (Figure 2), following ZITS (Dong et al., 2022), so the pathway initially changes nothing. The head's larger effect is on the bottleneck it supervises, because hidden-cell tokens must then encode the print's flow and period, and removing the head costs several degrees of hidden-region flow (Section 5.1, Appendix K.2).

3.5 Phase-free spectral loss

Knowing the field is useless if the objective punishes the decoder for drawing ridges. Under a pixel loss, a ridge pattern that flows correctly but sits half a period out of place costs as much as flat skin, so a regression model learns to draw flat skin. We therefore compare the *amplitude* of the Fourier spectrum instead, which carries orientation, period and contrast but not phase. For 32×32 Hann-windowed patches $\mathcal{P}$ lying mostly in the hidden region, we compare log-amplitude spectra, restricted by a mask B to the band of plausible ridge periods,

$$\mathcal{L}_{\text{spec}} = \frac{1}{|\mathcal{P}|} \sum_{P \in \mathcal{P}} \big\| B \odot \big(\log(1 + |\mathcal{F}\hat{x}_P|) - \log(1 + |\mathcal{F}x_P|) \big) \big\|_1. \tag{3}$$

so a correctly flowing pattern half a period off scores well. B keeps periods of 4 to 10.7 px, and a patch enters $\mathcal{P}$ when more than half of its central 16-px block is hidden. The gains hold across ridge-period terciles and a 15% magnification of the crops (Appendix F.3). The decoder's residual blocks are FFC blocks as in LaMa, but the ablation of Section 5.1 credits the gain to the loss, not to this choice of block.

3.6 Sweep references and objective

For each target frame we sample $R{=}2$ reference crops of the same finger and session, preferring a *different* pose sweep and otherwise a frame at least 8 frames away, and encode them with the shared encoder. Their bottleneck tokens, tagged with a learned per-view embedding, join the target's (Figure 2), and an attention mask switches off references dropped during training (probability 0.25 each). This is the simplest reference-guided design (Cao et al., 2024a) and needs no alignment or flow, since a coarse affine prior trained in changes nothing (Appendix I). Target and references share the same geometric augmentation, so only the finger's real pose change separates them. Because every MMFV frame shows the whole fingertip, the window is synthetic and the references always contain the hidden region, the favorable end of the sweep setting, and we bound the other end by degrading the references (Table 13, Appendix E), removing them (Table 35, Appendix K.3) and occluding them (Appendix D.1). The objective scores the two regions separately, since hidden-region error is an order of magnitude larger and would otherwise dominate the gradient,

$$\mathcal{L} = \ell(\hat{x}, x; m) + \lambda_h\, \ell(\hat{x}, x; 1-m) + \lambda_s \mathcal{L}_{\text{MS-SSIM}} + \lambda_{\text{str}} \mathcal{L}_{\text{str}} + \lambda_{\text{spec}} \mathcal{L}_{\text{spec}} + \lambda_{\text{aux}} \mathcal{L}_{\text{aux}}, \qquad (4)$$

where ℓ combines the Charbonnier loss (a smooth L_1), squared error and a Sobel-gradient term over the given mask, $\mathcal{L}_{\text{MS-SSIM}}$ is the multi-scale structural-similarity loss, and $\mathcal{L}_{\text{aux}}$ supervises the degradation encoder. Weights and the difficulty ramp over visibility and severity are in Appendix A. Equation (4) is the *entire* training signal, and no discriminator, perceptual network or diffusion process enters SweepFill or any ablation. Where generative fingerprint work buys the perceptual side of the tradeoff (Blau & Michaeli, 2018) with a GAN (Engelsma et al., 2023; Zhu et al., 2023; Grosz & Jain, 2025), SweepFill changes *what* is predicted and *how* it is scored, so added ridges are constrained by the observation and the references.

4 Evaluation protocol

We answer RQ3 with experiments at two levels, because image metrics show where a completion fails but not whether it is worth anything to a matcher, while identity metrics show whether it is but not why (Engelsma et al., 2023; Grosz & Jain, 2025; Pan et al., 2026). Appendix B gives the full specification.

Image Level. PSNR and SSIM say little about whether ridges were drawn at all, so over the hidden region we add three measures. They are computed on 16-px blocks that are ridged in the ground truth (coherence above 0.25) and hidden together with an 8-px margin, so that no visible pixel or window edge enters a scored block's estimate, and both estimators pad by reflection at the crop border. *Orientation error* (OF in tables) is the mean angle between the ridge flow of the output and that of the ground truth. The flat fill of the input scores chance, $45°$, and orientation extractors on clean contact prints report errors of a few degrees (Schuch et al., 2017), a reference for scale rather than a floor. *Frequency MAPE* is the mean relative error of the dominant ridge period. *Band contrast* ("band" in tables) is the ratio of the spectral amplitude at ridge periods in the output to that of the ground truth, so 1 is true contrast and 0 is flat skin, and because energy at the wrong orientation also counts, Appendix K.1 repeats it inside a $\pm 20°$ wedge around the true ridge direction.

Identity Level. A matcher compares a *probe* against enrolled *gallery* images. Our gallery and probes come from *different recording sessions*, in the spirit of a Type-II comparison (Wijewardena et al., 2023), so a completion is never matched against the impression it came from. Each of the 32 held-out (subject, finger) identities gives three clean gallery crops from one session and three probes from the other, degraded and windowed at $w \in \{1.0, 0.7, 0.5, 0.3\}$, and the model's references come only from other sweeps of the *probe* video. We report EER, the true-accept rate when 1% of impostors are falsely accepted (TAR at 1% FAR), and rank-1, the fraction of probes whose best-scoring gallery identity is their own. Two independent matchers are used, **Ridgeformer** (Pandey et al., 2025), a deep network trained on contactless data, and **LEADER+Bozorth3**, the learned LEADER minutiae extractor of pyfing (Cappelli, 2026b) followed by NBIS Bozorth3 minutiae matching (NIST, 2010), both after identical C2CL-style preprocessing (Grosz et al., 2022; Jawade et al., 2022). Intervals come from a bootstrap that resamples whole identities (2000 resamples, Appendix B.1). We measure every number ourselves with the cited implementations and transcribe none.

Minutiae Fidelity and Leakage. A matcher's gain does not say whether the hidden region holds the *right* detail, so, following Hussein et al. (2026), we match the minutiae of the completion against

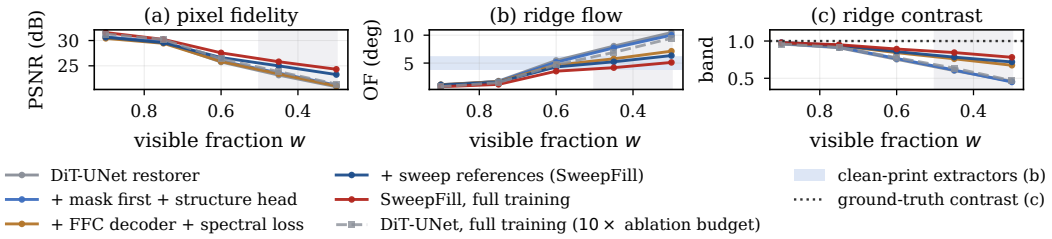

Figure 3: **Hidden-region quality against the visible fraction** on the validation sweep, solid at equal training budget and dashed with full training. Flow and contrast (b, c) are scored on blocks hidden with an 8-px margin, which exist for $w \leq 0.6$. At every window the restorer's hidden region has the contrast of flat skin, while SweepFill draws ridges at about two thirds of true contrast with half the orientation error. The gray band marks more than half hidden, the blue band in (b) where orientation extractors plateau (Schuch et al., 2017), and the dotted line in (c) true contrast.

those of the ground-truth frame inside the hidden region (15 px, 30°) and report the fraction of true minutiae *recovered* and of produced minutiae that are *spurious*. We then ask whether completion manufactures identity, that is, how often a completion is accepted against a *different* identity at the 1%-FAR threshold, and how often two completions of different identities match each other.

Baselines. On identical probes we compare against *(i)* the degraded partial probe itself, which a deployed system would otherwise match, *(ii)* a classical pipeline that diffuses the orientation field into the hidden region and grows ridges by contextual Gabor filtering (Hong et al., 1998; Feng et al., 2013), *(iii)* LaMa (Suvorov et al., 2022), zero-shot and fine-tuned on our crops and masks at the equal budget, under our pixel losses and under its own (Appendix H), and *(iv)* mosaicking, which registers the same two sweep frames SweepFill reads to the probe's window and pastes them in (Appendix K.4).

5 Experiments

This section answers RQ3 under that protocol. The *equal budget* is eight epochs over a 66k-crop subsample with identical optimizer, schedule, loss weights and hardware, so every *rung* of the cumulative ablation, each model adding one component, sees the same data (Appendix A), and SweepFill is then trained on all 213,655 crops for 20 epochs. All numbers come from the 10 held-out subjects, image metrics from the *validation sweep*, a fixed 1,727-crop sample of them, with mixed degradation at severity 0.5 unless noted.

5.1 Component ablation

The components divide by what they buy (Figure 3, and Table 29 in Appendix K.1, at w=0.3). The phase-free spectral loss makes the qualitative change, taking hidden-region orientation error from 21.1° to 14.8° and band contrast from 0.12 to 0.52, so flat skin becomes ridges. The structure head pays off once the decoder is allowed to draw, since removing it from the full model costs 17.4° against 13.2°, and it acts through the bottleneck it supervises rather than through its injected field, which the decoder barely reads (Appendix K.2). Mask-first tokenization is what lets the head read the hidden region at all, and the full model with the conventional order reaches 16.4°. References add 2.3 dB and a further step in flow, and the FFC decoder changes nothing measurable (13.3° without it). What the references contribute is specific to the finger. Removing them at test time costs 2.8 dB, yet SweepFill stays 6.8° and +0.39 band ahead of the fully trained DiT-UNet restorer, and a different subject's sweep recovers almost none of the PSNR and none of the orientation loss (Table 14, Appendix E).

5.2 Comparison and baselines

At w=0.3, SweepFill reaches 24.31 dB, 11.01° and 0.67 band in the hidden region against 21.39 dB, 20.07° and 0.15 for the fully trained DiT-UNet restorer (Figure 3, per-window numbers in Table 27, Appendix K.1). The pixel gap widens as the window shrinks, the signature of structure and references rather than texture extended past the boundary. The completion baselines trail further (Table 30,

Table 1: **Identity evaluation on MMFV** (Malhotra et al., 2023), cross-session, with 32 held-out identities, 96 probes and two matchers, the deep matcher Ridgeformer (Pandey et al., 2025) and the minutiae matcher LEADER+Bozorth3 (Cappelli, 2026b; NIST, 2010). TAR is at 1% FAR and R-1 is rank-1. Here and throughout, ↓ and ↑ give the direction of improvement, bold the best value in its comparison group, blue SweepFill or the setting it uses, and gray a ceiling rather than a competitor, and the indented row repeats SweepFill with its references degraded like the probe. Table 32, Appendix K.3, adds the baselines, Appendix B.1 the intervals.

	deep matcher (Ridgeformer)			minutiae matcher (LEADER+Bozorth3)		
probe	EER% ↓	TAR% ↑	R-1% ↑	EER% ↓	TAR% ↑	R-1% ↑
clean full-view frame *(ceiling)*	4.2	90.6	94.8	7.0	83.3	89.6
degraded probe, w=1.0	11.5	61.5	81.2	29.9	50.0	58.3
degraded probe, w=0.7	15.8	40.6	59.4	31.2	50.0	52.1
degraded probe, w=0.5	20.8	33.3	53.1	28.7	49.0	54.2
degraded probe, w=0.3	26.0	14.6	21.9	38.1	33.3	40.6
SweepFill (ours), w=1.0	**8.3**	**78.1**	**90.6**	**20.3**	**69.8**	**77.1**
SweepFill (ours), w=0.7	**7.3**	**66.7**	**88.5**	**19.0**	**64.6**	**77.1**
SweepFill (ours), w=0.5	**9.4**	**64.6**	**83.3**	**22.8**	**60.4**	**67.7**
SweepFill (ours), w=0.3	**11.6**	**57.3**	**79.2**	**23.2**	**60.4**	**64.6**
references degraded like the probe, w=0.3	14.4	52.1	77.1	25.0	58.3	67.7

Appendix K.1, and Figure 10, Appendix F). Zero-shot LaMa reaches 26.8° in the hidden region against 10.9° for SweepFill on the same probes, and the classical pipeline paints ridges in wrong directions (23.4°). Fine-tuned under our losses, LaMa draws flat skin outside the window (band 0.12), and under its own adversarial losses it draws near-true contrast (0.719) with flow still 16.81° off.

5.3 Matcher utility of the completions

Completion turns a probe the matcher largely fails on into one it passes (Table 1, and Figure 11 in Appendix K.3 for every window). With the deep matcher, completing a 30%-visible probe lowers EER from 26.0% to 11.6% and raises rank-1 from 21.9% to 79.2%. A *paired* bootstrap puts the change at -14.4 EER points ([-19.7, -6.1]) and +57 rank-1 points ([+42, +71]), and both intervals exclude zero at every window with part of the print hidden (Table 34, Appendix K.3), and at w=0.3 the intervals stay disjoint when whole subjects are resampled (Appendix B.1). The minutiae matcher agrees in direction (38.1% to 23.2% EER), and with references degraded like the probe, as in a real capture, completion still reaches 14.4% EER and 77.1% rank-1 (indented row). Other completions help the deep matcher less (Table 32, Appendix K.3). The DiT-UNet restorer reaches 19.6% EER (26.6% under the minutiae matcher, close to SweepFill's 23.2%), fine-tuned LaMa 18.6%, the classical pipeline is worse than not completing, and mosaicking the same two frames into the hidden region reaches 16.7% EER and 54.2% rank-1 (Appendix K.4). Only a restorer fine-tuned with a PatchGAN beats SweepFill, under the minutiae matcher alone (19.9% EER, Appendix I). The gain survives every robustness check we ran (Appendices B.2, D, D.1, G, J.1 and J.2).

Completion composes with multi-frame fusion. A video also yields other frames, so we compare completion with score-level fusion, taking the best score over the probe and the same two frames SweepFill receives (Table 35, Appendix K.3). Completing the poor frame *before* fusing lowers EER from 10.4% to 8.3% at w=0.3 and from 8.3% to 5.2% at w=0.7, on par with fusing two *clean* frames (6.1%). On 96 probes no fused EER gain is individually significant, while the rank-1 gains of +11.5 and +8.3 points at w=0.7 and 0.5 are ([+4.2, +19.8], [+2.1, +15.6]), and the +2.1-point gain at w=0.3 is not. With a single photograph and no video, reference-free completion takes EER from 26.0% to 17.6% and rank-1 from 21.9% to 55.2% (paired interval [-14.8, -4.0]).

5.4 Minutiae fidelity and identity leakage

Only 21% of true minutiae inside the hidden region are recovered at w=0.3 and about half of those produced are spurious (Table 2), so the gains of Section 5.3 come from restored ridge flow and texture rather than recovered minutiae (Appendix F.1). Generated flow is within 10.5° RMSE of the clean frame (46.3°, chance, for the restorer), but its phase is uncorrelated with the truth (0.011 against 0.55 where visible, Appendix F.2). Across both datasets and all windows, a completion is accepted against

Table 2: **Minutiae fidelity and impostor leakage.** The left columns give the hidden region's true minutiae that are recovered (15 px, 30°) and the produced minutiae that are spurious. The right columns give, at one 1%-FAR threshold calibrated on clean frames and held fixed, acceptance against a *wrong* identity (FA) and matches between completions of different identities (dup), on MMFV and ISPFDv2 (Malhotra et al., 2020). At w=1.0 nothing is hidden, so only the leakage columns apply.

	hidden-region minutiae			leakage, MMFV		leakage, ISPFDv2	
window	#GT/img	recovered% ↑	spurious% ↓	FA%	dup%	FA%	dup%
w=1.0	–	–	–	0.73	0.72	0.49	1.10
w=0.7	6.2	24.63	44.79	0.60	0.83	0.50	1.29
w=0.5	12.9	23.18	48.66	0.57	0.99	0.19	1.69
w=0.3	25.6	21.33	54.17	0.26	1.95	0.11	2.43

a wrong identity in 0.11 to 0.73% of cross-identity comparisons at 1% FAR, and at most 2.43% of completion pairs from different identities match, against 10.9% of uncompleted w=0.3 probe pairs and 1.3% of clean frames. Completion therefore removes the spurious similarity of near-empty probes instead of manufacturing identity (Figure 12, Appendix K.3). A *different subject's* references pull the completion toward that subject ($+0.120$, $p < 0.001$) without crossing the threshold, an attack surface (Appendix E.3).

5.5 External validity

On ISPFDv2 finger selfies (Malhotra et al., 2020), from a different device, background and protocol (Table 37, Appendix K.5), the deep matcher was trained on ISPFDv2 but not MMFV. Fine-tuned in domain on 2,980 crops from 51 unused subjects, SweepFill improves every window under both matchers, bar one deep-matcher EER tie at w=1.0. At w=0.3 the deep matcher's EER falls from 28.8% to 15.8% and the minutiae matcher's from 43.8% to 30.0%, and excluding four identities with fallback references changes little (Appendix B). Zero-shot transfer is partial (Table 5, Appendix B.1), as the deep matcher improves at w=0.3 but degrades at w=1, a gap that fine-tuning removes.

6 Discussion, limitations and conclusion

Discussion. The results separate what a partial print determines from what it leaves open. Ridge flow and period can be recovered far beyond the window, and a matcher already profits from them, since the deep matcher's EER falls by more than half at w=0.3 while only 21% of the hidden minutiae are recovered (Sections 5.3 and 5.4). Completion therefore acts as a structural prior placed in front of a matcher, and neither matcher is retrained. The capture sweep supplies this prior at no extra cost, because a phone already records the frames SweepFill reads, and the 2.3° they add appears only when they come from the same finger (Section 5.1). The most general finding is the token-ordering requirement. Any generator that replaces unseen patches with a mask token before a transformer can be checked in one forward pass, because the pairwise cosine similarity of its hidden tokens is 1.0 under the failure (Section 3.3).

Limitations. The completions are plausible rather than recovered, since their flow is close to the truth but not their phase, and few true hidden minutiae are reproduced (Section 5.4). The identity evidence is small (32 MMFV identities from 8 subjects, 92 ISPFDv2 identities), so fused EER gains are not individually significant (Section 5.3). Occlusion is synthetic, as MMFV has no truly partial frames (Appendices D and D.2). The method depends on its references, since ones occluded where the probe is hidden lose their benefit (Appendix D.1), another subject's frames pull the completion toward that subject (Section 5.4), and a new device needs in-domain fine-tuning (Section 5.5).

Conclusion. SweepFill completes degraded, partially visible contactless fingerprints from the capture sweep they came from. It predicts ridge structure before pixels, draws the hidden region under a phase-free spectral loss and reads two other frames of the sweep, on a bottleneck whose hidden tokens stay distinct. When 30% of the print is visible, the flat skin a strong restorer draws outside the window becomes ridges of nearly correct flow, and the deep matcher's EER falls from 26.0% to 11.6% without manufacturing identity. The phase gap is largely a property of the sweep (Appendix K.4), and closing it needs frames at nearby poses or a deformation model.

Ethics statement

This work generates biometric imagery. All experiments use existing research datasets under their licenses and consent procedures, and we redistribute no biometric data. Because completed prints could in principle be used to fabricate evidence or to attack a recognition system, we measure that risk directly (Section 5.4). At a 1% false-accept threshold, completions are accepted as a *different* identity in at most 0.73% of cross-identity comparisons, near-duplicate rates among completions stay at or below 2.43%, and completions built from another subject's frames drift toward that subject (0.120 in matcher similarity at w=0.3) without being accepted in any of 96 trials (one-sided 95% upper bound 3.1%), a margin we regard as thin. Generated ridges are not recovered (Section 5.4), so the output must not be presented as a person's fingerprint, least of all in forensic settings.

Reproducibility statement

Appendix A gives the architecture, optimizer and data-pipeline settings, Appendix B the evaluation protocol, including the subject split, probe construction and the preprocessing applied before each matcher, and Appendix H the versions and provenance of all third-party tools and the one deviation made in reconstructing a matcher backbone. SweepFill and every ablation are trained by one regression trainer on Equation (4), and the only adversarially trained models are the PatchGAN control of Appendix I and the LaMa baseline fine-tuned with its own losses (Appendix H). The supplementary material contains the code and data that define the data and evaluation protocols. It includes the MMFV alignment and cropping step with its configuration and the frame count of every aligned sequence, the subject split, degradation operator and visibility masks, and the identity protocol as a list of the held-out subjects and of every gallery, probe and reference frame with each probe's visible window, together with a script that renders these images without our model. It also includes the image-level metrics, the identity metrics (EER, TAR, rank-1, minutiae fidelity, impostor leakage), the bootstrap, fusion and flow-and-phase analyses, wrappers for the two matchers, the Gabor and mosaicking baselines, and the reconstruction, structure and phase-free spectral losses of Equation (4). MMFV, Ridgeformer and NBIS are not redistributed and must be obtained from their authors. The rendered windows, gallery, clean probes and references match those used in the paper exactly, and the degraded probes match up to the random noise realization. The model, the rest of its training code and the trained weights will be released with the acceptance of the paper. Meanwhile, our anonymous GitHub repository is available at `https://anonymous.4open.science/r/sweepfill_fingerprint-49D2`.

Generative AI disclosure

A large language model assistant, Claude 5.1 Fable, was used to revise text and analyze the scripts that generate the tables.

References

Marc-André Blais, Andy Couturier, and Moulay A. Akhloufi. Deep learning for partial fingerprint inpainting and recognition. In *International Conference on Image Analysis and Recognition (ICIAR)*, pp. 223–232, 2020. doi: 10.1007/978-3-030-50347-5_20.

Yochai Blau and Tomer Michaeli. The perception-distortion tradeoff. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. arXiv:1711.06077.

Chenjie Cao, Yunuo Cai, Qiaole Dong, Yikai Wang, and Yanwei Fu. LeftRefill: Filling right canvas based on left reference through generalized text-to-image diffusion model. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024a. arXiv:2305.11577.

Chenjie Cao, Chaohui Yu, Fan Wang, Xiangyang Xue, and Yanwei Fu. MVInpainter: Learning multi-view consistent inpainting to bridge 2d and 3d editing. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 37, 2024b. arXiv:2408.08000.

Raffaele Cappelli. No feature left behind: Filling the gap in fingerprint frequency estimation. *IEEE Access*, 12:153605–153617, 2024. doi: 10.1109/ACCESS.2024.3481507. benchmark and code: `https://github.com/raffaele-cappelli/FFE`.

Raffaele Cappelli. Unleashing the power of simplicity: A minimalist strategy for state-of-the-art fingerprint enhancement. *arXiv preprint arXiv:2603.19004*, 2026a. code: `https://github.com/raffaele-cappelli/pyfing`.

Raffaele Cappelli. pyfing: Fingerprint recognition in python (version 0.6). `https://github.com/raffaele-cappelli/pyfing`, 2026b. LEADER minutiae extractor; the toolkit lists its description as a paper under review.

Lu Chi, Borui Jiang, and Yadong Mu. Fast fourier convolution. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pp. 4479–4488, 2020. `https://dl.acm.org/doi/10.5555/3495724.3496100`.

Zhe Cui, Jianjiang Feng, and Jie Zhou. Monocular 3d fingerprint reconstruction and unwarping. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(7):8679–8695, 2023. doi: 10.1109/TPAMI.2022.3233898. arXiv:2205.00967.

Ali Dabouei, Sobhan Soleymani, Hadi Kazemi, Seyed Mehdi Iranmanesh, Jeremy Dawson, and Nasser M. Nasrabadi. ID preserving generative adversarial network for partial latent fingerprint reconstruction. In *IEEE International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pp. 1–10, 2018. doi: 10.1109/BTAS.2018.8698580. arXiv:1808.00035.

Qiaole Dong, Chenjie Cao, and Yanwei Fu. Incremental transformer structure enhanced image inpainting with masking positional encoding. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. arXiv:2203.00867.

Yongjie Duan, Jianjiang Feng, Jiwen Lu, and Jie Zhou. Orientation field estimation for latent fingerprints with prior knowledge of fingerprint pattern. In *IEEE International Joint Conference on Biometrics (IJCB)*, 2021. doi: 10.1109/IJCB52358.2021.9484334.

Joshua J. Engelsma, Steven A. Grosz, and Anil K. Jain. PrintsGAN: Synthetic fingerprint generator. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(5):6111–6124, 2023. doi: 10.1109/TPAMI.2022.3204591. arXiv:2201.03674.

Jianjiang Feng, Jie Zhou, and Anil K. Jain. Orientation field estimation for latent fingerprint enhancement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(4):925–940, 2013. doi: 10.1109/TPAMI.2012.155.

Steven A. Grosz and Anil K. Jain. Universal fingerprint generation: Controllable diffusion model with multimodal conditions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47 (2):1028–1041, 2025. doi: 10.1109/TPAMI.2024.3486179. arXiv:2404.13791.

Steven A. Grosz, Joshua J. Engelsma, Eryun Liu, and Anil K. Jain. C2CL: Contact to contactless fingerprint matching. *IEEE Transactions on Information Forensics and Security*, 17:196–210, 2022. doi: 10.1109/TIFS.2021.3134867. arXiv:2104.02811.

Xiongjun Guan, Jianjiang Feng, and Jie Zhou. Toward the whole picture: Accumulative fingerprint mapping and reconstruction for small-area mobile sensors. *arXiv preprint arXiv:2606.15574*, 2026.

Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. arXiv:2111.06377.

Lin Hong, Yifei Wan, and Anil K. Jain. Fingerprint image enhancement: Algorithm and performance evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8):777–789, 1998. doi: 10.1109/34.709565.

Mao-Hsiu Hsu, Yung-Ching Hsu, and Ching-Te Chiu. Inpainting diffusion synthetic and data augment with feature keypoints for tiny partial fingerprints. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 7(3):396–409, 2025. doi: 10.1109/TBIOM.2024.3517330.

Noor Hussein, Anil K. Jain, and Karthik Nandakumar. Intra-finger variability of diffusion-based latent fingerprint generation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (Foundation and Generative Models in Biometrics)*, 2026. arXiv:2604.10040.

Eddy Ilg, Özgün Çiçek, Silvio Galesso, Aaron Klein, Osama Makansi, Frank Hutter, and Thomas Brox. Uncertainty estimates and multi-hypotheses networks for optical flow. In *European Conference on Computer Vision (ECCV)*, 2018. arXiv:1802.07095.

Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. arXiv:1611.07004.

Bhavin Jawade, Deen Dayal Mohan, Srirangaraj Setlur, Nalini Ratha, and Venu Govindaraju. RidgeBase: A cross-sensor multi-finger contactless fingerprint dataset. In *IEEE International Joint Conference on Biometrics (IJCB)*, pp. 1–9, 2022. doi: 10.1109/IJCB54206.2022.10007936. arXiv:2307.05563.

Yuwei Jia, Yutang Lu, Zhe Cui, and Fei Su. FingerSplat: Contactless fingerprint 3d reconstruction and generation based on 3d gaussian splatting. *arXiv preprint arXiv:2509.15648*, 2025.

Zexi Jia, Chuanwei Huang, Zheng Wang, Hongyan Fei, Song Wu, and Jufu Feng. Finger recovery transformer: Toward better incomplete fingerprint identification. *IEEE Transactions on Information Forensics and Security*, 19:8860–8874, 2024. doi: 10.1109/TIFS.2024.3419690.

Liming Jiang, Bo Dai, Wayne Wu, and Chen Change Loy. Focal frequency loss for image reconstruction and synthesis. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021. arXiv:2012.12821.

Wenbo Li, Zhe Lin, Kun Zhou, Lu Qi, Yi Wang, and Jiaya Jia. MAT: Mask-aware transformer for large hole image inpainting. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. arXiv:2203.15270.

John Libert, John Grantham, Bruce Bandini, Kenneth Ko, Shahram Orandi, and Craig Watson. Interoperability assessment 2019: Contactless-to-contact fingerprint capture. NIST Interagency Report NISTIR 8307, National Institute of Standards and Technology, May 2020. URL `https://doi.org/10.6028/NIST.IR.8307`.

Taorong Liu, Liang Liao, Delin Chen, Jing Xiao, Zheng Wang, Chia-Wen Lin, and Shin'ichi Satoh. TransRef: Multi-scale reference embedding transformer for reference-guided image inpainting. *Neurocomputing*, 632:129749, 2025. arXiv:2306.11528.

Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. RePaint: Inpainting using denoising diffusion probabilistic models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. arXiv:2201.09865.

Aakarsh Malhotra, Anush Sankaran, Mayank Vatsa, and Richa Singh. On matching finger-selfies using deep scattering networks. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2(4):350–362, 2020. doi: 10.1109/TBIOM.2020.2999850.

Aakarsh Malhotra, Mayank Vatsa, and Richa Singh. MMFV: A multi-movement finger-video database for contactless fingerprint recognition. In *International Workshop on Biometrics and Forensics (IWBF)*, pp. 1–6, 2023. doi: 10.1109/IWBF57495.2023.10156919.

NIST. NIST biometric image software (NBIS). National Institute of Standards and Technology, 2010. Release 5.0.0; includes MINDTCT and BOZORTH3.

Guy Ohayon, Tomer Michaeli, and Michael Elad. Posterior-mean rectified flow: Towards minimum MSE photo-realistic image restoration. In *International Conference on Learning Representations (ICLR)*, 2025. arXiv:2410.00418.

Zhiyu Pan, Xiongjun Guan, Jianjiang Feng, and Jie Zhou. Identity-consistent multi-pose generation of contactless fingerprints. *arXiv preprint arXiv:2605.03830*, 2026.

Shubham Pandey, Bhavin Jawade, and Srirangaraj Setlur. Ridgeformer: Mutli-stage contrastive training for fine-grained cross-domain fingerprint recognition. In *IEEE International Conference on Image Processing (ICIP)*, pp. 2480–2485, 2025. doi: 10.1109/ICIP55913.2025.11084677. arXiv:2506.01806; code and checkpoints at `https://github.com/KNITPhoenix/Ridgeformer`.

William Peebles and Saining Xie. Scalable diffusion models with transformers. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023. arXiv:2212.09748.

David Perera, Victor Letzelter, Théo Mariotte, Adrien Cortés, Mickael Chen, Slim Essid, and Gaël Richard. Annealed multiple choice learning: Overcoming limitations of winner-takes-all with annealing. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024. arXiv:2407.15580.

Ethan Perez, Florian Strub, Harm de Vries, Vincent Dumoulin, and Aaron Courville. FiLM: Visual reasoning with a general conditioning layer. In *AAAI Conference on Artificial Intelligence*, 2018. doi: 10.1609/aaai.v32i1.11671. arXiv:1709.07871.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 9351 of *Lecture Notes in Computer Science*, pp. 234–241, 2015. doi: 10.1007/978-3-319-24574-4_28.

Christian Rupprecht, Iro Laina, Robert DiPietro, Maximilian Baust, Federico Tombari, Nassir Navab, and Gregory D. Hager. Learning in an uncertain world: Representing ambiguity through multiple hypotheses. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2017. arXiv:1612.00197.

Chitwan Saharia, William Chan, Huiwen Chang, Chris A. Lee, Jonathan Ho, Tim Salimans, David J. Fleet, and Mohammad Norouzi. Palette: Image-to-image diffusion models. In *ACM SIGGRAPH*, 2022. doi: 10.1145/3528233.3530757.

Patrick Schuch, Simon-Daniel Schulz, and Christoph Busch. Intrinsic limitations of fingerprint orientation estimation. In *International Conference of the Biometrics Special Interest Group (BIOSIG)*, pp. 1–5, 2017. doi: 10.23919/BIOSIG.2017.8053513.

Peter Shaw, Jakob Uszkoreit, and Ashish Vaswani. Self-attention with relative position representations. In *Proceedings of NAACL-HLT*, pp. 464–468, 2018. `https://aclanthology.org/N18-2074/`.

Jianlin Su, Murtadha Ahmed, Yu Lu, Shengfeng Pan, Wen Bo, and Yunfeng Liu. Roformer: Enhanced transformer with rotary position embedding. *Neurocomputing*, 568:127063, 2024. doi: 10.1016/j.neucom.2023.127063.

Yuting Sun, Xiaojuan Chen, and Yanfeng Tang. Recovery of incomplete fingerprints based on ridge texture and orientation field. *Electronics*, 13(14):2873, 2024. doi: 10.3390/electronics13142873.

Roman Suvorov, Elizaveta Logacheva, Anton Mashikhin, Anastasia Remizova, Arsenii Ashukha, Aleksei Silvestrov, Naejin Kong, Harshith Goka, Kiwoong Park, and Victor Lempitsky. Resolution-robust large mask inpainting with Fourier convolutions. In *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2022. arXiv:2109.07161.

U.S. Department of Homeland Security, U.S. Department of Justice, and White House Office of Science and Technology Policy. Biometric technology report: Submitted in fulfillment of section 13(e) of executive order 14074. Technical report, U.S. Department of Homeland Security, December 2024. URL `https://www.dhs.gov/sites/default/files/2024-12/24_1230_st_13e-Final-Report-2024-12-26.pdf`.

Yi Wang and Jiankun Hu. Global ridge orientation modeling for partial fingerprint identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(1):72–87, 2011. doi: 10.1109/TPAMI.2010.73.

Kanishka P. Wijewardena, Steven A. Grosz, Kai Cao, and Anil K. Jain. Fingerprint template invertibility: Minutiae vs. deep templates. *IEEE Transactions on Information Forensics and Security*, 18:744–757, 2023. doi: 10.1109/TIFS.2022.3229587. arXiv:2205.03809.

Bin Xia, Yulun Zhang, Shiyin Wang, Yitong Wang, Xinglong Wu, Yapeng Tian, Wenming Yang, and Luc Van Gool. DiffIR: Efficient diffusion model for image restoration. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023. arXiv:2303.09472.

Yanming Zhu, Xuefei Yin, and Jiankun Hu. FingerGAN: A constrained fingerprint generation scheme for latent fingerprint enhancement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(7):8358–8371, 2023. doi: 10.1109/TPAMI.2023.3236876. arXiv:2206.12885.

APPENDIX

A IMPLEMENTATION DETAILS

Table 3: **Hyperparameters of SWEEPFILL**, read from `run_config.json` of the SWEEPFILL run and from the loss constants in the source. Ablations differ only in the flags named in Section 5.1 and in the epoch count and subsample of the equal-budget protocol.

setting	value
data and sampling	
crops, subjects (train/held-out)	213,655, 73/10
crop size, CLAHE	256^2, on
visible fraction w	$\mathcal{U}(0.3, 1.0)$ after the ramp, 45% fully visible
degradation severity s	$\mathcal{U}(0.1, 1.0)$, ramped over 2 epochs
architecture	
encoder width, multipliers	$48, 1, 2, 4, 8, 8$
DiT blocks, width, heads	10, 768, 12
bottleneck grid, token cell	16^2, 16 px
mask token, positional embedding	mask token inserted first, $t_i = \mu + p_i$
adaLN-Zero conditioning	visible fraction w, learned degradation embedding
skip connections	target encoder features, concatenated at each resolution
structure head output	$\cos 2\theta$, $\sin 2\theta$, $\log f$, coherence; one zero-init 3×3 adapter per stage
output head	zero-initialized, $\hat{x} = c \odot m + \tanh(\text{head})$
optimization	
epochs, batch per GPU, GPUs	20, 32, 5
learning rate, warmup steps	0.0003, 800
weight decay, grad clip, EMA	0.01, 1.0, 0.999
precision, seed	bf16, 1337
loss weights	
Charbonnier, squared error	0.25, 100
gradient, MS-SSIM	0.1, 0.25
hidden region λ_h	0.3
structure λ_{str}, spectral λ_{spec}	1, 0.5
degradation auxiliary	0.1
spectral loss	
patch, window, stride	32^2, Hann, 16 px
radial bins, ridge periods	3–8, 4.0–10.7 px
patch enters $\mathcal{P}$ when	$\geq 90\%$ hidden
references	
count, drop probability	2, 0.25
minimum intra-sequence gap	8 frames

PREPROCESSING. Every third frame of each MMFV sweep is processed independently. The finger is segmented by chroma/skin thresholding at a reduced working width, its long axis estimated by PCA on the mask, the tip end disambiguated by image-border contact, and the distal phalanx rotated upright, scaled to a constant finger width and cropped to 256×256 grayscale. Frames failing a Laplacian-variance sharpness test or a mask fill test are dropped (2,525 and 4,412 respectively). A MediaPipe hand-landmark path is attempted first but fires on only 0.02% of crops, since macro shots of a single fingertip rarely contain enough of the palm, so the geometric fallback produces essentially the entire corpus. The result is 240,796 crops in 1,872 sequences over 83 subjects, CLAHE-equalized at load time.

ARCHITECTURE. The encoder and decoder have 48, 96, 192, 384 and 384 channels at resolutions 256^2, 128^2, 64^2, 32^2 and 16^2. The bottleneck is 10 DiT blocks of width 768 with 12 heads and adaLN-Zero conditioning on the visible fraction w and the degradation embedding. Tokens whose 16×16 cell holds no visible pixel are replaced by the learned mask token before the positional embedding is added ($t_i = \mu + p_i$, Section 3.3). After the stack only the 256 target tokens are reshaped to the 16^2 grid, and the decoder concatenates the target's encoder features at each matching resolution, the references contributing through attention only. The zero-initialized output head predicts a $\tanh$-bounded residual added to the masked input, $\hat{x} = c \odot m + \tanh(\text{head}(\cdot))$.

The degradation encoder is a five-layer strided CNN (width 32) pooled to a global vector, with auxiliary heads for severity (MSE) and degradation type (six-way cross-entropy). The structure head is $\mathrm{LayerNorm} \to \mathrm{Linear}(768{\to}192) \to \mathrm{SiLU} \to \mathrm{Linear}(192{\to}4)$ per token, emitting $(\cos 2\theta, \sin 2\theta)$, log ridge frequency and a coherence logit. Structure injection resizes that field to each decoder stage and adds it through one zero-initialized $3{\times}3$ convolution. FFC blocks split channels evenly between local and global branches, the FFT branch running in `float32` regardless of autocast. Reference frames are encoded by the same encoder with an all-ones mask, and their tokens carry a learned per-view embedding and are concatenated to the target's before the DiT stack, with a key-padding mask for dropped references. SweepFill has 126.2M parameters, the DiT-UNet restorer 128.4M. Figure 2 draws the two pathways.

At inference SweepFill completes one $256{\times}256$ probe in 27 ms at batch size one in bfloat16 on an RTX PRO 4500 Blackwell, peaking at 1.2 GB of GPU memory. Switching the references off changes neither figure (27 ms), since the shared encoder still runs over the reference slots.

Structure ground truth. Orientation comes from the structure tensor of the clean target, smoothed and pooled to $16{\times}16$ blocks as doubled-angle vectors with coherence. Ridge frequency comes from Hann-windowed $32{\times}32$ FFT patches on a 16-px grid, where the image is high-pass filtered to suppress illumination, the radial power spectrum is restricted to bins for periods of about 4 to 11 px, and the peak refined by parabolic interpolation in log-power. A spectral peak-to-mean confidence gates the frequency loss. All ground truth is computed on GPU inside the training step with autocast disabled.

Optimization. AdamW ($\beta = (0.9, 0.95)$, weight decay 0.01), learning rate 0.0003 with 800 warmup steps and cosine decay, gradient clipping 1.0, EMA 0.999 (EMA weights are evaluated and saved), bf16 autocast, `torch.compile`, DDP. Batch size 32 per GPU. The loss weights are Charbonnier 0.25, squared error 100, gradient 0.1, MS-SSIM 0.25, hidden-region weight $\lambda_h{=}0.3$, structure 1.0, spectral 0.5 and auxiliary 0.1. Inside the structure loss, hidden blocks are up-weighted by $\gamma{=}2$. During sampling, $w \sim \mathcal{U}(w_{\min}, 1)$ with $w_{\min}$ ramped $1.0 \to 0.3$ and severity ramped $0.1 \to 1.0$ over the first two epochs, and 45% of samples fully visible. Each example takes $R{=}2$ references, dropped independently with probability 0.25, minimum intra-sequence gap 8 frames, preferring a different pose sweep of the same finger and session. Table 3 collects every setting, emitted from the run's stored configuration. The saved checkpoint is the one with the best validation PSNR, and the validation images come from the held-out subjects. For SweepFill this is the final epoch (20 of 20), so the selection never departs from the end of training. The loss weights, spectral band and reference rule were also fixed on these subjects, which the two further held-out splits of Appendix B.2 guard against. No adversarial, perceptual or diffusion component enters SweepFill or any of its ablations, since every one is the output of a single regression trainer optimizing Equation (4). The one adversarially trained model in this paper is the PatchGAN fine-tune of the DiT-UNet restorer in Appendix I, a control.

Compute. The equal-budget ablations are 8 epochs over a 66,097-crop subsample (at most 40 crops per sequence) on three GPUs, giving 528,776 sample visits each. Equal visits are not equal compute, because the reference rung encodes two extra frames and attends over 768 tokens, costing 2.8 GPU-hours against 1.6 for the restorer rung, 1.73× (logged epoch times, three GPUs). SweepFill is 20 epochs over all 213,655 crops on five GPUs, giving 4,273,100 visits (8.1× the equal budget), roughly 13 minutes per epoch, 21.7 GPU-hours in all. The DiT-UNet restorer's full-training run totals 5,341,375 visits over a two-stage chain (10.1×), so SweepFill uses 0.80× that. All runs used RTX PRO 4500 (32 GB) GPUs. Results at every visible fraction for clean and mixed degradation appear in Table 28, Appendix K.1, and per operator in Table 20, Appendix G.

B Evaluation protocol details

Splits. Subjects are partitioned by a seeded shuffle with a 0.12 validation fraction, giving 10 held-out subjects, and no crop, sequence or reference from a validation subject is ever seen during training. The 8 of them recorded in both sessions supply the 32 (subject, finger) identities of the identity protocol.

Probe construction. Section 4 gives the gallery and probe counts. The probe frames are spread over the middle 60% of the sweep, degraded with the "mixed" operator at severity 0.5 and windowed with a *centered* rectangle. Every method receives the identical degraded and windowed image, so all

Table 4: **Degradation model.** Each training example draws a visible fraction $w \sim U(0.3, 1)$ (fully visible with probability 0.45), a severity $s \sim U(0.1, 1)$, and one operator or a composition, sampled with the listed probabilities. Evaluation uses the composition at s=0.5 unless stated.

operator (probability)	parameters at severity s
none (0.22)	exact identity pair
Gaussian noise (0.18)	$\sigma \sim U(0.02, 0.11)\ s$
blur (0.16)	Gaussian, $\sigma \sim U(0.6, 2.2)\ s$; with prob. 0.35 motion, length $U(5, 15)\ s$ px, random angle
downsampling (0.14)	factor $1 + U(0.4, 3.0)\ s$, area filter down, random filter up
JPEG (0.12)	quality $U(75 - 45s,\ 92 - 20s)$
composition (0.18)	two or three of the above in random order (the evaluation setting)

systems in Table 1 and in Table 30, Appendix K.1, see the same input. Model references come only from other pose sweeps of the probe's own video.

Matcher preprocessing. Contact-trained tools assume 500 ppi and dark ridges on a light ground, while our crops are $\approx$ 320 ppi finger photographs, so before either matcher every image is inverted, enhanced (segmentation, orientation and frequency estimation, then contextual Gabor filtering) and rescaled by 1.56. Minutiae are extracted with a learned end-to-end extractor on the enhanced image and converted to `.xyt` for Bozorth3, while the deep matcher consumes the enhanced 500 ppi image directly.

ISPFDv2 adaptation. We use the `BI` subset (finger selfies of 75 subjects, 32 instances per identity). Fingers there lie roughly horizontal on cluttered colored backgrounds, so the chroma segmentation is re-tuned (Otsu on signed warmth $\max(C_r - C_b, 0)$ in place of Otsu on $|C_r - 128| + |C_b - 128|$, with an unchanged strict skin box), and everything downstream is the MMFV pipeline. Instances are walked in a fixed order and rejected on segmentation, ridge-visibility, fill or sharpness tests until nine usable crops per identity are found (51.9% of attempted instances are rejected, mostly outdoor and white-background captures), and 92 of 96 attempted identities from the first 24 subjects complete, with a median ridge period of 6.4 px, matching the MMFV scale. Four identities fall back to reusing enrollment crops as references and are flagged in the manifest. Because that gives the completion model gallery imagery, the protocol is also rescored without them (88 identities, Section 5.5), and the deep matcher moves by at most 0.6 EER points and the minutiae matcher by up to four, in the same direction for the raw and completed rows, so no comparison changes. At w=0.3 the fine-tuned completion then reaches 15.8% EER under the deep matcher and 29.6% under the minutiae matcher, against 28.5% and 40.1% for the raw probe.

B.1 Bootstrap confidence intervals

With only 32 MMFV identities, we quantify uncertainty by a cluster bootstrap over *identities* (not individual probes, which are correlated within a finger), recomputing the full protocol on each of 2000 resamples. Table 5 reports percentile intervals for both datasets, and Table 34, Appendix K.3 the paired test comparing the two systems directly. Fingers of one person share capture conditions, so we also resample whole *subjects* (8 clusters). At w=0.3 the deep-matcher EER interval is [21.3, 29.5] for the degraded probe against [9.1, 17.1] for the completion, and rank-1 [14.6, 30.2] against [64.6, 94.8]. The intervals stay disjoint, so the conclusion does not move.

B.2 Robustness to the held-out subject split

The identity protocol has 32 identities because 8 of our 10 held-out subjects were recorded in both sessions. Table 6 repeats the protocol on two further disjoint held-out splits (28 and 40 identities), each with a model trained at the equal-budget setting (eight epochs, 66k crops) on the remaining subjects. The first row is the published split with the equal-budget model, so all three rows are comparable (SweepFill in Table 1 trains 20 epochs on all crops). Each run draws its own degradation realizations from a freshly seeded generator, so the raw-probe rates here differ from those of Table 1 on the same split by up to three EER points, but within a run every method sees byte-identical probes, which is what the comparison rests on. Completion lowers the EER and raises rank-1 on every split at w=0.3, by 9 and 12 EER points on the first two and by 3 on the third, where the raw probes are already easier (25.8% raw EER) and the rank-1 gain (23 $\rightarrow$ 37%) is the clearer signal. At w=0.5 the

Table 5: **Cluster bootstrap over probe identities** (2000 resamples), both datasets. Intervals are percentile-based and bold marks the best value at each window, the ceiling excluded. On MMFV the w=0.3 completed-vs-degraded intervals are disjoint in both EER and rank-1, while the completed-w=0.3 and degraded-w=1.0 intervals overlap, so a completed 30% probe is statistically indistinguishable from an uncompleted full-view one. On ISPFDv2 the w=0.3 intervals are likewise disjoint, so the zero-shot transfer gain at the hard end is not a sampling artifact, whereas at w=1.0 the ordering reverses, which is the domain-gap effect of Section 5.5.

probe	EER% ↓	95% CI	rank-1% ↑	95% CI
MMFV, 32 *identities, deep matcher*				
clean full-view frame *(ceiling)*	4.17	[1.53, 7.29]	94.8	[89.6, 99.0]
degraded probe, w=1.0	11.46	[7.38, 15.04]	81.2	[72.9, 88.5]
degraded probe, w=0.7	15.79	[13.52, 18.73]	59.4	[49.0, 69.8]
degraded probe, w=0.5	20.83	[15.62, 25.00]	53.1	[41.7, 65.6]
degraded probe, w=0.3	26.04	[21.88, 31.15]	21.9	[12.5, 33.3]
Gabor extrapolation (Hong et al., 1998), w=0.7	14.58	[12.50, 17.78]	54.2	[41.7, 65.7]
Gabor extrapolation, w=0.5	25.00	[19.12, 32.17]	35.4	[21.9, 50.0]
Gabor extrapolation, w=0.3	40.51	[34.38, 48.27]	7.3	[1.0, 15.6]
LaMa (zero-shot) (Suvorov et al., 2022), w=0.7	11.61	[10.13, 14.97]	75.0	[65.6, 83.3]
LaMa (zero-shot), w=0.5	17.83	[14.58, 23.96]	55.2	[43.8, 65.6]
LaMa (zero-shot), w=0.3	22.75	[17.73, 28.12]	11.5	[4.2, 19.8]
LaMa, fine-tuned, w=0.7	14.58	[11.46, 16.77]	70.8	[60.4, 79.2]
LaMa, fine-tuned, w=0.5	16.67	[12.50, 20.33]	67.7	[58.3, 77.1]
LaMa, fine-tuned, w=0.3	18.58	[14.88, 21.89]	40.6	[29.2, 52.1]
LaMa, fine-tuned with its own losses, w=0.7	9.38	[6.32, 11.79]	72.9	[64.6, 81.2]
LaMa, fine-tuned with its own losses, w=0.5	13.54	[8.43, 18.75]	70.8	[59.4, 81.2]
LaMa, fine-tuned with its own losses, w=0.3	17.88	[13.54, 23.96]	46.9	[35.4, 58.3]
SweepFill (ours), w=1.0	**8.33**	[5.21, 12.52]	**90.6**	[84.4, 95.8]
SweepFill (ours), w=0.7	**7.29**	[5.14, 11.02]	**88.5**	[80.2, 94.8]
SweepFill (ours), w=0.5	**9.38**	[6.30, 14.18]	**83.3**	[72.9, 91.7]
SweepFill (ours), w=0.3	**11.63**	[9.34, 17.73]	**79.2**	[67.7, 90.6]
ISPFDv2, 92 *identities, deep matcher*				
clean full-view frame *(ceiling)*	3.80	[2.34, 6.00]	92.4	[87.5, 96.2]
degraded probe, w=1.0	**10.33**	[8.15, 13.04]	**68.5**	[61.4, 75.5]
degraded probe, w=0.7	**14.13**	[11.41, 17.39]	**58.7**	[51.6, 65.2]
degraded probe, w=0.5	17.41	[14.35, 19.57]	37.0	[28.8, 44.6]
degraded probe, w=0.3	28.77	[26.31, 31.99]	14.1	[9.2, 20.1]
SweepFill (ours), w=1.0	13.59	[10.41, 16.85]	59.2	[52.2, 65.8]
SweepFill (ours), w=0.7	15.22	[11.95, 17.94]	52.7	[46.2, 59.2]
SweepFill (ours), w=0.5	**15.32**	[13.59, 17.94]	**44.6**	[37.5, 51.6]
SweepFill (ours), w=0.3	**17.93**	[14.52, 21.74]	**31.0**	[24.5, 37.5]

third split shows a rank-1 gain without an EER change. The direction replicates while the size varies with the subjects, which is why the main text reports paired bootstrap intervals.

Table 6: **Three disjoint held-out splits**, equal-budget models, Ridgeformer EER and rank-1 of the raw partial probe and of the completion.

		w=0.7				w=0.5				w=0.3			
held-out split	ident.	EER% raw / SweepFill		R-1% raw / SweepFill		EER%		R-1%		EER%		R-1%	
split A (published, seed 1337)	32	15.6	11.5	71.9	81.2	18.5	13.5	49.0	78.1	29.1	20.0	27.1	58.3
split B (seed 2024)	28	17.9	13.1	69.0	82.1	20.2	15.3	54.8	78.6	32.0	20.2	15.5	47.6
split C (seed 7)	40	17.5	14.2	63.3	64.2	18.3	18.3	44.2	62.5	25.8	22.5	22.5	36.7

C Hidden-token collapse under post-positional masking

The proposition of Section 3.3 follows because the output at position i depends only on t_i and the multiset $\{t_1, \ldots, t_N\}$, and induction over depth extends it to the stack. Figure 5 measures it on the

validation sweep at w=0.3 with two structure-head models of the ablation ladder (eight epochs on the 66k-crop subsample each), which share parameter counts, data, schedule and compute and differ only in whether μ is added before or after p_i. Appendix C.1 adds the same measurement for rotary and relative-position attention.

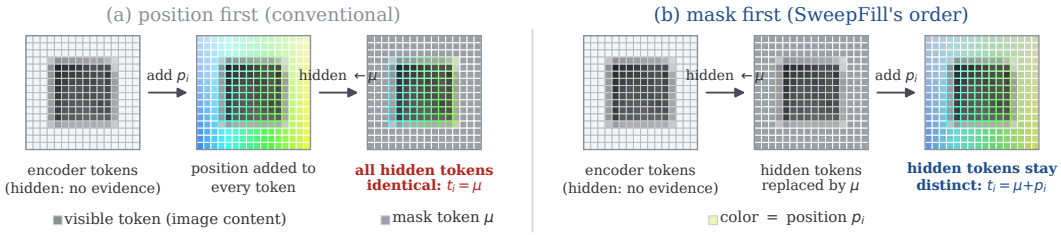

Figure 4: **Where the mask token enters**, on the real 16×16 token grid of a w=0.3 probe. Shading is image content, color the positional embedding p_i, and flat gray the mask token μ. (a) Inserted *after* position, all 156 hidden tokens equal μ. (b) Inserted *before*, they stay $\mu + p_i$ and differ from cell to cell (measured in Figure 5).

C.1 Position-encoding variants with rotary and relative-position attention

The proposition of Section 3.3 assumes position is added once, at the input. Rotary embeddings (Su et al., 2024) and relative-position biases (Shaw et al., 2018) inject position inside every attention layer through the query-key product, so two hidden tokens with identical content receive different attention weights and the collapse cannot occur. We test whether that suffices at the rung of Figure 5, Appendix C, that is, the structure-head model at equal budget (eight epochs, 66k crops, no references), under four choices of where position enters, namely a learned absolute embedding added before or after the mask token, no input embedding with 2-D axial RoPE, and no input embedding with a learned per-head relative bias over grid offsets. Table 7 reports the hidden-token cosine, the structure head's orientation error by block class and the output quality. The texture toy of Appendix C.2 repeats the four choices (Table 8), where both in-attention schemes match the input-side fix.

Table 7: **Four ways to give hidden tokens a position**, structure-head rung, equal budget, w=0.3. Cosine is the mean pairwise cosine among the fully-hidden bottleneck tokens after the last block (1.0 = identical). OF is the orientation error over ridged blocks read from the structure head and measured on the output, by distance from the window in 16-px blocks.

	hidden-token	structure-head OF (°)			output OF (°)	hidden region	
where position enters	cosine	visible	hidden ≤ 2	hidden > 2	hidden > 2	PSNR ↑	band
position first, learned absolute (conventional)	1.000	4.6	32.3	35.9	9.3	21.04	0.152
mask first, learned absolute (SweepFill)	0.599	4.4	6.4	5.9	9.2	21.14	0.122
no input embedding, 2-D RoPE in attention	0.626	4.4	6.2	5.7	9.1	21.33	0.113
no input embedding, relative position bias	0.750	4.7	6.7	6.5	9.6	21.01	0.126

Both in-attention schemes remove the collapse completely, since the hidden tokens are no longer identical (cosine 0.626 and 0.750 against 1.0), and the structure head reads ridge flow far inside the hidden region at 5.7° (RoPE) and 6.5° (relative bias), against 5.9° with the mask token inserted before a learned absolute embedding and 35.9° with it inserted after, the error of the single constant orientation a collapsed bottleneck can emit, against 45° for a uniformly random guess. Hidden-region PSNR and band are within noise across the three non-collapsed variants, so combining two remedies is untested, because once the tokens are distinct, the readout sits at the floor set by the visible-region error. The failure is specific to absolute embeddings added once at the input, the most common construction in masked-bottleneck generators, and any of the three remedies is free, and the point is the order of operations, not a preference for one embedding. Output orientation error far from the window is similar for all four models, the collapsed one included, because the decoder's skip connections extend visible ridges regardless of what the bottleneck carries. The difference shows in what the structure head can read, and downstream in the rungs that use that field (Figure 3).

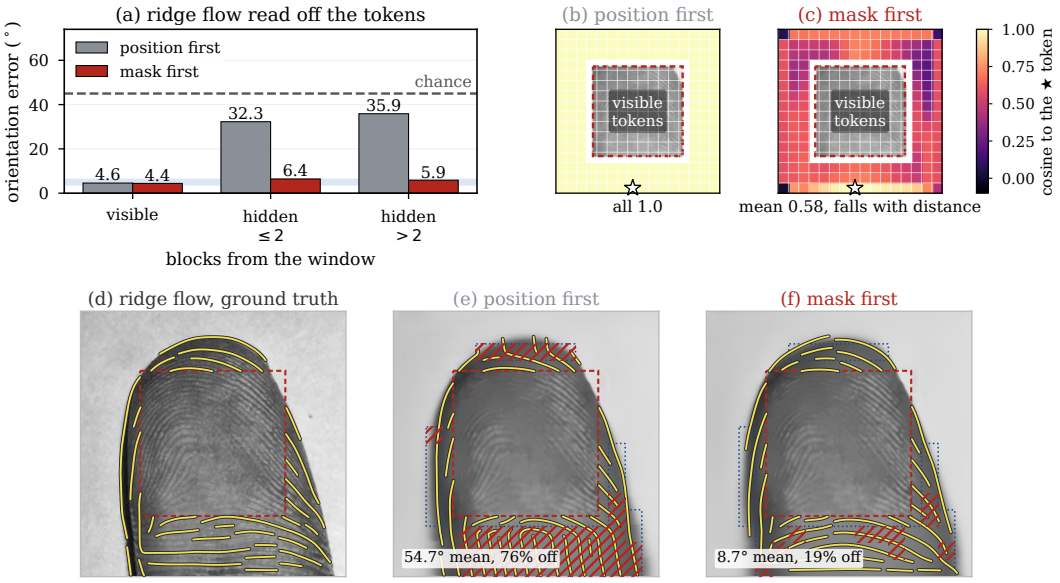

Figure 5: **What the order costs**, on two equal-budget structure-head models that differ only in it, on the probe of Figure 4. (a) Error of the model's own ridge-flow estimate read off the tokens, by distance from the window. Chance is $45°$, and the shaded band is the 4 to $6°$ plateau of the best extractors (Schuch et al., 2017). (b,c) Cosine of every fully hidden token to one reference token (★) after all ten blocks, and cells inside the dashed window hold visible input and are excluded. Position first makes every hidden token the same vector, so the map is identically 1.0, while mask-first similarity falls with distance, so those tokens carry their location. The number under each map is that map's own mean, over the cosines to the ★ token, whereas Table 7, Appendix C.1 reports the mean over all pairs of hidden tokens instead. (d-f) Streamlines of the orientation field the tokens encode, outside the window, drawn as in Figure 13, Appendix K.6. Position first predicts one orientation ($169°$) at all 156 hidden tokens, so its streamlines run parallel, whereas mask-first recovers the fingertip arcs of (d). As in Figure 1, hatched blocks of the scored hidden finger region (dotted) have encoded flow more than $15°$ from the ground truth's, and each panel states the mean error and the share of blocks off.

C.2 Collapse on natural textures

Section 3.3 argues that inserting the mask token after the positional embedding collapses every fully hidden token to one vector in *any* masked-bottleneck transformer. We test the claim on natural texture photographs (DTD, 5076 training images) with a small U-Net whose 8×8 bottleneck is a four-block transformer, trained for 3000 steps on 128-px crops with the paper's rectangular windows in four otherwise identical variants (Table 8), namely a learned absolute embedding with the mask token inserted after or before it, and no input embedding with position injected inside every attention layer, by 2-D rotary embedding or by a learned relative-position bias (Appendix C.1). An auxiliary linear head reads each block's mean intensity and dominant gradient orientation off the tokens. Only the position-first order collapses (cosine 1.000, readout $34.4°$ far from the window against $9.1°$ where the crop is visible), whereas the other three are indistinguishable on this toy (readout 14.8, 14.9 and $15.3°$ far inside the hidden region), so in-attention position is a valid alternative where the model is small enough to relearn location from the visible tokens. All four variants see identical data and reach the same hidden-region PSNR, so the collapse shows only in the tokens and in what the auxiliary head reads off them.

D Occlusion model

The protocol hides part of the print with a centered axis-aligned rectangle, the shape training used, which a deployment would not. This section relaxes that shape, then the window position and the assumption that the references image what it hides, and finally asks whether MMFV contains real partial captures. We replace the rectangle at test time, at the same area fraction and with the same

Table 8: **Where position enters, on DTD textures**, held-out crops at w=0.3, reporting the mean off-diagonal cosine among fully hidden tokens, the auxiliary-readout orientation error by distance from the window and the hidden-region PSNR.

	hidden-token	readout orientation error (°)			hidden
token order	cosine	visible	hidden ≤ 2	hidden > 2	PSNR (dB) ↑
position first (conventional)	1.000	9.1	35.7	34.4	17.51
mask first (SweepFill's order)	0.911	8.3	12.0	14.8	17.52
no input embedding, 2-D RoPE	0.924	7.5	11.6	14.9	17.59
no input embedding, relative bias	0.972	9.3	12.1	15.3	17.49

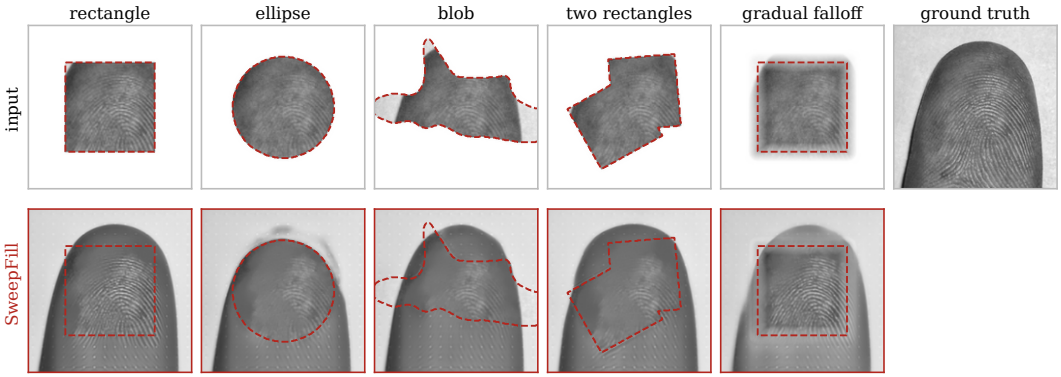

Figure 6: **Windows the model never saw.** One probe at w=0.3 under each mask shape (the top row shows the input as the model sees it, the bottom row SweepFill's completion, and the dashed line the window boundary, taken at half visibility for the fading one), ground truth at the right.

degradation draws, by an ellipse, an irregular blob (a thresholded smooth random field), the union of two rotated rectangles, and a rectangle whose visibility fades linearly to the constant over a 31-px band straddling its edge (the model receives the soft mask as its visibility channel, and metrics use the half-visibility contour). Figure 6 shows one probe under each, and Table 9 gives the identity and hidden-region numbers, with the rectangle from Table 1 for reference. Completion helps under every shape. At w=0.3 the deep matcher's EER falls from 30.2 to 16.7% for the ellipse, 38.5 to 19.9% for the blob, 30.2 to 15.6% for the two rectangles and 25.0 to 12.5% for the fading window, against 26.0 to 11.6% for the rectangle, but unseen shapes cost it, since orientation error rises from 10.9° to 13.5 to 16.8°, and under the minutiae matcher at w=0.3 the blob's gain shrinks to a third of the rectangle's while the ellipse's reverses, the completion scoring worse than the raw probe (Table 9). The blob is the hardest case for the raw probe, its visible area being scattered. The last two column pairs retrain the equal-budget two-reference model on rectangles only and on all five shapes, and test both on every shape. Training on the shapes closes the gap. At w=0.3 the ellipse goes from 24.0 to 15.8% EER and the blob from 26.0 to 16.7%, level with the rectangle (16.5%), with orientation errors of 15.1 to 14.3° against 19.1 to 16.5°, while the rectangle itself is unchanged. Mixed-shape training is as stable as rectangle training, since the spectral and structure terms of Equation (4) end at 0.108 and 0.559 against 0.111 and 0.562 for an equal-budget rectangle run, with monotone validation curves throughout. The sensitivity is one of training distribution, not of architecture or objective, so a deployment expecting irregular occlusion should train with it.

D.1 Effect of window position and reference coverage

The identity protocol fixes two things a deployment would not, namely that the window is centered and that the references always image the hidden region (Section 3.6). Table 10 relaxes each in turn, on the fully trained SweepFill, probes and protocol of Section 5.3.

Window position. Training draws the window at a random position and aspect ratio, whereas evaluation centers it. Drawing it at test time as in training leaves different content visible and makes the raw probe harder (29.2% against 26.0% EER at w=0.3 under the deep matcher). SweepFill

Table 9: **Mask shape at test time**, identity protocol, same probes and degradations as the rectangle rows. SweepFill was trained on rectangles only, while the equal-budget columns compare a rectangle-trained model with one trained on all five shapes.

	SweepFill (trained on rectangles)							equal budget, 8 epochs			
	Ridgeformer			LEADER+Bozorth3 EER%		hidden		rectangles only		all five shapes	
visibility mask	raw EER% ↓	SweepFill EER% ↓	SweepFill R-1% ↑	raw	SweepFill	OF° ↓	band	EER% ↓	OF° ↓	EER% ↓	OF° ↓
w=0.5											
rectangle (training shape)	20.8	9.4	83.3	28.7	22.8	23.1	0.640	16.7	31.0	15.6	28.6
ellipse	23.2	15.6	79.2	33.7	22.4	20.5	0.679	17.7	26.5	13.6	23.4
irregular blob	32.3	17.9	66.7	46.0	29.4	13.9	0.693	20.8	18.3	14.6	16.1
two rotated rectangles	25.1	15.6	76.0	34.4	22.3	15.7	0.704	19.8	20.0	15.6	17.8
rectangle, gradual falloff	20.8	7.5	81.2	32.4	17.9	27.1	0.503	13.5	32.8	13.5	28.0
w=0.3											
rectangle (training shape)	26.0	11.6	79.2	38.1	23.2	10.9	0.715	16.7	13.9	16.5	13.3
ellipse	30.2	16.7	50.0	28.8	30.4	16.8	0.617	24.0	19.1	15.8	15.1
irregular blob	38.5	19.9	53.1	43.6	38.2	13.2	0.661	26.0	16.5	16.7	14.3
two rotated rectangles	30.2	15.6	54.2	36.6	29.7	13.5	0.684	21.1	17.5	14.7	15.4
rectangle, gradual falloff	25.0	12.5	68.8	35.8	22.6	14.8	0.465	17.7	16.9	14.6	12.2

Table 10: **Window position and reference coverage**, identity protocol, fully trained SweepFill, 96 probes, with EER and rank-1 in percent under the deep matcher, EER under the minutiae matcher at w=0.3. Coverage is the per-probe fraction of the hidden region that the union of its occluded references images. Every row but the two random-window rows uses the published probes verbatim.

	w=0.7		w=0.5		w=0.3		minutiae
deep matcher	EER ↓	R-1 ↑	EER ↓	R-1 ↑	EER ↓	R-1 ↑	EER, w=0.3 ↓
position of the window							
degraded probe, centered window	15.8	59.4	20.8	53.1	26.0	21.9	38.1
SweepFill, centered window (the protocol)	7.3	88.5	9.4	83.3	11.6	79.2	23.2
degraded probe, random window	20.9	55.2	24.0	47.9	29.2	19.8	35.8
SweepFill, window at a random position and aspect ratio	8.3	80.2	11.5	81.2	12.5	69.8	28.6
fraction of the hidden region the reference union images							
0%, every reference occluded exactly like the probe	7.3	90.6	13.5	78.1	17.7	44.8	24.1
28%	7.3	89.6	13.5	81.2	14.7	57.3	24.1
54%	7.3	88.5	12.4	83.3	15.5	60.4	25.2
100%, references unoccluded (the protocol)	7.3	88.5	9.4	83.3	11.6	79.2	23.2
no references at all	7.5	87.5	12.5	82.3	17.6	55.2	–

moves with it, to 12.5% EER and 69.8% rank-1 against 11.6% and 79.2% with the centered window, so the completion's gain is intact, and the minutiae matcher keeps its lead (28.6% against 35.8% for the probe). The model depends on the window's position no more than on its shape.

How much the references must see. We vary how much of the hidden region the references image. Each reference is occluded by a window of the probe's size, pushed in the opposite direction for alternate references, so no single frame sees the hidden region and their union sees a measured fraction of it. At 0% every reference is blanked like the probe, and the two intermediate rows measure 28% and 54%. The result is graded rather than brittle, as Section 3.6 anticipates. Where little is hidden the references are irrelevant, since EER at w=0.7 is 7.3% at every coverage. At w=0.3 rank-1 rises monotonically with coverage, from 44.8% through 57.3% and 60.4% to 79.2%, so the references contribute exactly the part of the hidden region they image. With no coverage the completion falls back to the reference-free model in EER (17.7% against 17.6%) but not in rank-1, which drops from 55.2% without references to 44.8% with blanked ones, so a frame that images nothing useful costs something rather than nothing. A deployment that can detect the case should drop the references instead, and the single-frame path of Table 35, Appendix K.3 still improves on the degraded probe. The bracket is conservative, because a real adverse sweep occludes a different region in each frame, so their union would cover more than the aligned windows drawn here.

D.2 Naturally partial frames in MMFV

The window of the protocol is synthetic (Section 3). To look for a real counterpart we re-ran segmentation and alignment on every third probe-session frame of the held-out subjects, including the frames the preprocessor had rejected, and measured how much of the aligned canvas lies inside the photograph. Of the 14,498 frames, none has less than 80% of the canvas inside the image, while 263 have between 80 and 90%, and in all of them the finger itself is whole, the deficit being the canvas

Table 11: **References at test time**, SweepFill, validation sweep (held-out subjects, centered window, mixed degradation at severity 0.5, 512 probes), hidden region, reporting PSNR in dB, orientation error (OF) in degrees and band contrast relative to the ground truth.

	w=0.5			w=0.3		
references at test time	PSNR ↑	OF° ↓	band	PSNR ↑	OF° ↓	band
no reference	24.43	25.79	0.537	21.50	13.32	0.544
one cross-pose reference	26.06	22.42	0.611	23.99	11.37	0.658
two cross-pose references (default)	26.30	21.74	0.619	24.30	11.02	0.671
two same-sweep references, ±4 frames	27.06	21.05	0.628	25.08	10.39	0.680
two same-sweep references, ±12 frames	26.80	21.27	0.625	24.76	10.62	0.677
two same-sweep references, ±32 frames	26.50	21.57	0.619	24.42	10.96	0.670
two same-sweep references, ±64 frames	26.57	21.56	0.622	24.49	10.85	0.674
two same-sweep references, ±96 frames	26.51	21.47	0.621	24.53	10.84	0.675

Table 12: **Number of references in training**, equal budget (eight epochs, 66k crops), validation sweep, with identity columns from the protocol of Section 5.3.

	w=0.6			w=0.45			w=0.3			identity, w=0.3	
references in training and test	PSNR ↑	OF° ↓	band	PSNR ↑	OF° ↓	band	PSNR ↑	OF° ↓	band	EER% ↓	R-1% ↑
R=0	25.83	28.74	0.541	23.43	17.57	0.551	21.01	14.84	0.518	17.7	46.9
R=1	26.26	27.99	0.550	24.53	16.91	0.584	22.84	14.03	0.568	14.8	56.2
R=2 (SweepFill)	26.62	26.97	0.566	24.99	16.36	0.606	23.26	13.17	0.595	15.9	57.3
R=4	26.83	25.86	0.586	25.24	15.85	0.635	23.49	13.03	0.620	15.6	63.5

margin beyond a fingertip that touches the image border, which also makes the alignment unreliable. Real specular loss is equally rare. Of the 4,800 aligned probe-session frames, none has 5% or more of its area saturated in a connected blob. MMFV therefore contains no genuinely cropped or blown-out fingertips on which a completion could be scored, so the partial-visibility setting must be simulated, and a benchmark of real partial captures is left as future work (Section 6).

E Reference frames

Reading the other frames of the sweep is what separates SweepFill from a restorer. How much of the hidden region those frames must image is answered by the window controls in Appendix D.1, while the subsections below vary their number, their distance from the probe, their quality and their identity, and read their mutual disagreement as a confidence map.

E.1 Number and temporal distance of references

Table 11 varies the references given to the trained SweepFill without retraining, giving it none, one or both of its cross-pose pair, or both taken from the *same* sweep as the probe at $\pm g$ frames, g from 4 to 96 (the sweep changes pose continuously, so g is a pose gap). The second reference is worth less than the first (0.30 against 2.49 dB at w=0.3, for 2.80 dB from the pair), and the closer a same-sweep reference is to the probe the more it helps. At ±4 frames the pair beats the cross-pose default by 0.78 dB, while at ±96 frames the two are level (24.53 against 24.30 dB). Near frames show the hidden region from almost the probe's own viewpoint, so the limit on the copy is view change, not the amount of reference material. Table 12 retrains at equal budget with $R \in \{0, 1, 2, 4\}$ references. Hidden-region quality improves monotonically with R (21.01, 22.84, 23.26, 23.49 dB at w=0.3, with orientation error 14.84° to 13.03°), while the EER of these eight-epoch models saturates after the first reference (17.7, 14.8, 15.9, 15.6%, within the protocol's noise beyond R=1) and rank-1 keeps rising (Table 12). We kept R=2 for SweepFill, where the gain per reference flattens, since R=4 costs twice the reference encoding.

E.2 Reference composition with degraded, same-pose and wrong references

Two further controls test whether the result depends on a convenient choice of reference rather than on the sweep itself. Degrading the references with the same corruption model as the target costs almost nothing (25.28 against 25.34 dB hidden PSNR at a matched 6-epoch budget, Table 13), so the model does not need pristine auxiliary views. References drawn from the *same* pose sweep, still at least 8 frames away, help *more* (25.92 dB), for the same reason the near same-sweep frames of Table 11 do, since correspondence is easier when the viewpoint barely moves. The cross-pose setting used throughout the paper is therefore the harder of the two. Table 14 instead replaces the references with a different subject's sweep. Ridge contrast survives even though the hidden-region gain does not, so a wrong reference degrades gracefully rather than importing another print.

Table 13: **Reference composition**, matched budget (6 epochs, one GPU, 66k crops, which is smaller than the 8-epoch ladder of Table 29, Appendix K.1). Degrading the references costs almost nothing, while same-pose references help more than cross-pose ones, so the configuration used throughout the paper is the harder of the two.

reference composition	PSNR ↑	hid. PSNR ↑	hid. OF° ↓	hid. band	hid. MAPE
clean, cross-pose *(control)*	31.93	25.34	5.30	0.792	0.113
degraded, cross-pose	31.89	25.28	5.43	0.787	0.115
clean, same-pose ($\geq$8 frames away)	**32.20**	**25.92**	**5.00**	**0.805**	**0.108**

Table 14: **Reference controls at test time**, one trained SweepFill model, validation sweep, hidden region, mixed degradation. Only the references change, since weights, probes and windows are identical across the first three rows. The last row is the DiT-UNet restorer for scale.

	w=0.6			w=0.45			w=0.3		
references given to SweepFill	PSNR ↑	OF° ↓	band	PSNR ↑	OF° ↓	band	PSNR ↑	OF° ↓	band
own sweep (default)	**27.57**	**21.23**	**0.620**	**25.82**	**13.25**	**0.700**	**24.31**	**11.01**	**0.671**
switched off at test time	26.45	24.90	0.557	23.70	15.56	0.595	21.49	13.31	0.545
a *different subject*'s sweep	26.22	23.24	0.613	23.78	15.98	0.682	21.75	14.04	0.666
DiT-UNet restorer, full training	26.37	37.23	0.155	23.83	23.51	0.215	21.39	20.07	0.155

E.3 Wrong-reference identity control

Degrading gracefully at the image level does not show whether the completion picks up the *stranger's* identity. Every held-out probe is completed a second time with the reference pool taken from a different subject's probe video (the donor), deterministically assigned, and scored against the same enrollment gallery under the same protocol. Three quantities matter. For *utility*, rank-1 against the probe's own identity falls from 79.2% to 42.7% at w=0.3 and from 83.3% to 77.1% at w=0.5, so references matter most where the print shows least. For *leakage*, 0 of 96 wrong-reference completions are accepted as their donor at the 1%-FAR threshold, the highest scoring 0.650 against a threshold of 0.757. For *drift*, scoring a wrong-reference completion against the donor and comparing it with the *same probe's* correct-reference completion scored against the *same* donor, so that only the references differ, gives 0.120 at w=0.3 and 0.026 at w=0.5, both $p < 0.001$ under a paired sign-flip permutation test over probes. The construction rules out any explanation in terms of the donor being a generically attractive gallery entry. The two quantities say different things. The accept rate shows that at this operating point no completion impersonated the owner of the auxiliary frames in 96 trials, bounding the underlying rate below 3.1% (one-sided 95%, and since probes within a finger are correlated the true bound is looser), while the drift shows that the mechanism by which it could exists and grows as the window shrinks.

E.4 Reference disagreement as a confidence map

SweepFill takes two reference frames from the sweep by a fixed rule, but the sweep offers many more. We complete every identity-protocol probe with 6 distinct pairs (the published pair among them, with the degraded probe and window byte-identical across pairs) and take the per-pixel standard

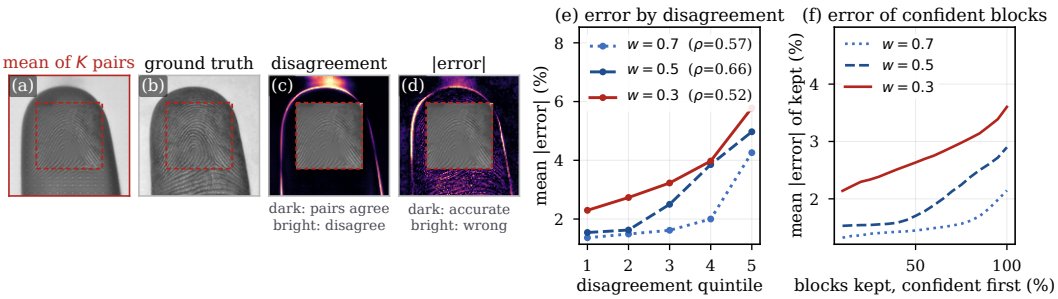

Figure 7: **Reference disagreement as a confidence map**, w=0.3. (a) Mean of 6 completions of one probe, (b) its ground truth, (c) their per-pixel standard deviation and (d) the absolute error of the mean. (c) and (d) color the hidden region only, while inside the dashed window each panel shows the completion, so restored ridges sit beside generated ones. (e) Over all hidden ridged 16-px blocks of all probes, mean absolute error by quintile of block disagreement, with the Spearman rank correlation in the legend. (f) Mean error of the most-confident fraction of those blocks, the part of the generated area a downstream consumer can keep.

deviation across the 6 completions as a *disagreement map*. Where the pairs agree the flow follows the references, whereas where they disagree the model is extrapolating.

The map tracks intensity error. Over 15,840 hidden ridged blocks at w=0.3 the rank correlation between block disagreement and block error against the ground truth is 0.52 (0.66 and 0.57 at w=0.5, 0.7), the least-certain fifth of the blocks carries a mean error of 5.8% of the intensity range against 2.3% for the most certain fifth, and the most confident half of the generated area has error 2.6% against 3.6% overall (Figure 7e,f). It is not a map of orientation error (rank correlation 0.05 at w=0.3), because the pairs disagree on ridge phase and contrast far more than on flow, which the references pin down. Two hoped-for uses do not follow. Averaging the 6 completions is no better than one (Table 15, second row of each block), since the mean matches the single fixed pair on the image metrics, trails it by about a point of EER under the deep matcher, and trails it under the minutiae matcher at w=0.5, so the disagreement is not noise. And gating the generated minutiae on the map does not sort true from spurious. At w=0.3, keeping the 70% most confident blocks removes a quarter of the true minutiae and a twentieth of the spurious ones (15.1% recovered and 52.1% spurious, against 20.2% and 54.4% ungated), which is worse than not gating at all. The map grades generated intensity, not minutiae fidelity.

Test-time reference selection. Closer-pose references help more, so the pair might be worth choosing per probe. Instead of the fixed pair we pick, out of every frame of the other sweeps, the two whose ridge-gradient pattern inside the visible window correlates best with the probe's. The third row of each block of Table 15 scores that choice on identical probes, gallery and protocol, and the selected pair is no better than the fixed one (Ridgeformer EER 13.5% against 11.6% at w=0.3, equal at w=0.7). Window similarity is measured where the probe is visible, while the completion needs agreement where it is not, and the fixed rule, which takes frames from the middle of the other sweeps, already gives views of the hidden region.

F Content of the hidden region

The band and orientation numbers of Section 4 are summaries. Three views open them up, namely the hidden-region spectra those summaries are computed from (Figure 8), a per-block map of where the generated flow is wrong (Figure 10), and the per-probe change in genuine matcher score (Figure 9, the probe-level view of Table 34, Appendix K.3). The subsections then separate the ridge signal SweepFill recovers from the part it does not (Appendix F.2), test the fixed ridge band against the spread of real ridge periods (Appendix F.3), and document a periodic decoder artifact (Appendix F.4).

Table 15: **Three ways of choosing the reference pair**, identity-protocol probes, with the degraded probes and windows held identical throughout, so only the references differ. The fixed pair is the one SweepFill uses everywhere else, the mean averages 6 completions made with 6 different pairs, and the selected pair scores every fourth frame of the other sweeps by the normalized correlation of Sobel gradient magnitude inside the visible window and takes the two best, at least eight frames apart. Neither alternative beats the fixed rule. Hidden-region orientation and band are undefined at w=0.7, where no block is hidden with the 8-px margin the metric needs.

	hidden region			Ridgeformer		LEADER+Bozorth3	
how the references are chosen	PSNR ↑	OF° ↓	band	EER% ↓	R-1% ↑	EER% ↓	R-1% ↑
w=0.7							
fixed pair (published, SweepFill)	30.10	–	–	7.3	88.5	19.0	77.1
mean of 6 pairs	30.11	–	–	7.3	87.5	18.4	80.2
pair selected by window similarity	30.13	–	–	7.3	87.5	20.2	72.9
w=0.5							
fixed pair (published, SweepFill)	26.94	23.09	0.640	9.4	83.3	22.8	67.7
mean of 6 pairs	27.07	22.91	0.632	10.3	84.4	27.3	65.6
pair selected by window similarity	26.89	23.21	0.638	11.4	84.4	24.5	67.7
w=0.3							
fixed pair (published, SweepFill)	25.00	10.90	0.715	11.6	79.2	23.2	64.6
mean of 6 pairs	25.16	10.69	0.691	12.6	80.2	22.3	66.7
pair selected by window similarity	24.61	11.61	0.703	13.5	75.0	22.5	64.6

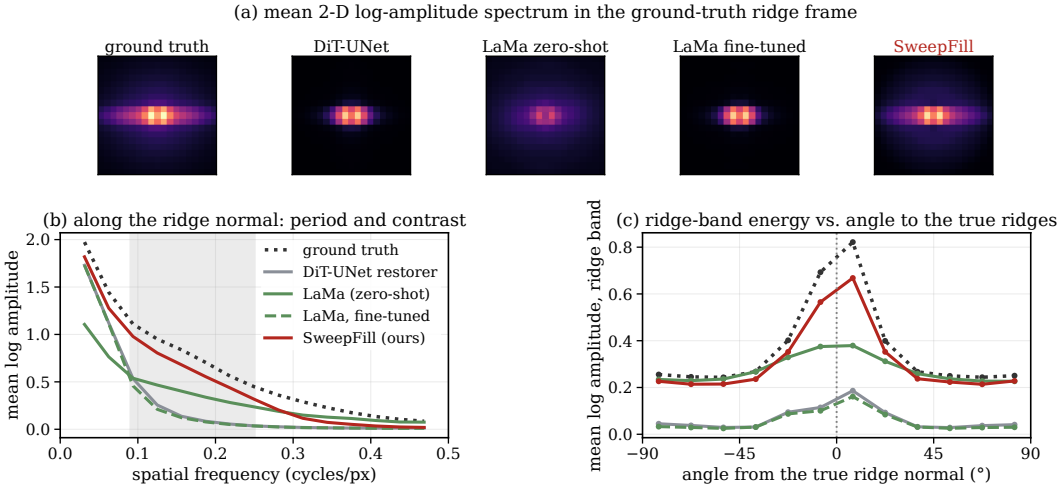

Figure 8: **Hidden-region spectra in the ground-truth ridge frame**, identity probes at w=0.3. Every 32-px block at least 90% hidden and ridged in the ground truth is rotated so the true ridge normal points along $+x$, Hann-windowed, and its log-amplitude spectrum averaged per method. (a) Averaged 2-D spectra, where correctly oriented ridges concentrate energy on the horizontal axis at the ridge frequency. (b) The profile along that axis, i.e. ridge period and contrast, and the gray band marks periods of 4 to 11 px. SweepFill reaches 0.63 mean log amplitude in the band, against 0.77 for the ground truth, 0.18 for the DiT-UNet restorer and 0.16 for fine-tuned LaMa. (c) Ridge-band energy against angle from the true normal. SweepFill peaks where the ground truth peaks and nearly as sharply, its energy reaching 2.1× its angular mean against 2.3× for the ground truth, while zero-shot LaMa is almost flat over angle at 1.4×, the signature of texture drawn at every orientation.

F.1 Restored versus hidden region in the matcher gain

A matcher gain could come from the window SweepFill restores or from the region it generates. Table 16 separates them with two composites built byte for byte from the published probes, namely the restored window with the hidden region blanked to the input constant, and the degraded window exactly as the matcher already had it with the hidden region added. The two matchers answer

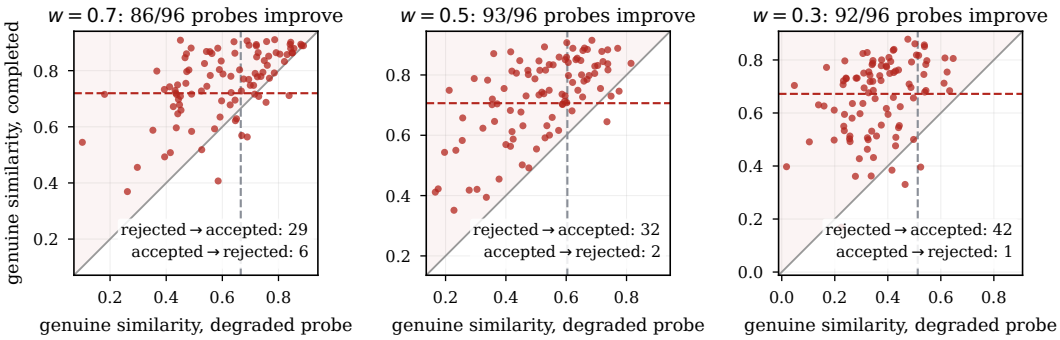

Figure 9: **What completion does to each probe.** Genuine similarity under the deep matcher for every MMFV probe, degraded (horizontal) against completed by SweepFill (vertical). Dashed lines are each role's own accept threshold at 1% FAR and the shaded half-plane marks improvement. At w=0.3, 92 of 96 probes score higher after completion, 42 cross from rejected to accepted and 1 crosses the other way, while at w=0.7 the counts are 86, 29 and 6.

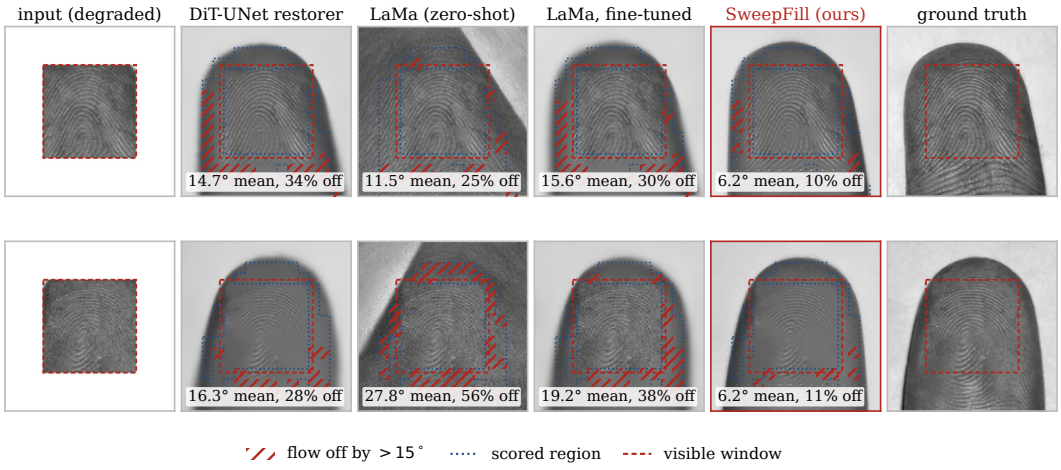

Figure 10: **Where the ridge flow goes wrong.** Two held-out probes at w=0.3, chosen by rule as those at the 40th and 60th percentile of SweepFill's per-probe error, so neither is a best case. The scored region (dotted) is the blocks the image metrics score (hidden with an 8-px margin and ridged in the ground truth), silhouette blocks excluded. A block is hatched red when its measured ridge orientation is more than 15° from the ground truth, against 45° for chance and a 5° plateau for the best extractors (Schuch et al., 2017). Each panel states the mean error over the scored region and the share of blocks flagged, which is about a fifth for SweepFill and more than half for every baseline. Flat skin and texture at the wrong angle are both flagged. The images are unmodified, so each method's smoothness is visible as it is.

differently. The deep matcher, which reads ridge flow and texture, gains mostly from the hidden region once more than half the print is hidden (at w=0.3, 27.1% EER from restoration alone, 16.9% from generation alone, 11.6% from both) and about equally from either when little is hidden (10.2% against 11.4% at w=0.7). The minutiae matcher gains from the restored window at every visibility (26.0% against 33.1% at w=0.3), which is what Section 5.4 predicts, because the minutiae of the hidden region are mostly spurious, so a minutiae matcher cannot use it, while its ridge flow is nearly correct, so a texture matcher can. Restoration and generation are complementary, and the identity gain of Table 1 is not attributable to restoring the visible window alone.

F.2 Flow versus phase in the hidden region

Section 5.4 bounds the completions with a minutiae count. We measure the same thing at the level of the ridge signal, separating the part SweepFill gets right from the part it does not, on the

Table 16: **Region-composition control**, identity protocol, same probes, gallery and fixed thresholds as Table 1. Each block scores what a matcher sees, namely the raw probe, the restored window with the hidden region blanked, the raw window with the hidden region added, and the full completion.

	Ridgeformer		LEADER+Bozorth3	
what the matcher sees	EER% ↓	R-1% ↑	EER% ↓	R-1% ↑
w=0.7				
degraded probe (input)	15.8	59.4	31.2	52.1
restored visible window, hidden region blank	10.2	81.2	20.7	75.0
degraded visible window, generated hidden region	11.4	78.1	29.1	56.2
full completion (SweepFill)	**7.3**	**88.5**	**19.0**	**77.1**
w=0.5				
degraded probe (input)	20.8	53.1	28.7	54.2
restored visible window, hidden region blank	15.6	63.5	20.7	69.8
degraded visible window, generated hidden region	13.5	72.9	28.7	56.2
full completion (SweepFill)	**9.4**	**83.3**	**22.8**	**67.7**
w=0.3				
degraded probe (input)	26.0	21.9	38.1	40.6
restored visible window, hidden region blank	27.1	27.1	26.0	62.5
degraded visible window, generated hidden region	16.9	65.6	33.1	52.1
full completion (SweepFill)	**11.6**	**79.2**	**23.2**	**64.6**

identity-protocol probes (96 probes and 867 scored blocks at w=0.3). A 16×16 block counts only where the answer is defined, that is, with no visible pixel in it, inside the finger silhouette eroded by one block, and carrying a well-defined ridge in the clean frame (orientation coherence ≥ 0.5, spectral peak-to-mean ≥ 2, period in $[4, 12]$ px).

Flow. Root-mean-square block orientation error against the clean frame, and the absolute error of the local ridge period.

Phase. Each scored block is demodulated with a complex Gabor filter tuned to the clean frame's local orientation and period, and phase agreement is $\left|\left\langle e^{i(\phi-\phi_{\mathrm{gt}})}\right\rangle\right|$ over blocks, which is 1 when the generated ridges sit at one fixed offset from the true ones in every block (zero offset included) and 0 when the offset is unrelated from block to block. It is the quantity dense phase-demodulation registration is built on, and what "plausible, not reconstructed" costs. In Table 17, SweepFill is the only method whose generated flow approaches the clean frame, 10.5° RMSE at w=0.3 against 46.3° for the same network without our components, with the ridge period within 0.78 px. *No* method recovers phase, since none exceeds 0.052, SweepFill included. The last row is the control, which applies the identical measurement to the *visible* part of the same probes, where SweepFill reaches 0.55, so the metric is not insensitive, and phase is recovered where there is evidence for it. The phase column must not be ranked across methods, since a completion with no ridge energy has an arbitrary phase on a near-zero response, and cannot be scored where the model is not drawing ridges at all. Singular-point localization is not reported because, with a centered window, the core is almost always inside it, and only 3 of the 96 probes have a clean-frame core in the hidden region. w=0.7 is absent because the hidden ring falls inside the eroded finger margin, so no block qualifies.

F.3 Ridge frequency and the ridge band

The spectral loss and the band metric use a fixed radial band (4 to 10.7 px). Table 18 splits the identity probes into terciles of the ground-truth ridge period measured over the hidden ridged blocks, and Table 19 magnifies the fingertip crops (probe and references alike, the central $1/z$ of the crop resized to full size) before degradation and windowing, so every ridge period grows by z, from 8.0 px to 9.2, 10.4 and 12.0 px while the band stays fixed, the last two magnifications carrying the ridges to and beyond its long edge. Across the natural spread of periods the orientation error and band contrast of SweepFill vary little, and the margin over the DiT-UNet restorer and fine-tuned LaMa holds in every tercile. Magnification moves the whole distribution and costs nothing while the ridges stay inside the band, since the orientation error at w=0.3 is 11.0° unmagnified and 9.3° at 1.15× (period 9.2 px). It rises as they reach the long edge of the band, to 12.1° at 1.3× and 17.9° at 1.5× (period 12.0 px,

Table 17: **Flow versus phase in the hidden region**, identity-protocol probes, hidden blocks with a well-defined clean-frame ridge. OF RMSE is the block orientation error against the clean frame, period the absolute ridge-period error, phase the Gabor phase agreement (1 = one consistent offset from the true ridges in every block, 0 = offsets unrelated across blocks). SweepFill recovers flow and period, but no method recovers phase. The control row repeats the phase measurement where the print is visible.

	w=0.5			w=0.3		
completion of the hidden region	OF RMSE°	period px	phase	OF RMSE°	period px	phase
Gabor extrapolation (Hong et al., 1998)	26.1	1.18	0.070	30.4	1.17	0.036
LaMa, zero-shot (Suvorov et al., 2022)	26.0	1.32	0.069	30.7	1.48	0.023
LaMa, fine-tuned	34.4	1.22	0.031	43.6	1.66	0.009
DiT-UNet restorer	27.4	1.02	0.030	46.3	1.76	0.052
SweepFill	9.5	0.80	0.005	10.5	0.78	0.011
control: SweepFill where the print is *visible*			0.575			0.548

beyond the long end of the band and outside the training distribution), while the DiT-UNet restorer stays worse at every scale (Table 19). A deployment at a different ridge scale should rescale to 500 ppi first, as the matcher pipeline already does, rather than rely on the band's width. We do not sweep the band bounds, the patch size or the loss weights. The band is the range of plausible ridge periods at the crop's scale (Section 3.5), not a tuned knob, since Table 18 shows every tercile of the data inside it and Table 19 stresses its edges. The 32-px patch is the smallest window that resolves the band's longest period with three cycles. The loss weights were fixed once (Appendix A) and held for every run, so the ablation ladder isolates each component at one weighting. Tuning them per experiment on the held-out subjects the paper reports would turn a robustness claim into a fit.

Table 18: **Hidden-region quality by ground-truth ridge period**, identity probes split into terciles per window. LaMa (Suvorov et al., 2022) is the fine-tuned variant of Table 30, Appendix K.1.

	mean period		SweepFill		DiT-UNet restorer		LaMa, fine-tuned	
ground-truth ridge period	(px)	probes	OF° ↓	band	OF° ↓	band	OF° ↓	band
w=0.5								
short period	9.4	32	21.9	0.632	38.5	0.110	38.7	0.093
medium period	9.6	32	23.1	0.638	39.4	0.133	39.5	0.105
long period	9.9	32	24.3	0.648	40.2	0.140	42.9	0.106
w=0.3								
short period	8.6	32	9.3	0.663	19.9	0.137	18.9	0.099
medium period	8.9	32	11.3	0.722	21.1	0.163	21.3	0.121
long period	9.2	32	12.1	0.761	21.0	0.182	21.5	0.132

F.4 Periodic lattice artifact in the hidden region

SweepFill's output carries a faint periodic artifact, visible in Figure 1 and in Figure 13, Appendix K.6, which consists of isolated bright pixels on a 16-px lattice, about 4 to 6 gray levels above their neighbors, confined to the hidden region and scaling with its area (about 1.3 levels at w=0.7, 5 at w=0.3). Its source is the spectral loss, not the decoder, since at w=0.3 the lattice stands 0.2 gray levels above its surroundings with the structure head alone, 0.2 with the FFC decoder added and 5.1 with the spectral loss on a plain decoder instead (6.3 for SweepFill). The loss compares Hann-windowed patches on a fixed 16-px grid, and a bright pixel at each patch center, where the window peaks, adds broadband amplitude at almost no pixel cost. Shifting the grid by a random offset at every step removes that incentive in expectation, and at equal budget it lowers the lattice from 5.7 to 2.1 gray levels, at a cost of 14.0° against 13.2° hidden-region orientation error, so we report the trade rather than retrain with it. We measured its effect rather than removing it from the images we report. Replacing only the lattice pixels and their four neighbors by a 3×3 median on the w=0.3 identity probes changes hidden-region PSNR from 25.00 to 25.05 dB, orientation error from 10.90° to 10.84° and band contrast from 0.715 to 0.643, so roughly 0.07 of the band values reported for SweepFill at w=0.3 is broadband energy from the lattice, well inside every band gap we claim. The deep matcher moves from 11.6% EER and

Table 19: **Magnified crops**, validation sweep. The ridge period grows with the magnification (the period column is the verified scale factor applied to the $1\times$ estimate, since the 32-px block estimator cannot resolve periods beyond its band) while the band stays fixed. Band contrast is left blank where the ridges lie outside the band it measures. The upper block is SweepFill and the lower block the DiT-UNet restorer.

	w=0.5				w=0.3			
magnification	period (px)	PSNR ↑	OF° ↓	band	period (px)	PSNR ↑	OF° ↓	band
SweepFill								
×1.00	8.0	26.30	21.75	0.619	8.0	24.30	10.98	0.671
×1.15	9.2	24.32	11.43	0.635	9.2	22.49	9.29	0.630
×1.30	10.4	19.03	12.73	0.747	10.4	16.52	12.10	0.698
×1.50	12.0	15.50	18.83	–	12.0	13.24	17.86	–
DiT-UNet restorer								
×1.00	8.0	24.42	38.38	0.123	8.0	21.35	20.09	0.155
×1.15	9.2	21.56	20.50	0.186	9.2	18.49	17.51	0.138
×1.30	10.4	16.04	21.56	0.334	10.4	13.85	20.95	0.202
×1.50	12.0	13.58	26.97	–	12.0	12.44	26.09	–

Table 20: **Identity protocol by degradation type**, severity 0.5, same probes and gallery for every column. Ridgeformer EER / rank-1 for the raw partial probe, the two equal-budget rungs without references, SweepFill without and with its references, Bozorth3 EER for the raw probe and SweepFill, and the hidden-region orientation error and band contrast from the reference-free rung to SweepFill.

	Ridgeformer EER% / rank-1%					LEADER+Bozorth3 EER%		hidden, no ref. → SweepFill	
degradation	raw partial	struct. only	+ FFC + spec.	SweepFill, no ref.	SweepFill	raw	SweepFill	OF° ↓	band
w=0.5									
occlusion only	11.5 / 70.8	10.4 / 78.1	11.5 / 83.3	10.4 / 87.5	9.4 / 87.5	12.1	17.7	34.3→22.3	0.562→0.657
noise	14.6 / 51.0	10.4 / 81.2	11.6 / 85.4	10.4 / 86.5	7.5 / 88.5	17.1	21.5	34.5→22.5	0.571→0.657
blur (Gaussian / motion)	14.6 / 67.7	12.5 / 67.7	15.6 / 79.2	12.5 / 83.3	11.5 / 86.5	26.4	19.8	34.6→22.4	0.549→0.629
downsampling	14.5 / 57.3	10.4 / 74.0	11.4 / 81.2	10.2 / 88.5	9.6 / 89.6	26.1	17.2	34.7→22.7	0.557→0.638
JPEG	12.3 / 68.8	10.6 / 75.0	11.5 / 83.3	10.6 / 87.5	9.3 / 89.6	12.0	17.5	34.5→22.3	0.559→0.657
w=0.3									
occlusion only	20.8 / 39.6	14.7 / 57.3	14.6 / 62.5	13.5 / 62.5	10.6 / 85.4	22.0	18.1	15.7→10.6	0.601→0.748
noise	22.9 / 36.5	15.6 / 55.2	15.7 / 51.0	14.8 / 61.5	10.5 / 86.5	23.6	20.5	16.0→10.8	0.600→0.740
blur (Gaussian / motion)	22.9 / 32.3	17.6 / 50.0	17.7 / 57.3	15.9 / 65.6	11.5 / 83.3	31.9	16.9	16.0→10.7	0.582→0.727
downsampling	24.0 / 29.2	15.6 / 51.0	15.6 / 53.1	14.7 / 66.7	10.6 / 84.4	27.1	20.2	16.1→10.7	0.581→0.728
JPEG	21.9 / 37.5	15.6 / 55.2	14.6 / 60.4	13.5 / 67.7	10.4 / 84.4	21.3	18.1	15.7→10.6	0.597→0.746

79.2% rank-1 to 12.5% and 76.0%, one to three probes in 96 and well inside the paired bootstrap intervals of Table 34, Appendix K.3.

G Results by degradation type

Table 20 repeats the identity protocol with a single degradation operator at severity 0.5 inside the window (occlusion only, noise, blur, downsampling, JPEG), on probes paired across all models, adding the equal-budget rungs without references (structure head only, and structure head with the Fourier decoder and spectral loss) and SweepFill with its references switched off. The pattern is the same under every operator. In the per-operator image metrics of these models, not tabulated here, the Fourier decoder and spectral loss lower the hidden-region orientation error from about 22° to 16° and lift the band contrast from about 0.14 to 0.60 under every operator, as in Table 29, Appendix K.1, but move the deep matcher's EER by at most a point at w=0.3 and *raise* it by one to three points at w=0.5, most under blur, while raising rank-1. The identity gain at w=0.3 comes from the structure head with the mask-first order (20.8 to 14.7% EER with occlusion alone) and from the references (13.5 to 10.6%). The references contribute most under noise, blur and downsampling, where the visible half is itself uncertain, and least with occlusion alone or JPEG. Under the minutiae matcher the largest gain is for blur (31.9 to 16.9% EER), the operator that destroys the most minutiae inside the window.

H Third-party tools and versions

- **Deep Matcher.** Ridgeformer (Pandey et al., 2025), public `phase1_scratch` checkpoint, loaded with zero missing or unexpected keys. We run the `timm` 0.5.0 wheel the repository ships, but that wheel's vision-transformer pools to the class token, while the repository's own model and fusion code require the full post-LayerNorm sequence of 197 tokens, since it averages the backbone output over the token axis and slices 197 rows from it, neither of which a pooled vector admits. The authors evidently ran a further local edit that was not released, and we restore exactly that behavior by returning the sequence unpooled. Nothing else changes, because for `vit_large_patch16_224_in21k` the pre-logits layer is the identity, so our forward differs from the shipped one only in not pooling, and we load the backbone without its ImageNet weights since the fine-tuned checkpoint supplies every one of them. Our absolute numbers depend on this reconstruction, so the matcher is used only for *relative* comparisons in which every method is scored identically. Input rotation was calibrated on clean gallery/probe pairs before use. The repository states that the checkpoint was trained on RidgeBase, the HKPolyU contactless-to-contact database, ISPFDv1 and ISPFDv2, so the deep matcher never saw MMFV, on which the main results are measured. ISPFDv2 it did see, and we cannot establish identity-level overlap with our evaluation identities, so the ISPFDv2 deep-matcher numbers are read as relative comparisons on identical probes, with the minutiae matcher, which has no training set, as the independent check.
- **Minutiae Matcher.** NBIS (NIST, 2010) Bozorth3, built from the NIST release, matching minutiae from the LEADER extractor listed below, and Mindtct is not used anywhere in the pipeline.
- **Enhancement, orientation, frequency, minutiae extraction.** The `pyfing` toolkit, version 0.6 (Cappelli, 2026b), namely its LEADER extractor (whose paper the toolkit lists as under review at the time of writing) and its published orientation, frequency and enhancement methods (Cappelli, 2024; 2026a), used for the matcher preprocessing chain, the classical baseline, and as the second opinion on orientation-error magnitudes.
- **Inpainting Baseline.** LaMa (Suvorov et al., 2022), traced `big-lama` weights, and two independently distributed traces produce bit-identical output. Zero-shot, visible pixels are copied back verbatim after inpainting, so the baseline is never penalized for altering the observed region. The *fine-tuned* LaMa row trains the same 51.0M-parameter generator, from those weights, on the equal-budget ablation distribution, that is, the 66k-crop subsample for eight epochs, the same window sampler, degradation model, difficulty ramp and pixel objective (Charbonnier, squared error, gradient, MS-SSIM, hidden weight 0.3), without LaMa's high-receptive-field perceptual and adversarial losses, so the row tests the LaMa architecture under our objective. A second fine-tune, *LaMa with its own losses*, uses the published big-lama objective instead (masked ℓ_1 on the known region, the ADE20K ResNet50-dilated perceptual loss, a PatchGAN adversarial loss with R_1 penalty and feature matching, at the official weights and implementations) on the same data and budget. Neither has a structure head, spectral loss or references, and both use AdamW at 10^{-4} with warmup and cosine decay, EMA 0.999, and batch-norm statistics frozen at their pretrained values. Because it is trained on the whole image it also restores the window, so its full output is scored without copying visible pixels back.

I Alternative designs that do not enter SweepFill

This appendix documents the alternatives evaluated around the design of Section 3, namely an adversarial control separating ridge realism from ridge correctness, three literature remedies for the minutiae gap of Section 5.4, and our reason for not reporting a diffusion baseline. None enters SweepFill, and together they localize what remains to local, non-rigid correspondence with the reference frames at ridge resolution.

Adversarial Fine-tuning buys Contrast, not Correctness. We fine-tuned the trained DiT-UNet restorer with a spectral-norm 70×70 PatchGAN for 6 epochs, with adversarial weight 0.02 and the adversarial loss applied only over the hidden region, and scored it with the harness of Table 27, Appendix K.1 (Table 21), with the result the perception-distortion account predicts. Band contrast at w=0.3 rises from 0.142 to 0.392, almost threefold, because the discriminator refuses flat skin and so the decoder draws ridges. Their flow does not follow, since orientation error stays near the restorer's at every window of Table 21, and about twice SweepFill's, improving at w=0.6 and 0.45

Table 21: **Adversarial control**, hidden region on the validation sweep (512 probes per condition, mixed degradation). The PatchGAN fine-tune and the DiT-UNet restorer it started from differ only by the discriminator, and SweepFill is shown for scale.

	w=0.6			w=0.45			w=0.3		
method	PSNR ↑	OF° ↓	band	PSNR ↑	OF° ↓	band	PSNR ↑	OF° ↓	band
DiT-UNet restorer (initialization of the fine-tune)	26.26	37.36	0.146	23.71	23.68	0.198	21.27	20.04	0.142
+ PatchGAN fine-tune (Isola et al., 2017)	26.31	34.25	0.345	23.76	22.99	0.450	21.24	20.52	0.392
SweepFill (ours)	**27.57**	**21.23**	**0.620**	**25.82**	**13.25**	**0.700**	**24.31**	**11.01**	**0.671**

Table 22: **Multiple-hypothesis training** (K=3, relaxed winner-takes-all, equal budget). "Committed" is the hypothesis available at test time, "oracle" the best of three by hidden-region PSNR, "selector" a training-free choice by agreement with the model's own predicted ridge flow. Win rate is the share of samples each hypothesis wins.

	hidden PSNR			hidden OF°			win rate
window	committed	oracle	selector	committed	oracle	selector	(%)
w=0.6	25.85	26.57	25.57	28.92	27.98	27.96	43/21/36
w=0.45	23.39	24.28	22.95	17.85	17.30	17.48	45/20/35
w=0.3	21.51	22.04	20.35	14.92	14.63	14.72	58/14/28

and *worsening* to $20.52°$ from $20.04°$ at the hardest setting, where hidden-region PSNR is flat too (21.24 against 21.27 dB). The matchers split (Table 32, Appendix K.3). Under the deep matcher the fine-tune tracks the restorer it started from, 9.2% against 10.4% EER at w=0.7 and 21.9% against 19.6% at w=0.3, far behind SweepFill (11.6%, with rank-1 52.1% against 79.2%). The minutiae matcher reverses the order, reaching 17.9, 20.2, 19.9% EER at w=0.7/0.5/0.3 against 19.0, 22.8, 23.2% for SweepFill, and 70.8% rank-1 against 64.6% at w=0.3. Sharper ridges give the minutiae extractor more to work with even when their flow is no more correct. An adversarial loss supplies a prior over what fingerprints look like, not evidence about *which* fingerprint this is, whereas SweepFill reaches 0.67 band at $11.01°$ on the same sweep, better on both axes, because its structure head and sweep references add evidence. Stacking a discriminator on SweepFill would trade the flow the deep matcher rewards for the contrast the minutiae extractor rewards without making generated minutiae more true (Section 5.4), the wrong direction for a biometric.

Multiple Hypotheses do not help once you must commit. For a multi-modal hidden region, the natural response is K hypotheses trained with relaxed winner-takes-all (Rupprecht et al., 2017; Perera et al., 2024), committing to one ridge phase instead of averaging. We added K=3 hypothesis codes modulating the decoder (ϵ=0.05, hidden region only). Diversity is real, since win rates run from 43/21/36% at w=0.6 to 58/14/28% at w=0.3 with no dead hypothesis, and the oracle is 0.5 to 0.9 dB above the committed one (Table 22). But the committed hypothesis, all that is available at test time, matches the single-hypothesis model at three times the decoder cost, and a training-free selector by agreement with the model's own structure head recovers orientation while losing up to 1.2 dB. The oracle gap quantifies an ambiguity this mechanism cannot cash in, and a learned scorer (Ilg et al., 2018) remains open.

A Coarse Alignment prior gains little on the pairs the model uses. Reference-guided inpainting on uncanonicalized views reports its largest gain from explicit alignment (+1.31 dB from patch alignment in TransRef (Liu et al., 2025), while MVInpainter (Cao et al., 2024b) gains +0.3 dB from grouped flow and loses with dense flow). We implemented the standard cheap version, rotation and scale from phase correlation of log-polar magnitude spectra then translation, gated by the correlation peak, and measured it on the 96 identity-protocol probe frames by ridge-band correlation, the Pearson correlation of the high-passed images the aligner works on. On synthetic transforms it is accurate, since a $12°$ rotation is recovered to within $0.12°$ at the median and $0.34°$ at worst, taking the correlation from 0.08 to 0.97. On real pairs the gain depends on pose. A same-pose reference from the same sweep, at least eight frames away, gains +0.28 at the median, whereas a cross-pose reference, the kind SweepFill consumes, gains only +0.04, and confidence gating (0.06, keeping 60% of pairs) raises this to +0.25 by keeping mostly same-pose pairs. Under plain-image correlation, dominated by silhouette and shading, the same warps read as *losses* (-0.01 same-pose, -0.22 cross-pose), which

Table 23: **A patch-correlation warp at the 64×64 decoder stage**, identity protocol, against SweepFill, a control fine-tuned with the same recipe and no module, and the same fine-tune with the coarse affine alignment prior applied to every reference. Hidden-region metrics come from the validation sweep at w=0.3, and recovered / spurious hidden-region minutiae and both matchers' EER from the identity protocol.

	hidden, w=0.3			w=0.7		w=0.5		w=0.3		EER%, w=0.3	
model	PSNR ↑	OF° ↓	band	rec.%	spur.%	rec.%	spur.%	rec.%	spur.%	Ridgef.	Boz.
SweepFill	24.31	11.01	0.671	24.6	44.8	23.2	48.7	21.3	54.2	11.6	23.2
SweepFill, fine-tuned two more epochs (control)	24.30	11.02	0.669	25.1	44.6	26.1	44.5	21.0	51.4	13.3	19.7
+ local correspondence at 64×64, same fine-tune	24.27	10.96	0.670	25.9	46.8	26.5	49.0	20.2	52.5	12.5	21.9
+ affine alignment of the references, same fine-tune	24.31	10.99	0.671	23.0	44.7	25.2	48.0	21.4	52.6	13.5	21.3

misleadingly suggests harm. Our preprocessing already canonicalizes rotation and scale per frame, so the affine part of a cross-view mismatch is small and the remainder is non-rigid (perspective, skin deformation), outside a global affine's model class. Trained in by the same two-epoch fine-tune from SweepFill as the controls below, every reference warped onto the degraded probe at train and test time (Table 23), the prior moves hidden-region orientation error from $11.02°$ to $10.99°$, recovered minutiae from 21.0% to 21.4% and the deep matcher's EER from 13.3% to 13.5% against the control sharing its recipe, every difference inside the protocol's noise.

Finer access to the references does not recover minutiae either. The remaining option is a local, non-rigid correspondence to the references at ridge resolution, which we test at two points in the network. The first is an explicit warp at the 64×64 decoder stage (4-px cells), built from a patch-wise correlation between decoder features and each reference's encoder features over a ±4-cell search window, a soft-argmax displacement per cell, and the warped reference features injected with that displacement and its confidence through a zero-initialized convolution, so the model equals SweepFill at initialization. We fine-tuned it from SweepFill for two epochs on all crops and, as a control, fine-tuned the unchanged SweepFill with the same recipe (Table 23). Hidden-region quality is unchanged ($10.96°$ orientation error against $11.02°$ at w=0.3), as is minutiae recovery, with 20.2% recovered and 52.5% spurious at w=0.3 against 21.0% and 51.4% for the control and 21.3% and 54.2% for SweepFill, all within the protocol's noise, and both matchers agree.

The second place is attention rather than warping, since if 16-px bottleneck tokens are too coarse to transfer 6-px ridge detail, a finer decoder stage can attend to the references. We added global cross-attention from the 32×32 decoder stage to reference encoder features at the same resolution, zero-initialized so the model is provably identical to SweepFill at initialization (verified by a maximum output difference of 0), then fine-tuned from the full SweepFill model on all data. The result is flat, with hidden PSNR 27.61 against 27.57 dB and band 0.622 against 0.620 at w=0.6 on the common sweep, and minutiae recovery unchanged, 24.5/23.8/21.4% at $w=0.7/0.5/0.3$ against 24.6/23.2/21.3% for SweepFill (Table 2). Both fail in the same way. A ±16-px learned warp and 8-px decoder cells attending globally have the reach to find the matching ridge, but no reason to prefer it, because the hidden-region pixel and gradient terms of Equation (4) reward agreement with the target, but not enough to teach correspondence, so both modules settle for a plausible ridge over the right one. What remains untried is not capacity or global attention but training the correspondence against an objective of its own, photometric or feature-space agreement on the visible region, where the warp can be checked. Appendix K.4 bounds what that could buy on these frames. Registered to the clean frame itself, the references recover 30% of hidden minutiae at phase coherence 0.07, so a learned correspondence could at best add a few points.

Why there is no Diffusion Baseline. A diffusion inpainter (Lugmayr et al., 2022; Saharia et al., 2022) fills the hidden region by *sampling* ridge phase from a learned prior, as does the reference-guided inpainter of Section 2 closest to our design, on a pretrained diffusion backbone (Cao et al., 2024a). That is the one design this paper sets out not to use. Ridge phase more than a few periods from the window is unconstrained by the observation (Section 3.5), so a sample from a prior places minutiae the evidence does not support, and the minutiae-fidelity and leakage metrics of Section 5.4 charge for exactly that, as the adversarial control above shows. SweepFill takes the other route of Section 3.6, keeping every added ridge tied to the visible print and the sweep. A diffusion baseline would also be a second model family at a different budget, since no fingerprint diffusion model is public and RePaint needs an unconditional one trained first, so the fine-tuned LaMa of Table 30, Appendix K.1

Table 24: **Cost of the realistic order at the input**, 168 held-out probe frames, mixed severity 0.5, identical corruption draw on both sides, each scored against the same clean target over the finger. PSNR in dB, orientation error in degrees.

	enhance, then degrade		degrade, then enhance		
degradation	PSNR ↑	OF° ↓	PSNR ↑	OF° ↓	ΔPSNR ↑
sensor noise	30.7	3.4	22.0	12.3	-8.7
JPEG	36.7	1.3	32.6	3.8	-4.1
blur	36.9	9.8	35.5	10.0	-1.4
downsampling	33.4	4.2	32.1	4.5	-1.3
mixed (the protocol default)	30.2	9.5	25.8	16.9	-4.4

is the strongest inpainter we could bring to identical data, masks and losses. We therefore treat a domain-tuned diffusion inpainter, scored under the same protocol, as a separate study rather than a missing row.

J Matcher-side preprocessing

Two choices sit between the raw photograph and the matcher, namely where the contrast enhancement of Appendix A sits relative to the corruption, and which enhancement is used at all. Neither is forced by the method, so we vary both.

J.1 Order of enhancement and degradation

Our degradation model enhances the clean frame, then degrades it, whereas a camera does the reverse, enhancing an already degraded image. Since enhancement is a local contrast stretch, it amplifies noise as readily as ridges, so the orders are not interchangeable. We re-run the study in the reverse order.

The realistic order is harder, and selectively so. Table 24 degrades 168 held-out probe frames both ways with an identical corruption draw, each scored against the same clean target. Reversing the order costs 30.2 → 25.8 dB and 9.5° → 16.9° of orientation accuracy under the mixed degradation used throughout. Sensor noise dominates the loss (30.7 → 22.0 dB) because enhancement lifts noise into the ridge band, pushing band contrast to 1.51 against a true value of one, so it over-draws rather than restores. Blur and downsampling barely move (1.4 dB), leaving no high frequencies to amplify. Rescaling each probe to the target's mean and variance still leaves 26.9 dB, which is lost evidence, not a contrast shift.

The result holds with and without retraining. Table 25 scores every method on those harder probes under the identity protocol of Section 5.3. The paper's weights, out of domain here, still take a 30%-visible probe from 31.5% EER to 13.5% and rank-1 from 21.9% to 76.0%, and still lead the DiT-UNet restorer, by more than the margin of Section 5.3. Retraining both models under the realistic order, at the equal budget and with full training, favors SweepFill, since with full training the completion reaches 12.6% EER and 81.2% rank-1, ahead of the original SweepFill, which pay a domain-shift penalty a retrained model does not. The minutiae matcher agrees (25.8% against the restorer's 34.2%).

The order is expensive for restoration and nearly free for generation. The validation sweep shows the split directly. On the harder probes the original SweepFill loses 4.9 dB of full-image PSNR (33.8 → 28.9 dB at w=1), since the visible window it copies and cleans has lost detail. In the hidden region it loses only 0.38 dB (23.93 dB at w=0.3), with orientation error 11.01° to 11.38° and band contrast 0.67 to 0.66. Ridge flow and period outside the window come from the structure head and the reference frames, which do not read the noise level of the visible pixels, so noisier observations burden the copy-and-clean path and spare the generate path. Retraining recovers most of the restoration loss, 2.6 dB over the original SweepFill on the same probes, while generation stays put (24.07 dB, 10.91°). Retraining also supplies a comparison the main text cannot, with SweepFill and the DiT-UNet restorer on the same data, for the same epochs, under the same degradation. SweepFill leads by 10.3 EER points and 40.6 rank-1 points at w=0.3, wider than the margin Section 5.3 reports

Table 25: **Identity protocol under the realistic order.** Every row is scored on the same degrade-then-enhance probes (32 held-out identities, 96 probes, deep matcher), with EER and rank-1 in percent. The top block holds the weights of Section 5, evaluated out of domain, and the bottom both models retrained under this order at equal budget and with full training.

	w=0.7		w=0.5		w=0.3	
probe scored by the deep matcher	EER ↓	R-1 ↑	EER ↓	R-1 ↑	EER ↓	R-1 ↑
degraded probe	21.9	45.8	22.8	33.3	31.5	21.9
SweepFill, original	9.4	83.3	13.5	79.2	13.5	76.0
DiT-UNet restorer, original	14.6	65.6	20.6	55.2	25.0	36.5
SweepFill, retrained (equal budget)	10.2	81.2	16.7	70.8	20.8	54.2
DiT-UNet restorer, retrained (equal budget)	15.6	65.6	18.8	56.2	24.0	34.4
SweepFill, retrained (full training)	8.4	88.5	8.3	85.4	12.6	81.2
DiT-UNet restorer, retrained (full training)	14.8	67.7	19.9	60.4	22.9	40.6

Table 26: **Matcher gains under other preprocessing pipelines**, identity protocol, EER% of the raw partial probe and of SweepFill's completion, both matchers.

	w=0.7				w=0.5				w=0.3			
	Ridgeformer		LEADER+Bozorth3		Ridgeformer		LEADER+Bozorth3		Ridgeformer		Bozorth3	
pipeline before matching (EER%) ↓	raw	SweepFill	raw	SweepFill	raw	SweepFill	raw	SweepFill	raw	SweepFill	raw	SweepFill
CLAHE (published pipeline)	15.8	7.3	31.2	19.0	20.8	9.4	28.7	22.8	26.0	11.6	38.1	23.2
gallery without CLAHE, probes with	15.6	6.2	26.8	20.9	20.7	8.3	28.1	23.3	27.1	11.3	38.7	28.5
+ global histogram equalization	11.5	7.3	31.6	23.1	19.0	8.6	28.8	21.8	26.0	8.3	33.7	27.5
+ unsharp masking	13.5	6.3	26.4	17.4	18.8	9.6	24.4	22.6	22.9	8.5	37.4	19.4
+ oriented Gabor enhancement	17.5	12.5	37.4	27.0	24.1	12.5	33.5	29.2	38.3	14.8	35.9	29.2
+ downsampling to 128 px	13.6	9.6	33.4	28.3	21.9	11.6	32.1	31.7	32.4	16.4	37.7	28.6

against a restorer that saw more training samples. Section 5 keeps the original SweepFill, on which every other study is measured, and the retrained models support the same conclusion.

J.2 Alternative preprocessing pipelines

The identity numbers are measured after the alignment and CLAHE of Appendix A. Table 26 re-runs both matchers with a different pipeline between the images and the matcher, keeping the completions fixed. The alternatives are a gallery enrolled *without* CLAHE against probes with it (a mismatched deployment), and global histogram equalization, unsharp masking, oriented Gabor enhancement or a downsampling to 128 px applied to every image. Absolute error rates move with the pipeline, by up to 12 EER points for the raw probe under Gabor enhancement, but the completion stays ahead under every pipeline and both matchers. The deep matcher's gain at w=0.3 ranges from 14.4 to 23.6 EER points, while the minutiae matcher's gain shrinks most under global histogram equalization and oriented Gabor enhancement, both of which redraw ridge contrast over the whole image and so act on the generated ridges as much as on the observed ones.

K Additional results

This appendix holds the tables and figures the main text refers to, grouped the way the experiments are, namely image-level quality (Appendix K.1), the identity protocol in full (Appendix K.3), the second benchmark (Appendix K.5), and the qualitative panels (Appendix K.6). Apart from the structure-field study of Appendix K.2, every number comes from a run the main text has already introduced, reported over more windows, conditions or systems than the main text has room for.

K.1 Image-level quality

Table 27 gives the per-window numbers behind Figure 3, and Table 28 expands SweepFill across every visibility and degradation condition, the restoration rows at w=1 included. Table 29 is the equal-budget ladder of Section 5.1, with the single-component cells below the rule that separate what each part contributes. Table 30 scores the completion baselines on the identity-protocol probes rather

Table 27: **Hidden-region quality on the validation sweep** (512 probes per condition, mixed degradation). "OF" is the mean orientation error in degrees, "band" the band contrast relative to the ground truth, and "orient." the same contrast counted only within $\pm 20^\circ$ of the true ridge direction, which ridges at the wrong angle or isotropic noise cannot earn. All three are scored on blocks hidden with an 8-px margin (Section 4), which exist only for $w \leq 0.6$. Bold is best per column. The DiT-UNet restorer (Section 3.2) appears at equal budget and fully trained, the latter having seen $10.1\times$ the equal-budget sample visits and $1/0.80 \approx 1.25\times$ those of SweepFill (Appendix A).

	w=0.6				w=0.45				w=0.3			
method	PSNR ↑	OF° ↓	band	orient.	PSNR ↑	OF° ↓	band	orient.	PSNR ↑	OF° ↓	band	orient.
DiT-UNet restorer, equal budget	25.76	37.33	0.154	0.130	23.20	24.84	0.180	0.163	20.84	21.54	0.158	0.120
DiT-UNet restorer, full training (10× the equal budget)	26.37	37.23	0.155	0.148	23.83	23.51	0.215	0.212	21.39	20.07	0.155	0.140
SweepFill (ours), equal budget	26.62	26.97	0.566	0.468	24.99	16.36	0.606	0.531	23.26	13.17	0.595	0.525
SweepFill (ours), full training	**27.57**	**21.23**	**0.620**	**0.537**	**25.82**	**13.25**	**0.700**	**0.645**	**24.31**	**11.01**	**0.671**	**0.620**

Table 28: **Full sweep for SweepFill** on the validation sweep, covering every visibility × degradation condition with 512 probes each. "in PSNR" scores the degraded input against the clean target and is undefined (100) for the identity "clean" operator at full visibility. RMSD is the root-mean-square orientation error, reported beside the mean for comparability with the orientation-field literature. Orientation, frequency and band are scored on blocks hidden with an 8-px margin (Section 4) and are undefined for $w \geq 0.75$, where no such block exists.

		whole crop			hidden region					
w	degradation	in PSNR ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	OF° ↓	RMSD°	MAPE	band
1.0	clean	100.00	41.86	0.9769	–	–	–	–	–	–
1.0	mixed	31.15	33.77	0.8888	–	–	–	–	–	–
0.9	clean	20.13	38.30	0.9700	31.57	0.9473	–	–	–	–
0.9	mixed	19.71	33.21	0.8846	31.58	0.9396	–	–	–	–
0.75	clean	16.29	34.96	0.9399	30.22	0.8519	–	–	–	–
0.75	mixed	16.13	32.00	0.8648	30.22	0.8475	–	–	–	–
0.6	clean	13.54	31.01	0.9016	27.60	0.8123	20.95	32.02	0.131	0.645
0.6	mixed	13.47	29.61	0.8367	27.57	0.8097	21.23	32.32	0.129	0.620
0.45	clean	10.90	28.24	0.8588	25.87	0.7779	13.12	24.31	0.110	0.726
0.45	mixed	10.86	27.48	0.8039	25.82	0.7751	13.25	24.45	0.109	0.700
0.3	clean	8.88	25.86	0.7979	24.41	0.7322	10.76	20.95	0.123	0.703
0.3	mixed	8.87	25.41	0.7543	24.31	0.7285	11.01	21.19	0.123	0.671

than on the validation sweep, so it is comparable with the identity tables of Appendix K.3 and not with Table 27.

K.2 How the structure head acts, and a decoder that follows it

Removing the structure head from the full equal-budget model raises hidden-region orientation error at w=0.3 from 13.2° to 17.4° (Table 29), and its own field is within 5.74° of the ground truth's where the print is hidden. We asked how that field reaches the output by replacing it, at test time and with everything else fixed, before it enters the decoder (Table 31). In SweepFill the decoder barely reads it, since zeroing the field moves the output's orientation error by 0.8°, rotating it by 90° in the hidden region by 1.7°, and even the ground-truth field only brings it to 10.72°. The head's contribution is therefore made during training, as a target that forces the bottleneck tokens to encode flow and period for the whole print, which the decoder then draws from.

The decoder can be made to follow the field. We fine-tuned SweepFill for two epochs, injecting the ground-truth field in place of the predicted one for half of the samples and adding a loss that ties the output's block orientation to the injected field in the hidden region, against a control fine-tuned with the same recipe and neither change. The field then steers the ridges. Rotated by 90° it takes the output's error to 50.6°, zeroed to 32.1°, and the ground-truth field brings the same weights from 10.5° to 6.9°. With the predicted field, orientation error moves from 11.02° (control) to 10.55° and band contrast from 0.669 to 0.708, at 24.30 against 24.21 dB hidden PSNR, while the deep matcher's EER at w=0.3 is unchanged (12.5% against 12.5% for the control). We therefore keep SweepFill as trained. What the experiment adds is a measured lever, since a decoder that follows its field turns every degree gained in structure prediction into output flow, and the ground-truth field shows how far that could go.

Table 29: **Equal-budget ablation**, components cumulative, identical data, schedule and compute (pos is mask token before positional embedding, str structure head, spc FFC decoder and spectral loss, and ref sweep references). "restore" is whole-crop PSNR at w=1, while other columns are hidden-region metrics, with "OF" and "band" as in Table 27 (1 ideal, 0 flat skin for "band"). Bold is best per column within the ladder. The second block holds three single-component cells at the same budget, references without the structure and spectral components, and the FFC decoder and the spectral loss alone. The third removes one component at a time from the full model ("spec" is spectral loss on a plain decoder). Figure 3 plots the ladder's end points across every window.

	component				restore	hidden region @ w=0.6			hidden region @ w=0.3		
model	pos	str	spc	ref	PSNR ↑	PSNR ↑	OF° ↓	band	PSNR ↑	OF° ↓	band
DiT-UNet restorer					32.90	25.76	37.33	0.154	20.84	21.54	0.158
+ mask first	✓				32.97	25.82	38.31	0.143	21.13	21.13	0.134
+ structure head	✓	✓			**33.10**	25.99	38.92	0.140	21.14	21.09	0.122
+ FFC decoder & spectral loss	✓	✓	✓		32.88	25.83	28.74	0.541	21.01	14.84	0.518
+ sweep references (SweepFill)	✓	✓	✓	✓	32.95	**26.62**	**26.97**	**0.566**	**23.26**	**13.17**	**0.595**
mask first + sweep references only	✓			✓	32.98	26.61	37.68	0.153	23.05	20.69	0.201
structure head + FFC decoder only	✓	✓	FFC		32.90	25.92	37.98	0.161	21.10	20.67	0.120
structure head + spectral loss only	✓	✓	spec		33.08	25.83	27.88	0.544	20.99	14.67	0.525
full model, position first		✓	✓	✓	32.93	26.46	28.88	0.536	22.97	16.36	0.560
full model without structure head	✓		✓	✓	32.89	26.55	30.74	0.551	23.10	17.39	0.552
full model without FFC decoder	✓	✓	spec	✓	33.14	26.62	26.00	0.569	23.22	13.31	0.582

Table 30: **Baselines**, hidden-region image metrics on the identity-protocol probes (96 probes at windows $w \in \{0.5, 0.3\}$, since at w=0.7 no block is hidden with the 8-px margin the flow metrics need). With identical inputs per method, bold marks the best per column, with band and "orient." scored by distance from 1, since 1 is true contrast. "orient." counts band energy only within $\pm 20°$ of the true ridge direction. The classical pipeline's band exceeds 1 because it paints high-contrast ridges in the wrong places, as its orientation error and its lower oriented contrast show. LaMa fine-tuned on our data with our pixel losses at the equal budget (Appendix H) restores the window (its PSNR) but, like the DiT-UNet restorer, draws flat skin outside it (its band), while with LaMa's own losses it draws near-true contrast whose flow remains further from truth than SweepFill's.

	w=0.5				w=0.3			
method	PSNR ↑	OF° ↓	band	orient.	PSNR ↑	OF° ↓	band	orient.
Gabor extrapolation (Hong et al., 1998)	9.25	35.66	2.695	2.014	10.29	23.44	1.891	1.443
LaMa (zero-shot) (Suvorov et al., 2022)	18.03	32.05	**0.968**	**0.748**	11.14	26.80	**0.804**	0.612
LaMa, fine-tuned (equal budget)	24.81	40.37	0.101	0.086	21.43	20.56	0.117	0.112
LaMa, fine-tuned with its own losses (equal budget)	24.05	28.96	0.658	0.558	21.15	16.81	0.719	0.658
SweepFill (ours)	**26.94**	**23.09**	0.640	0.536	**25.00**	**10.90**	0.715	**0.660**

K.3 The identity protocol in full

Table 1 carries only the degraded probe and SweepFill. Table 32 adds every other completion on the same probes, gallery and fixed thresholds, and Figure 11 plots both datasets against the visible fraction with the bootstrap intervals drawn. The remaining three sharpen the reading. Table 33 moves the operating point to the strictest false-accept rate this gallery resolves and carries identification out to rank-5, Table 34 is the paired bootstrap the main text quotes, and Table 35 sets completion against the multi-frame fusion a video capture could do instead. Figure 12 shows the score distributions behind the leakage columns of Table 2.

K.4 Mosaicking the sweep frames

The classical way to fill a missing area is to paste it in from another impression. We build that baseline from the same two frames SweepFill reads. Each reference is registered to the probe by an affine ECC fit on gradient images over the probe's visible window, and the better-correlated one is warped and pasted outside the window behind a 6-px feathered seam. Two oracles bound what registering these frames could deliver at all, because they register to the clean frame instead, one a global affine fit and the other the same fit followed by a shift of up to 6 px per 32-px block. Table 36 scores them on the w=0.3 identity probes.

Table 31: **The injected structure field, replaced at test time**, validation sweep at w=0.3. Orientation error of the output against the ground truth, and against the field that was injected ("follow"). Left is SweepFill as trained, right is after the structure-following fine-tune.

	SweepFill			structure-following fine-tune		
field given to the decoder	OF° ↓	follow°	PSNR ↑	OF° ↓	follow°	PSNR ↑
predicted by the head (default)	11.0	16.7	24.32	10.5	14.7	24.20
zeroed	11.8	45.0	24.25	32.1	45.0	17.83
rotated by 45° in the hidden region	11.9	43.2	24.35	31.5	21.9	18.61
rotated by 90° in the hidden region	12.7	71.1	24.37	50.6	28.4	16.91
ground-truth field (oracle)	10.7	18.7	24.35	6.9	14.8	24.87

Table 32: **Identity evaluation on MMFV, all systems**, the rows of Table 1 plus every other completion on identical probes, gallery and protocol, namely the DiT-UNet restorer, its PatchGAN fine-tune (the adversarial control of Appendix I), classical Gabor extrapolation, zero-shot LaMa, and LaMa fine-tuned on our data under our losses and under its own. Bold marks the best value at each window, the ceiling excluded. The indented rows repeat SweepFill with references degraded like the probe, and SweepFill, with clean or degraded references, takes every EER and rank-1 under the deep matcher, whereas under the minutiae matcher the PatchGAN control takes most cells, the split Appendix I accounts for. Plotted in Figure 11.

	deep matcher (Ridgeformer)			minutiae matcher (LEADER+Bozorth3)		
probe	EER% ↓	TAR% ↑	R-1% ↑	EER% ↓	TAR% ↑	R-1% ↑
clean full-view frame *(ceiling)*	4.2	90.6	94.8	7.0	83.3	89.6
degraded probe, w=1.0	11.5	61.5	81.2	29.9	50.0	58.3
degraded probe, w=0.7	15.8	40.6	59.4	31.2	50.0	52.1
degraded probe, w=0.5	20.8	33.3	53.1	28.7	49.0	54.2
degraded probe, w=0.3	26.0	14.6	21.9	38.1	33.3	40.6
DiT-UNet restorer, w=0.7	10.4	49.0	83.3	26.0	55.2	67.7
DiT-UNet restorer, w=0.5	14.6	33.3	60.4	27.2	57.3	61.5
DiT-UNet restorer, w=0.3	19.6	24.0	46.9	26.6	49.0	57.3
restorer + PatchGAN fine-tune (Isola et al., 2017), w=0.7	9.2	**67.7**	83.3	**17.9**	**68.8**	71.9
restorer + PatchGAN fine-tune, w=0.5	12.5	53.1	77.1	**20.2**	**65.6**	**70.8**
restorer + PatchGAN fine-tune, w=0.3	21.9	39.6	52.1	**19.9**	**61.5**	**70.8**
Gabor extrapolation (Hong et al., 1998), w=0.7	14.6	36.5	54.2	28.4	38.5	45.8
Gabor extrapolation, w=0.5	25.0	20.8	35.4	34.5	36.5	41.7
Gabor extrapolation, w=0.3	40.5	2.1	7.3	40.3	19.8	21.9
LaMa (zero-shot) (Suvorov et al., 2022), w=0.7	11.6	46.9	75.0	28.4	45.8	56.2
LaMa (zero-shot), w=0.5	17.8	35.4	55.2	28.0	49.0	55.2
LaMa (zero-shot), w=0.3	22.7	6.2	11.5	31.0	31.2	38.5
LaMa, fine-tuned, w=0.7	14.6	42.7	70.8	31.3	47.9	56.2
LaMa, fine-tuned, w=0.5	16.7	40.6	67.7	27.1	50.0	57.3
LaMa, fine-tuned, w=0.3	18.6	21.9	40.6	32.8	37.5	47.9
LaMa, fine-tuned with its own losses, w=0.7	9.4	58.3	72.9	25.5	47.9	58.3
LaMa, fine-tuned with its own losses, w=0.5	13.5	**65.6**	70.8	25.9	51.0	56.2
LaMa, fine-tuned with its own losses, w=0.3	17.9	31.2	46.9	33.6	37.5	43.8
SweepFill (ours), w=1.0	**8.3**	**78.1**	**90.6**	**20.3**	**69.8**	**77.1**
SweepFill (ours), w=0.7	**7.3**	66.7	**88.5**	19.0	64.6	**77.1**
SweepFill (ours), w=0.5	**9.4**	64.6	83.3	22.8	60.4	67.7
SweepFill (ours), w=0.3	**11.6**	**57.3**	**79.2**	23.2	60.4	64.6
references degraded like the probe, w=0.7	**7.3**	63.5	**88.5**	19.1	65.6	76.0
references degraded like the probe, w=0.5	10.2	64.6	**84.4**	27.7	62.5	66.7
references degraded like the probe, w=0.3	14.4	52.1	77.1	25.0	58.3	67.7

The deployable mosaic helps the deep matcher far less than completion (16.7% EER and 54.2% rank-1, against 11.6% and 79.2% for SweepFill) and ties it under the minutiae matcher (23.8% against 23.2%), because a 30% window is too little to register a frame taken at another pose. Knowing the answer does not close the gap either. Registered to the clean frame by the affine oracle, the references reach 11.5% EER but only 67.7% rank-1, and even with the per-block shift their hidden-region flow is 16.3° off against SweepFill's 10.9°. The shifted oracle does place more true minutiae (30% against 21%), at 70% spurious, but its phase coherence with the clean frame is 0.07, far below the 0.55 of the visible region. Between two poses the finger deforms non-rigidly, so even a registration that sees the target recovers little of the exact ridge pattern, so the phase gap of Appendix F.2 is largely a property

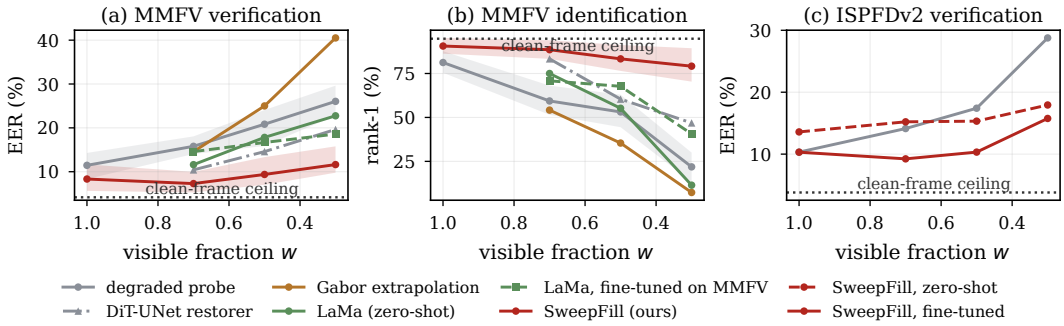

Figure 11: **Matcher performance against the visible fraction.** (a,b) MMFV, deep matcher, where completion improves identification at every window and verification wherever part of the print is hidden, more so as the window shrinks. Shaded bands are 95% cluster-bootstrap intervals over probe identities (Appendix B.1) for the degraded probe and SweepFill, both tabulated in Table 5, Appendix B.1. Dash-dotted triangles mark the DiT-UNet restorer and dashed squares LaMa fine-tuned on our data. (c) ISPFDv2, where the in-domain fine-tune (solid) improves every window while the zero-shot model (dashed) crosses the degraded probe near w=0.6, the domain gap of Section 5.5.

Table 33: **Stricter operating point and the full identification curve**, MMFV, deep matcher. With 32 enrolled identities the gallery affords 2,976 impostor comparisons, so 0.1% is the strictest false-accept rate this protocol resolves. Completion helps at both rates and every rank, and rank-5 recovers most of what rank-1 misses, so residual failures are near-misses rather than collapses.

		TAR% at FAR		identification rate%		
probe	EER% ↓	1%	0.1%	rank-1 ↑	rank-2	rank-5
clean full-view frame *(ceiling)*	4.17	90.6	68.8	94.8	97.9	99.0
degraded probe, w=1.0	11.46	61.5	31.2	81.2	89.6	95.8
SweepFill (ours), w=1.0	**8.33**	**78.1**	**49.0**	**90.6**	**94.8**	**97.9**
degraded probe, w=0.7	15.79	40.6	18.8	59.4	81.2	94.8
SweepFill (ours), w=0.7	**7.29**	**66.7**	**37.5**	**88.5**	**94.8**	**96.9**
degraded probe, w=0.5	20.83	33.3	16.7	53.1	65.6	85.4
SweepFill (ours), w=0.5	**9.38**	**64.6**	**40.6**	**83.3**	**90.6**	**95.8**
degraded probe, w=0.3	26.04	14.6	5.2	21.9	30.2	54.2
SweepFill (ours), w=0.3	**11.63**	**57.3**	**15.6**	**79.2**	**85.4**	**95.8**

Table 34: **Paired cluster bootstrap**, completed minus degraded on the *same* probe identities (2000 resamples), stricter than overlap of two separately-estimated intervals since the same identities drive both terms. Every window improves identification, while verification improves significantly wherever part of the print is hidden, but not at w=1.0, where the task is pure restoration.

	Δ EER% (lower is better)			Δ rank-1%		
window	estimate	95% CI	excludes 0	estimate	95% CI	excludes 0
w=1.0	-3.12	[-7.16, +1.24]	–	+9.4	[+3.1, +15.6]	✓
w=0.7	-8.50	[-11.46, -5.11]	✓	+29.2	[+18.8, +39.6]	✓
w=0.5	-11.46	[-15.24, -5.21]	✓	+30.2	[+17.7, +42.7]	✓
w=0.3	-14.42	[-19.69, -6.13]	✓	+57.3	[+41.7, +70.8]	✓

of the sweep, and the flow SweepFill generates is closer to the truth than any frame of it we could register.

Table 35: **Completion against multi-frame fusion**, deep matcher, MMFV. Max-score score-level fusion over the other frames is the strongest deployed alternative. Top is a single degraded frame, where fusion is unavailable and SweepFill runs reference-free. Bottom is the full capture, with fusion and SweepFill given the same two degraded auxiliary frames (the completion's references), so inputs are identical. Bold marks the better of each pair. Paired cluster-bootstrap intervals (2000 resamples) on each pair's EER difference at w=0.7/0.5/0.3 are -3.1 $[-7.0, +0.4]$, -2.2 $[-6.5, +0.5]$, -2.1 $[-5.5, +0.5]$ (fused) and -8.3 $[-11.1, -4.8]$, -8.3 $[-13.6, -2.5]$, -8.4 $[-14.8, -4.0]$ (single frame), and on the fused rank-1 difference they are [+4.2, +19.8], [+2.1, +15.6] and [-2.1, +6.2].

configuration	EER% ↓	TAR% @ 1% FAR ↑	rank-1% ↑
single degraded frame, no auxiliary frames			
degraded probe, w=0.7	15.79	40.6	59.4
probe completed by SweepFill (reference-free), w=0.7	**7.48**	**63.5**	**87.5**
degraded probe, w=0.5	20.83	33.3	53.1
probe completed by SweepFill (reference-free), w=0.5	**12.50**	**55.2**	**82.3**
degraded probe, w=0.3	26.04	14.6	21.9
probe completed by SweepFill (reference-free), w=0.3	**17.64**	**32.3**	**55.2**
full capture: probe and two auxiliary frames, max-score fusion			
two auxiliary frames only, clean	6.15	84.4	93.8
two auxiliary frames only, degraded	12.47	65.6	78.1
degraded probe + two auxiliary frames, w=0.7	8.33	67.7	81.2
probe completed by SweepFill, then fused, w=0.7	**5.21**	**80.2**	**92.7**
degraded probe + two auxiliary frames, w=0.5	10.53	67.7	78.1
probe completed by SweepFill, then fused, w=0.5	**8.33**	**75.0**	**86.5**
degraded probe + two auxiliary frames, w=0.3	10.42	65.6	78.1
probe completed by SweepFill, then fused, w=0.3	**8.33**	**71.9**	**80.2**

Table 36: **Mosaicking from the same two frames**, identity protocol at w=0.3 (96 probes). The mosaic registers each reference on the probe's window and pastes it, whereas the two oracle rows (gray) register to the clean frame and cannot be run at test time. Columns give hidden-region orientation error (OF, Section 4), minutiae recovered and spurious in the hidden region (15 px, 30°, pooled over probes), and phase coherence with the clean frame (Appendix F.2).

	Ridgeformer		LEADER+Bozorth3			minutiae (%)		
completion of the hidden region	EER% ↓	R-1% ↑	EER% ↓	R-1% ↑	OF° ↓	recovered ↑	spurious ↓	phase
degraded probe, no completion	26.0	21.9	38.1	40.6	45.0	–	–	–
mosaic: references registered on the window, pasted	16.7	54.2	23.8	57.3	22.9	25	75	0.010
oracle: affine registration to the clean frame	11.5	67.7	27.2	61.5	16.9	27	72	0.036
oracle: affine + per-block shift to the clean frame	13.5	65.6	28.3	50.0	16.3	30	70	0.073
SweepFill	**11.6**	**79.2**	**23.2**	**64.6**	**10.9**	**21**	**55**	**0.011**

Table 37: **External validity on ISPFDv2** (Malhotra et al., 2020), with 92 identities, both matchers and probe construction as in Table 1. Bold is best of three systems per row. Fine-tuned on 2,980 crops from 51 subjects disjoint from these identities, SweepFill improves every window under both matchers bar one deep-matcher EER tie at w=1.0, whereas zero-shot with MMFV weights it helps at low visibility, hurts at full visibility, and barely moves the minutiae matcher.

	degraded probe		SweepFill, zero-shot		SweepFill, fine-tuned	
window	EER% ↓	R-1% ↑	EER% ↓	R-1% ↑	EER% ↓	R-1% ↑
deep matcher (Ridgeformer)						
clean full-view frame *(ceiling)*	3.8	92.4	–	–	–	–
w=1.0	**10.3**	68.5	13.6	59.2	**10.3**	**71.7**
w=0.7	14.1	58.7	15.2	52.7	**9.2**	**63.6**
w=0.5	17.4	37.0	15.3	44.6	**10.3**	**62.0**
w=0.3	28.8	14.1	17.9	31.0	**15.8**	**41.8**
minutiae matcher (LEADER+Bozorth3)						
clean full-view frame *(ceiling)*	4.7	96.2	–	–	–	–
w=1.0	27.9	58.2	32.6	57.1	**19.4**	**67.4**
w=0.7	32.3	48.4	33.1	52.2	**23.6**	**66.8**
w=0.5	35.6	40.8	33.9	48.4	**23.5**	**59.8**
w=0.3	43.8	23.9	42.0	30.4	**30.0**	**47.8**

K.5 External validity on ISPFDv2

Table 37 is the second benchmark of Section 5.5 under both matchers, zero-shot and after the in-domain fine-tune. The deep matcher's training data include ISPFDv2 but not MMFV (Appendix H), so on this dataset it is the minutiae matcher, which has no training set at all, that reads independently.

K.6 Qualitative panels

Figure 13 follows one held-out subject as the window closes, beside the restorer and the field the structure head predicts for the whole print, and Figure 14 repeats that layout on further subjects at w=0.3 and shows the reference frames the model reads. Figure 15 is the w=1 case, where nothing is hidden and the network only restores.

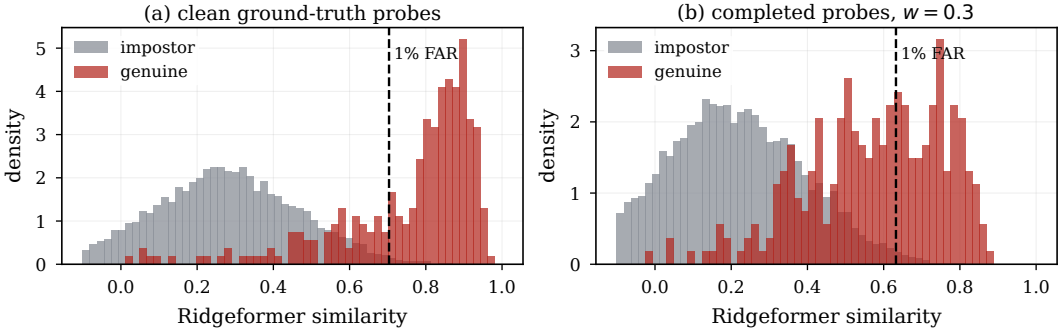

Figure 12: **Genuine and impostor score distributions** under the deep matcher, both panels sharing one threshold calibrated at 1% FAR on the clean probes, the fixed threshold of Table 2. Completion shifts genuine scores down relative to clean frames yet leaves *less* impostor mass above that threshold, not more, which is what the leakage columns of Table 2 report. At the same threshold, 10.9% of uncompleted w=0.3 probe pairs from different identities match on MMFV (10.0% on ISPFDv2), against 1.9% of completion pairs and 1.3% of clean frames.

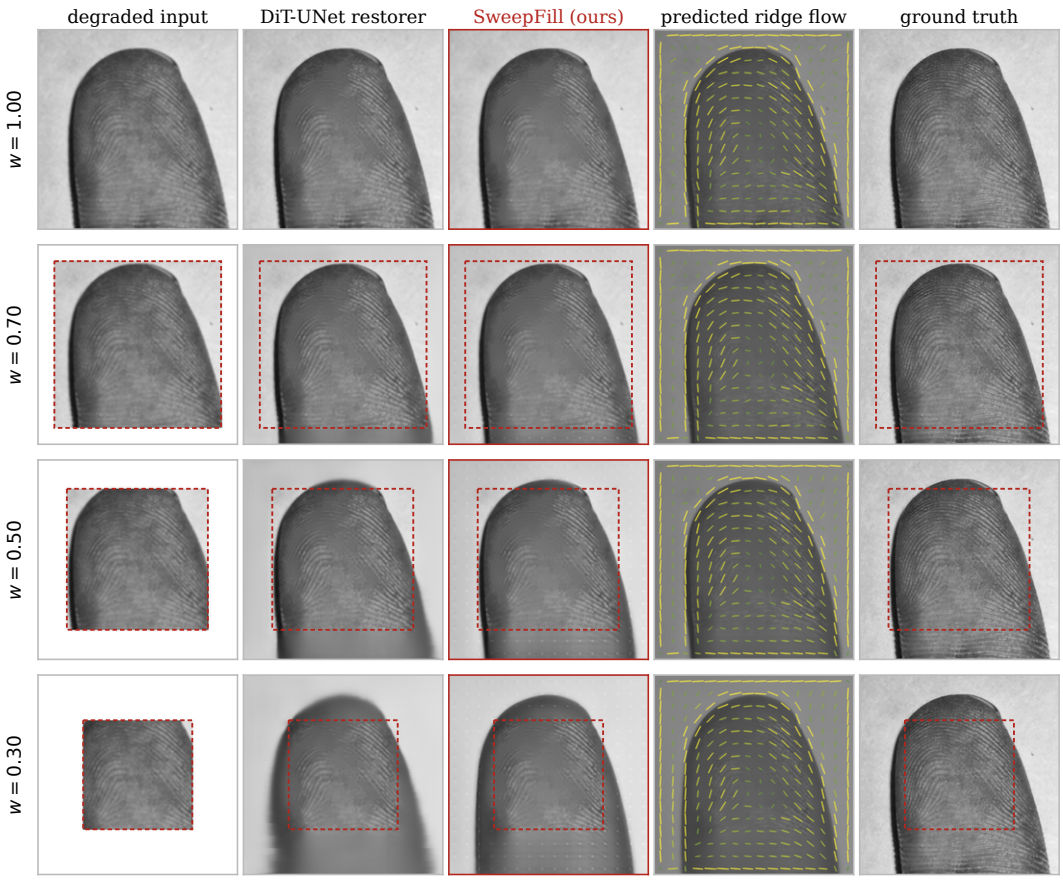

Figure 13: **One held-out subject as the window shrinks.** The DiT-UNet restorer fades to ridge-less skin below w=0.7, whereas SweepFill continues coherent ridge flow, guided by the whole-print field from its structure head (fourth column, where stroke direction is predicted orientation and length confidence). Dashed outlines mark the visible window. At w=0.3 the lower fingertip stays smoother than ground truth, the residual gap quantified in Section 5.4.

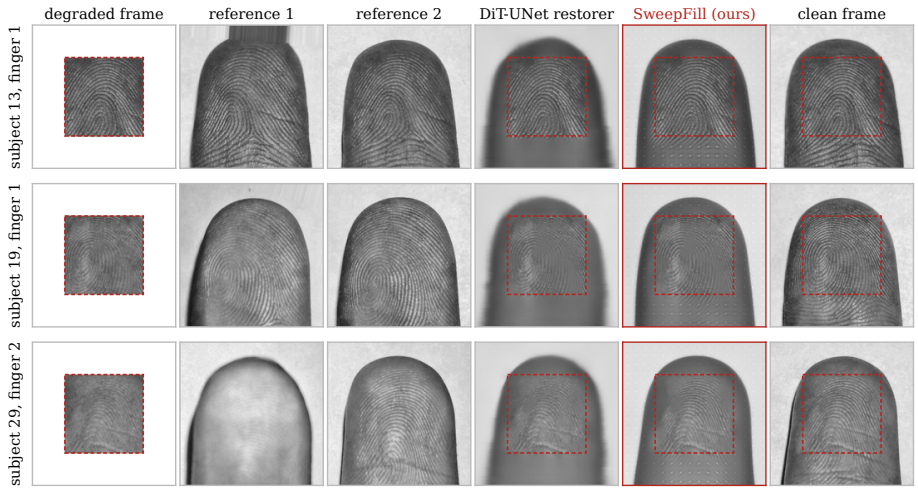

Figure 14: **Additional held-out subjects at** w=0.3, the panels of Figure 1 without its ridge-flow row, showing the degraded frame, the two other sweep frames SweepFill reads as references, the two completions and the clean target. Dashed outlines mark the visible window, and everything outside is generated.

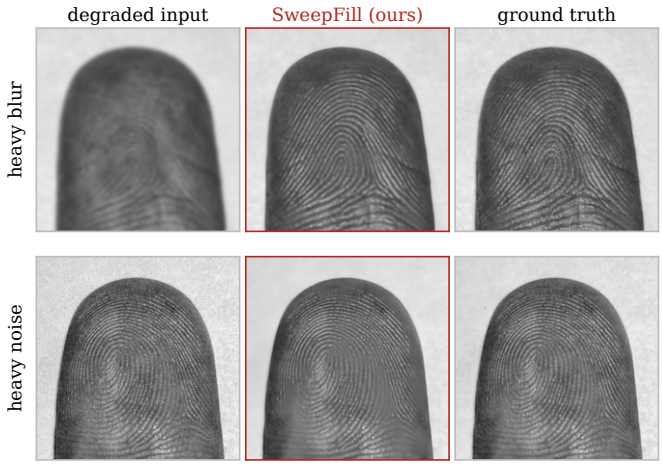

Figure 15: **Restoration at full visibility.** With nothing hidden the same network acts as a restorer, with the completion machinery idle while the residual head cleans the observed print.