FORENSICFACE: RELIABLE 3D FACE TEXTURE RECONSTRUCTION VIA CONFIDENCE GUIDED DIFFUSION

Anonymous authors
Paper under double-blind review

ABSTRACT

Reliable 3D face texture reconstruction from a single 2D facial image is of critical importance in criminal investigations. Existing methods usually use visible texture directly as a generative condition, but they overlook differences in observation confidence across facial regions. As a result, unreliable projections may be retained incorrectly, while genuine local details such as moles and skin tone variations may be smoothed or rewritten during generation. To address these problems, we present ForensicFace, a framework for reliable single image 3D face texture reconstruction using a canonical UV representation. First, we propose texel level observation confidence estimation to model the reliability of input texture continuously, addressing the inability of binary visibility masks to distinguish the reliability of different observed regions. We then design an observation confidence guided diffusion model that controls generative freedom according to observation confidence. Regions with high observation confidence preserve more of the input observation, whereas regions with low observation confidence or no direct observation rely more on the generative prior for completion. This design prevents reliable observations from being overwritten by the generative prior while reducing the risk of directly retaining unreliable projections. Finally, we estimate generation confidence from consistency across repeated stochastic samples to characterize the stability of generated regions. Experiments on FaceScape and Stirling ESRC show that ForensicFace outperforms existing methods in PSNR, SSIM, LPIPS, and CSIM while maintaining stable texture reconstruction quality under large pose variations.

1 INTRODUCTION

Reliable 3D face texture reconstruction from a single 2D facial image is particularly important in criminal investigations, where facial appearance is represented in canonical UV space to support cross pose analysis, 3D visualization, and manual examination. Forensic use requires not only complete and natural textures but also faithful preservation of genuine input details. To address these issues, recent methods have moved from low dimensional statistical models and model fitting toward GAN and diffusion based generative approaches that infer unobserved surfaces from limited visible regions and substantially improve completeness, realism, detail quality, and identity consistency (Gecer et al., 2019; 2021; Bai et al., 2023; Li et al., 2024; Yang et al., 2025). Stronger generative priors and identity constraints further improve visual fidelity and natural appearance (Qiu et al., 2026; Xiao et al., 2026). However, existing methods still focus primarily on the final complete texture, with comparatively less attention to whether observed evidence is preserved correctly during generation. This is critical in forensic analysis because identity cues such as moles, local pigmentation, skin tone variation, and fine texture may be smoothed, modified, or regenerated. The central challenge is therefore to complete missing texture without compromising trustworthy input observations.

Single image texture reconstruction faces three key challenges. First, large pose variation, self occlusion, contour projection distortion, and dense correspondence errors can cause substantial differences in observation quality even within visible regions. A binary visible or invisible partition therefore cannot accurately describe the reliability of the input evidence. Second, even when more reliable regions can be identified, the generative model must still determine how strongly it should follow each input observation. Excessive observation constraints can preserve erroneous projections, whereas excessive generative freedom can overwrite genuine local details. Third, texture on the blind side and in other unobserved regions is inferred mainly from the generative prior, and a single

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.

generated result cannot indicate whether the model is stable at a given location. The generation confidence of prior completed regions must therefore be characterized explicitly.

To address these challenges, we propose **ForensicFace**, a confidence guided framework for reliable 3D face texture reconstruction. ForensicFace first establishes traceable correspondences from texels in canonical UV space to source image pixels and constructs continuous observation confidence by integrating semantic validity, geometric visibility, source correspondence reliability, projection distortion, boundary effects, and imaging factors. We then introduce a confidence guided observation preservation strategy that adaptively injects the input observation at every reverse diffusion step according to observation confidence. Regions with high observation confidence remain strongly constrained by the original evidence, regions with low observation confidence permit limited generative adjustment, and unobserved regions are completed by the complete texture prior. For locations that depend primarily on prior generation, we further estimate generation confidence from dispersion across random seeds, thereby explicitly distinguishing observation from generation.

Our main contributions are summarized as follows:

- We propose conventional binary visibility to a continuous observation confidence representation that characterizes the reliability of individual input texels.
- We use observation confidence to adaptively control generative freedom during diffusion, preserving reliable input details while completing missing texture.
- We estimate the stability of generated regions from consistency across multiple random seeds, providing a spatial confidence description for results completed by the generative prior.

2 Related Work

Single image face texture reconstruction. Single image face texture reconstruction aims to reconstruct complete facial appearance from limited 2D evidence, with unobserved regions usually completed by statistical models or generative priors. GANFIT and OSTeC established representative pipelines that use generative texture priors to reconstruct complete high quality facial appearance (Gecer et al., 2019; 2021). With the development of canonical UV data and stronger generative models, subsequent work increasingly focused on high fidelity texture generation and identity preservation (Bai et al., 2023; Li et al., 2024). Recent methods such as FreeUV and OMGTex further reduce the dependence on complete UV supervision and explicit geometric modeling (Yang et al., 2025; Qiu et al., 2026; Xiao et al., 2026). Overall, existing approaches have substantially improved completeness, realism, and identity consistency, but they mainly optimize the quality of the final complete texture and devote less attention to the distinct origins of observed and generated content.

Observation constraints during generation. Generative models may alter existing evidence while completing missing content, making preservation of input constraints a central problem in image editing and restoration. SDEdit regulates the relation between input structure and the generative prior by controlling the starting state of diffusion, whereas RePaint repeatedly injects known regions during reverse diffusion (Meng et al., 2022; Lugmayr et al., 2022). Subsequent studies extend diffusion priors to general inverse problems and conditional image restoration and explore more flexible forms of observation guidance (Kawar et al., 2022; Fei et al., 2023). DynFaceRestore further considers spatial variation in degradation severity during face restoration (Do et al., 2025). In contrast, observation quality in single image facial UV reconstruction varies strongly across space because of pose, occlusion, projection distortion, and correspondence error. We therefore focus on adapting the strength of observation constraints according to texel level observation confidence.

Generation confidence. When insufficient input evidence prevents a unique output, prior work commonly characterizes predictive uncertainty through model ensembles, Bayesian approximations, or repeated stochastic predictions (Lakshminarayanan et al., 2017; Kendall & Gal, 2017). With diffusion models, recent studies further analyze uncertainty in image space, semantic space, and denoising trajectories and use uncertainty information for generation control and 3D reconstruction (Franchi et al., 2025; De Vita & Belagiannis, 2025; Bose et al., 2025). These studies demonstrate the value of spatially characterizing output reliability when observations are incomplete. In this work, generation confidence denotes texel level consistency across stochastic samples generated under identical input conditions and is used to describe stability in regions without direct observation.

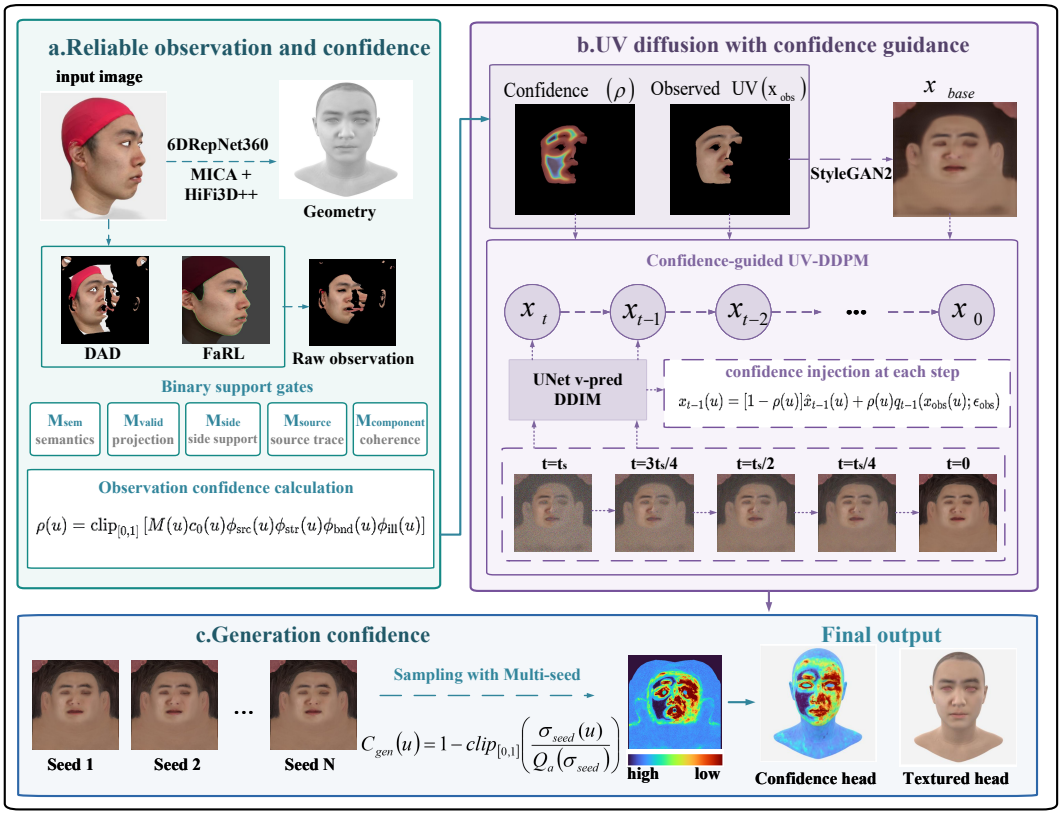


Figure 1: Overview of ForensicFace. (a) The input image is mapped to traceable observations in canonical UV space, where texel level observation confidence is computed. (b) During diffusion, input observations are injected adaptively according to confidence, preserving reliable texture while completing missing regions. (c) Under fixed input conditions, multiple random seeds are sampled and generation confidence is estimated from consistency across samples.

In summary, previous studies have advanced complete texture generation, observation constraints during generation, and uncertainty modeling, but these directions are usually treated separately. Single image 3D face texture reconstruction still lacks a unified framework that addresses three questions together: which input observations should be preserved, how strongly different observations should constrain generation, and how stable generated regions are when direct observation is unavailable. Accordingly, ForensicFace unifies texel level observation confidence, confidence guided diffusion, and multi seed generation confidence in one texture reconstruction framework.

3 METHOD

Figure 1 shows the overview workflow of ForensicFace. ForensicFace reconstructs complete 3D face texture as follows: Given a single facial image, the system first uses existing head pose, 3D geometry, and dense correspondence estimators to establish an image-to-UV mapping and obtain an initial observed texture in canonical UV space. Because these projections have different quality, we compute observation confidence for each texel to describe whether the input observation at that location should be preserved. We then obtain a complete UV texture basis and continuously inject the input observation during diffusion according to observation confidence. Regions with high observation confidence mainly follow the original input, regions with low observation confidence permit model adjustment, and unobserved regions are completed mainly by the generative prior. Finally, while keeping the input conditions fixed, we repeat generation with multiple random seeds and compute generation confidence from consistency across the resulting samples. The system outputs a complete UV texture, a textured 3D head, and the corresponding generation confidence.

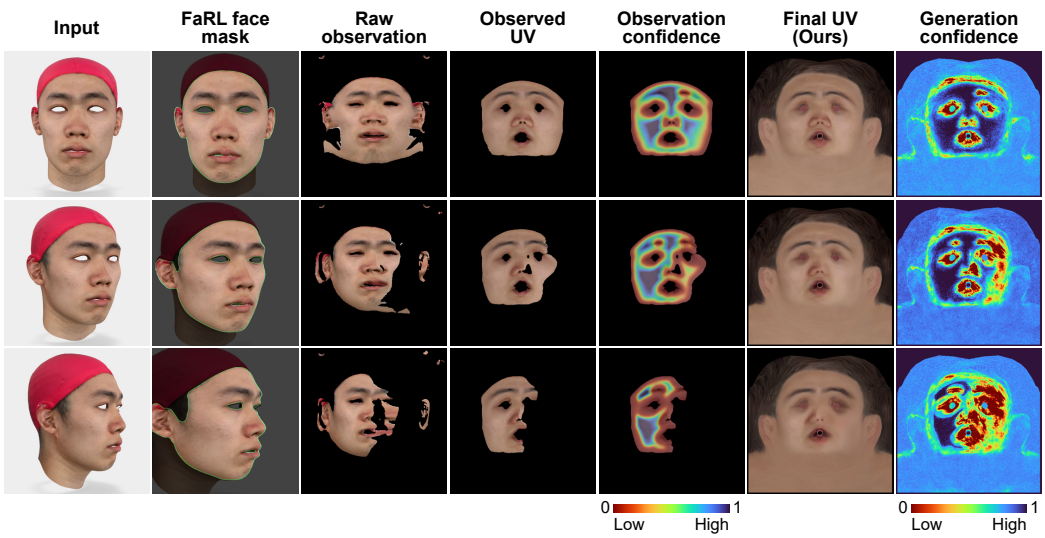


Figure 2: Reliable texture reconstruction under different input poses. As yaw increases, directly observed facial regions become smaller and observation confidence ρ contracts accordingly. ForensicFace still reconstructs a complete UV texture and uses generation confidence C_{gen} to indicate the stability of prior completed regions under repeated sampling. Observation confidence and generation confidence are both displayed from low to high.

3.1 Observation Confidence Modeling

Using existing head pose, parametric geometry, and dense correspondence models (Hempel et al., 2024; Zielonka et al., 2022; Martyniuk et al., 2022), we establish a fixed image-to-UV mapping and obtain candidate observations in canonical UV space. For each valid texel u, the system stores its observed color $x_{\text{obs}}(u)$ and the corresponding source image coordinate $T_{\text{src}}(u)$.

Colors obtained by direct back projection cannot all be treated as reliable observations. Projection stretching near contours, occlusion, hair, shadows, and local highlights can introduce unstable colors into UV space. After applying FaRL face parsing to exclude background, hair, and evident non skin regions (Zheng et al., 2022), we retain only texels that are semantically valid, geometrically visible, validly mapped, traceable to source pixels, and part of the main facial connected component. These conditions jointly define a binary support mask $M(u)$. The binary support mask removes evident projection errors but cannot represent continuous variation in observation quality within the retained region. We therefore define texel level observation confidence as

$$\rho(u) = \text{clip}_{[0,1]}[M(u)c_0(u)\phi_{\text{src}}(u)\phi_{\text{str}}(u)\phi_{\text{bnd}}(u)\phi_{\text{ill}}(u)]\,. \tag{1}$$

Here, c_0 describes basic geometric visibility and local projection quality, ϕ_{src} measures source correspondence reliability, ϕ_{str} suppresses unstable texture caused by back projection stretching, ϕ_{bnd} introduces smooth attenuation near observation boundaries, and ϕ_{ill} reduces the weight of severe shadow, overexposure, and specular reflection. The value $\rho(u) \in [0, 1]$ represents the strength with which a texel should follow the input observation. A larger value indicates stronger observation confidence, whereas a smaller value permits greater generative adjustment.

3.2 Observation Confidence Guided Complete UV Reconstruction

A single image provides only partial surface observations, so complete texture reconstruction still requires a generative prior. We train a diffusion prior, UV DDPM (Ho et al., 2020), on FFHQ-UV to model the distribution of complete facial textures in canonical UV space. During inference, we first perform W^+ inversion in the StyleGAN2 texture space and apply limited identity adaptation with Pivotal Tuning Inversion (Karras et al., 2020; Abdal et al., 2019; Roich et al., 2022) to obtain a complete UV basis x_{base}. This basis provides global color distribution and initial structure for unobserved regions. UV DDPM then performs diffusion based coordination over this basis so that the complete texture prior and reliable observations can be integrated within a single sampling trajectory. Following SDEdit (Meng et al., 2022), we first add noise to the complete basis at an intermediate

time step t_s:

$$x_{t_s} = q_{t_s}(x_{\text{base}}; \epsilon_{\text{base}}), \tag{2}$$

where $q_t(\cdot; \epsilon)$ denotes the forward diffusion process and ϵ_{base} is the initialization noise for the basis. Reverse diffusion starts from t_s rather than from pure noise.

The key operation in ForensicFace is confidence guided observation injection at every reverse diffusion step. Let $\hat{x}_{t-1}$ denote the unconstrained denoising candidate produced by the current DDIM step (Song et al., 2021), and let $q_{t-1}(x_{\text{obs}}; \epsilon_{\text{obs}})$ denote the noisy observation constructed with fixed observation noise ϵ_{obs} and aligned with time step $t-1$. The update is

$$x_{t-1}(u) = \big[1 - \rho(u)\big]\hat{x}_{t-1}(u) + \rho(u) q_{t-1}(x_{\text{obs}}(u); \epsilon_{\text{obs}}). \tag{3}$$

This update formulates observation preservation and texture completion as continuous spatial control. When $\rho(u) \approx 1$, the sampling trajectory remains close to the input observation at that location and fine identity details are less likely to be rewritten by the generative prior. When $0 < \rho(u) < 1$, the model can make limited adjustments to local exposure, color, and boundary discontinuity. When $\rho(u) = 0$, the current texel is determined entirely by the complete texture prior. Compared with post generation texture pasting or unconstrained generation, progressive injection coordinates observed and generated regions throughout reverse diffusion while continuously constraining reliable input evidence.

3.3 Generation Confidence

Observation confidence constrains visible regions, but a single image still cannot uniquely determine texture on the blind side and in rear regions. For locations completed mainly by the generative prior, a visually natural output does not imply that the model makes a stable estimate. ForensicFace therefore fixes x_{obs}, ρ, x_{base}, the sampler, and all other conditions, while varying only the diffusion random seed. For the result $x^{(n)}$ generated by seed n, we compute the standard deviation across N generated textures for each color channel and average the three RGB channels to obtain the texel level multi seed dispersion $\sigma_{\text{seed}}(u)$. Generation confidence is defined as

$$C_{\text{gen}}(u) = 1 - \text{clip}\left(\frac{\sigma_{\text{seed}}(u)}{Q_\alpha}, 0, 1\right), \tag{4}$$

where Q_α is a robust high quantile scale determined from the σ_{seed} distribution on an independent FFHQ-UV calibration set and fixed before testing. A larger C_{gen} indicates stronger consistency across samples under identical input conditions, whereas a smaller value indicates greater sensitivity to stochastic generation. We use the medoid with the smallest average distance to the remaining samples as the representative final UV texture and report C_{gen} as its generation confidence.

Observation confidence and final generation confidence describe two different information sources. The observation confidence ρ measures how strongly the input evidence should be preserved, whereas the generation confidence C_{gen} measures the consistency of model completion across repeated stochastic samples. ForensicFace separates the final texture into traceable observations and prior completion while marking locations with high and low generation confidence within completed regions.

4 Experiments

4.1 Experimental Settings

Datasets. We train the texture prior and related models on approximately 45,000 canonical facial UV textures from FFHQ-UV (Bai et al., 2023). Evaluation uses two complementary 3D face datasets that are excluded from training. For FaceScape (Yang et al., 2020), we select the first 359 identities to ensure complete texture ground truth and avoid interference from privacy mosaics. For each identity, the neutral expression scan is rendered at yaw angles of $0°$, $30°$, and $60°$, producing 1,077 test samples. For Stirling ESRC (Hancock & Tiddeman, 2013), released by the University of Stirling Psychological Image Collection, we select 133 identities with both textured scans and registered meshes (Zielonka et al., 2022; Li et al., 2017). Only the neutral scan is used for each identity, and inputs are rendered at $0°$, $30°$, and $60°$ using a fixed orthographic camera and a consistent rendering configuration, producing 399 test samples. The main evaluation therefore covers 492 identities and 1,476 rendered inputs across two independent scan sources. Every method receives only RGB input; ground truth geometry, texture, visibility, and target rendering are reserved exclusively for evaluation.

Table 1: Single image texture reconstruction on FaceScape and Stirling ESRC. All methods use the same pose input generation and frontal rendering protocol. Upward arrows indicate that larger values are better, and downward arrows indicate that smaller values are better.

	FaceScape				Stirling ESRC			
Method	PSNR ↑	SSIM ↑	LPIPS ↓	CSIM ↑	PSNR ↑	SSIM ↑	LPIPS ↓	CSIM ↑
Deep3DFaceRecon	18.341	0.690	0.216	0.220	19.602	0.779	0.073	0.231
FFHQ-UV	19.597	0.703	0.187	0.191	21.230	0.770	0.082	0.187
UV-IDM	18.572	0.701	0.224	0.192	19.422	0.781	0.073	0.263
FreeUV	17.083	0.667	0.201	0.335	16.382	0.715	0.098	0.397
OMGTex	17.759	0.638	0.212	0.141	18.645	0.654	0.083	0.138
ForensicFace	**20.874**	**0.710**	**0.182**	**0.373**	**24.792**	**0.814**	**0.050**	**0.464**

Implementation details. We train UV DDPM as the complete texture diffusion prior on approximately 45,000 FFHQ-UV textures. Optimization uses AdamW (Loshchilov & Hutter, 2019) with a learning rate of 1×10^{-4}, a batch size of 12, and 150k training steps. StyleGAN2 initialization uses pretrained weights released with AvatarTex and TexHub (Qiu et al., 2026). During inference, we use the DDIM sampler (Song et al., 2021) with 50 sampling steps and an initial noise strength of 0.40. Generation confidence is computed from $N = 5$ random seeds. All experiments use a single 32 GB vGPU.

Compared methods and evaluation protocol. We compare Deep3DFaceRecon (Deng et al., 2019b), FFHQ-UV (Bai et al., 2023), UV-IDM (Li et al., 2024), FreeUV (Yang et al., 2025), and OMGTex (Xiao et al., 2026), using official code or pretrained weights provided by the authors. To isolate texture reconstruction differences, all outputs are converted to the canonical FLAME UV representation (Li et al., 2017) and rendered on the corresponding neutral FLAME geometry with the same camera, illumination, valid facial mask, and 512×512 resolution. For Stirling ESRC, we use the corresponding FLAME registered meshes released with MICA (Zielonka et al., 2022). All 3D renderings use the PyTorch3D rasterizer.

Metrics. The main comparison evaluates pixel fidelity, structural consistency, perceptual quality, and identity consistency using PSNR, SSIM, LPIPS (Wang et al., 2004; Zhang et al., 2018), and CSIM. CSIM denotes cosine similarity between ArcFace identity features (Deng et al., 2019a) extracted from the reconstructed rendering and the ground truth rendering. Larger values indicate stronger identity consistency. Higher PSNR, SSIM, and CSIM are better, whereas lower LPIPS is better.

4.2 Comparison with Existing Methods

Table 1 reports the quantitative comparison on FaceScape and Stirling ESRC, and Figure 3 shows the qualitative results. Our findings are as follows. (1) ForensicFace consistently achieves the best results on all four metrics over both datasets. Compared with the strongest competing value for each metric, the relative improvements in (PSNR, SSIM, LPIPS, CSIM) are (6.5%, 1.0%, 2.7%, 11.3%) on FaceScape and (16.8%, 4.2%, 31.5%, 16.9%) on Stirling ESRC, respectively, where the LPIPS percentage denotes relative reduction. (2) The improvements are more pronounced on Stirling ESRC. In particular, ForensicFace improves PSNR by 16.8% and reduces LPIPS by 31.5% over the strongest competing results, while the corresponding gains on FaceScape are 6.5% and 2.7%. This suggests that preserving reliable observations remains effective across different scan sources and texture distributions. As shown in Figure 3, existing methods can reconstruct complete facial textures, but the generative prior often reinterprets visible input regions through local color changes, detail smoothing, or texture transfer. ForensicFace instead keeps reliable observations continuously constrained by the input during diffusion while allowing greater generative freedom in low confidence and unobserved regions, reducing unnecessary modification of reliable details without directly preserving poor projections.

The effect of observation preservation is particularly evident for local identity details. Figure 4 compares ForensicFace with two representative baselines selected from Table 1. On FaceScape, FFHQ-UV achieves the strongest competing results on the three reconstruction quality metrics, namely PSNR, SSIM, and LPIPS, whereas FreeUV achieves the highest competing CSIM and thus provides the strongest identity consistency baseline. We therefore use them to examine local observation preservation under both high reconstruction quality and high identity similarity. For the 60° input of subject 122, a mole on the visible side has explicit observation support, whereas the unobserved side contains no corresponding evidence. FFHQ-UV fits a complete texture through a

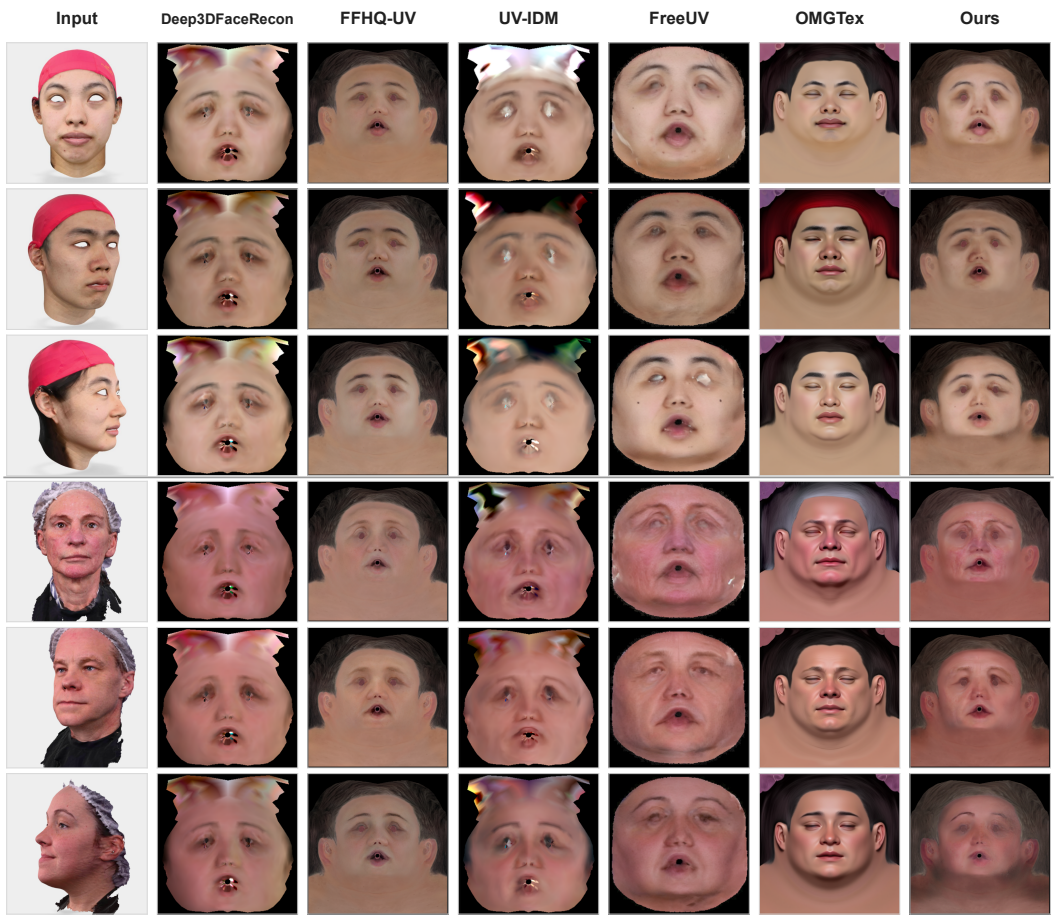


Figure 3: Qualitative comparison across methods. ForensicFace reconstructs complete UV textures while more consistently preserving local appearance directly supported by the input image.

GAN texture decoder in latent space, and its generative texture basis emphasizes global appearance and identity consistency. Small local input details can therefore be weakened during complete texture fitting (Bai et al., 2023). FreeUV can capture fine appearance from the input effectively, but its objective is to generate complete texture from incomplete UV conditions and does not explicitly constrain the observation origin of local features. A mole extracted from the visible side may consequently be synthesized again by the prior on the unobserved side (Yang et al., 2025). In contrast, ForensicFace preserves the correspondence between each local observation and its source image location and uses observation confidence to limit modification of reliable regions. It therefore retains the genuinely observed mole without extending that feature to the side where no supporting evidence exists.

4.3 Ablation Analysis

Core components. To evaluate the contribution of each component, we test the *Full* model and three variants on the same 359 FaceScape identities at 0°, 30°, and 60°, yielding 1,077 inputs. All variants use the same base prior, StyleGAN basis, random seed, and sampling budget. The results are reported in Table 2. The *Full* model achieves 20.874 dB PSNR, 0.710 SSIM, and 0.182 LPIPS. Compared with *Full*, removing (confidence, reliable extraction, UV DDPM) decreases PSNR by (0.45%, 0.35%, 0.59%), decreases SSIM by (1.13%, 1.69%, 0.85%), and increases LPIPS by (0.55%, 0.55%, 2.20%), respectively. Hence, removing UV DDPM causes the largest degradation in PSNR and LPIPS, while removing reliable extraction has the largest impact on SSIM. Figure 5 further shows that binary confidence creates sharper observed and generated transitions, while removing UV DDPM weakens global consistency and local transition quality. These results indicate that reliable extraction, continuous confidence control, and diffusion based coordination make complementary contributions.

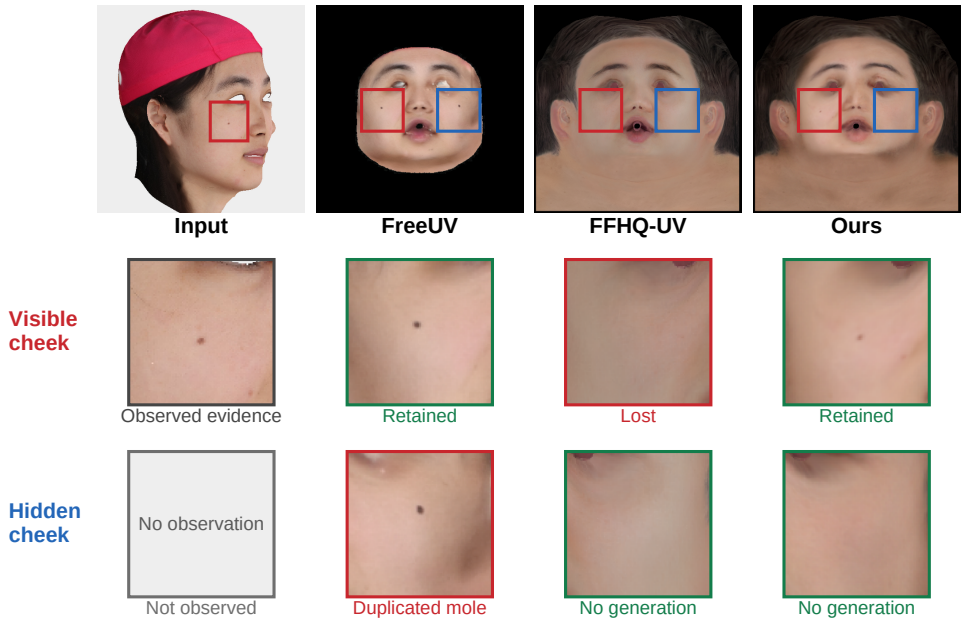


Figure 4: Preservation of local observations. A mole supported by genuine evidence on the visible side should be preserved, whereas the unobserved side should not reproduce that feature solely because of a symmetry prior.

Table 2: Ablation on 359 FaceScape identities rendered at $0°$, $30°$, and $60°$, under the same trained prior and inference budget.

Variant	PSNR ↑	SSIM ↑	LPIPS ↓
Full	**20.874**	**0.710**	**0.182**
w/o Confidence	20.783	0.702	0.183
w/o Reliable Extraction	20.799	0.698	0.183
w/o UV DDPM	20.749	0.704	0.186

4.4 Robustness Analysis

To evaluate the impact of pose variation, we vary the input yaw angle from $0°$ to $90°$ in $15°$ increments on 359 FaceScape identities, yielding 2,513 test samples under identical model weights and inference settings. Each metric is normalized to its mean at $0°$, which is set to 100%, and the results are reported in Figure 6. When varying the yaw angle over $(0°, 15°, 30°, 45°, 60°, 75°, 90°)$, the PSNR retentions are $(100.0\%, 99.8\%, 99.6\%, 99.1\%, 98.8\%, 99.1\%, 98.3\%)$, the SSIM retentions are $(100.0\%, 99.9\%, 99.9\%, 99.5\%, 99.3\%, 98.6\%, 99.6\%)$, the LPIPS retentions are $(100.0\%, 99.8\%, 99.9\%, 99.7\%, 99.6\%, 98.5\%, 96.6\%)$, and the CSIM retentions are $(100.0\%, 96.0\%, 87.0\%, 66.1\%, 57.7\%, 46.5\%, 40.2\%)$, respectively.

When varying the yaw angle, PSNR, SSIM, and LPIPS remain consistently stable, with average retentions of 99.1%, 99.5%, and 99.0% over the six non-frontal poses. Even at $90°$, they retain 98.3%, 99.6%, and 96.6%, respectively. In contrast, CSIM decreases from 96.0% at $15°$ to 40.2% at $90°$, with the largest adjacent reduction of 20.9 percentage points occurring from $30°$ to $45°$. The separation is already clear at $45°$, where PSNR, SSIM, and LPIPS retain 99.1%, 99.5%, and 99.7%, while CSIM retains 66.1%; at $75°$, the corresponding retentions are 99.1%, 98.6%, 98.5%, and 46.5%. Across all poses, the maximum deviations of PSNR, SSIM, and LPIPS from the frontal reference are only 1.7, 1.4, and 3.4 percentage points, respectively. These results indicate that increasing pose mainly reduces directly recoverable identity evidence rather than overall texture quality. This behavior is consistent with the design of ForensicFace: reliable observations remain constrained by the input, whereas increasingly unobserved regions are completed by the generative prior without being treated as direct identity evidence.



Figure 5: Qualitative ablation of the main components. Removing reliable observation extraction introduces erroneous projections, replacing continuous ρ with binary confidence creates hard boundaries, and removing UV DDPM weakens global coordination between observation evidence and the complete texture prior.

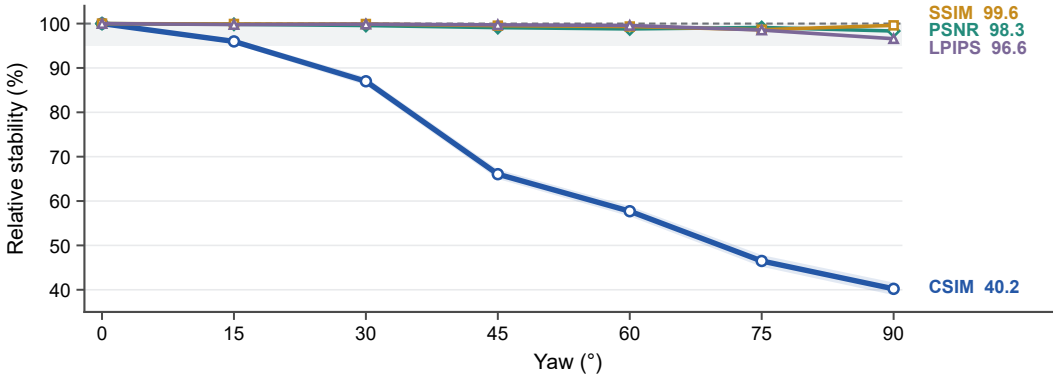


Figure 6: Pose robustness on FaceScape. Each metric is normalized to 100 percent at $0°$, and the curves show relative stability retention. PSNR, SSIM, and LPIPS remain stable under large pose variation, whereas CSIM decreases as observable identity information is reduced.

5 Conclusion

ForensicFace addresses the need for reliability and traceability in single image 3D face texture reconstruction for criminal investigations by explicitly separating input observations from prior completion. Canonical UV space provides the common texture representation for this process, while existing geometry and correspondence estimators are used only to establish the mapping from image evidence to surface texels. The reconstruction target therefore remains complete 3D face texture, with the proposed components focusing on the reliability, preservation, and provenance of texture evidence rather than geometry optimization. Texel level observation confidence estimates the reliability of input evidence and adaptively controls observation preservation during diffusion, while generation confidence from cross seed consistency characterizes regions without direct observation. Results on FaceScape and Stirling ESRC show favorable overall performance across pixel, structural, perceptual, and identity consistency metrics. Even at $90°$ yaw, PSNR, SSIM, and LPIPS retain high stability, whereas CSIM decreases gradually as observable identity information is reduced, showing stable texture completion without treating prior generated content as genuine identity observation. Together with source pixel mapping, observation confidence and generation confidence provide clearer information provenance for subsequent manual verification and reliable 3D face analysis.

AI USE STATEMENT

AI tools were used for language refinement, structural editing, and LaTeX formatting assistance. All technical content, experimental design, numerical results, citations, and final statements were verified by the authors, who retain full responsibility for the manuscript.

ETHICS STATEMENT

This work involves facial data and identity sensitive applications. Experiments use only datasets under their corresponding licenses and applicable dataset agreements. All facial images and reconstructed face examples displayed in this paper are selected only from samples whose dataset terms permit public display and dissemination in research publications.

REFERENCES

R. Abdal, Y. Qin, and P. Wonka. Image2StyleGAN: How to embed images into the StyleGAN latent space? In *ICCV*, 2019.

H. Bai, D. Kang, H. Zhang, J. Pan, and L. Bao. FFHQ-UV: Normalized facial UV-texture dataset for 3D face reconstruction. In *CVPR*, 2023.

A. J. Bose et al. Uncertainty-aware diffusion-guided refinement of 3D scenes. In *ICCV*, 2025.

M. De Vita and V. Belagiannis. Diffusion model guided sampling with pixel-wise aleatoric uncertainty estimation. In *WACV*, 2025.

J. Deng, J. Guo, N. Xue, and S. Zafeiriou. ArcFace: Additive angular margin loss for deep face recognition. In *CVPR*, 2019a.

Y. Deng, J. Yang, S. Xu, D. Chen, Y. Jia, and X. Tong. Accurate 3D face reconstruction with weakly-supervised learning: From single image to image set. In *CVPRW*, 2019b.

H.-P. Do, Y.-W. Chen, Y.-C. Liao, C.-W. Hsiao, H.-Y. Wang, W.-C. Chiu, and C.-C. Huang. DynFaceRestore: Balancing fidelity and quality in diffusion-guided blind face restoration with dynamic blur-level mapping and guidance. In *ICCV*, 2025.

B. Fei, Z. Lyu, L. Pan, J. Zhang, W. Yang, T. Luo, B. Zhang, and B. Dai. Generative diffusion prior for unified image restoration and enhancement. In *CVPR*, 2023.

G. Franchi, N. Belkhir, D. N. Trong, G. Xia, and A. Pilzer. Towards understanding and quantifying uncertainty for text-to-image generation. In *CVPR*, 2025.

B. Gecer, S. Ploumpis, I. Kotsia, and S. Zafeiriou. GANFIT: Generative adversarial network fitting for high fidelity 3D face reconstruction. In *CVPR*, 2019.

B. Gecer, J. Deng, and S. Zafeiriou. OSTeC: One-shot texture completion. In *CVPR*, 2021.

P. J. B. Hancock and B. Tiddeman. Stirling ESRC 3D Face Database. Psychological Image Collection at Stirling (PICS), University of Stirling, 2013.

T. Hempel, A. A. Abdelrahman, and A. Al-Hamadi. Toward robust and unconstrained full range of rotation head pose estimation. IEEE TIP, 33:2377–2387, 2024.

J. Ho, A. Jain, and P. Abbeel. Denoising diffusion probabilistic models. In *NeurIPS*, 2020.

T. Karras et al. Analyzing and improving the image quality of StyleGAN. In *CVPR*, 2020.

B. Kawar, M. Elad, S. Ermon, and J. Song. Denoising diffusion restoration models. In *NeurIPS*, 2022.

A. Kendall and Y. Gal. What uncertainties do we need in bayesian deep learning for computer vision? In *NeurIPS*, 2017.

B. Lakshminarayanan, A. Pritzel, and C. Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. In *NeurIPS*, 2017.

H. Li et al. UV-IDM: Identity-conditioned latent diffusion model for face UV-texture generation. In *CVPR*, 2024.

T. Li et al. Learning a model of facial shape and expression from 4D scans. ACM TOG, 36(6): 194:1–194:17, 2017.

I. Loshchilov and F. Hutter. Decoupled weight decay regularization. In *ICLR*, 2019.

A. Lugmayr et al. RePaint: Inpainting using denoising diffusion probabilistic models. In *CVPR*, 2022.

T. Martyniuk, O. Kupyn, Y. Kurlyak, I. Krashenyi, J. Matas, and V. Sharmanska. DAD-3DHeads: A large-scale dense, accurate and diverse dataset for 3D head alignment from a single image. In *CVPR*, 2022.

C. Meng et al. SDEdit: Guided image synthesis and editing with stochastic differential equations. In *ICLR*, 2022.

Y. Qiu, Z. Xiao, Y. Zuo, Z. Ye, W. Chen, and X. Han. AvatarTex: High-fidelity facial texture reconstruction from single-image stylized avatars. In *3DV*, 2026.

D. Roich, R. Mokady, A. H. Bermano, and D. Cohen-Or. Pivotal tuning for latent-based editing of real images. ACM TOG, 41(6):143:1–143:13, 2022.

J. Song, C. Meng, and S. Ermon. Denoising diffusion implicit models. In *ICLR*, 2021.

Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error visibility to structural similarity. IEEE TIP, 13(4):600–612, 2004.

Z. Xiao, Y. Qiu, Z. Ye, and X. Han. OMGTex: One-stage multi-style facial texture reconstruction without geometry guidance. In *CVPR*, 2026.

H. Yang et al. FaceScape: A large-scale high quality 3D face dataset and detailed riggable 3D face prediction. In *CVPR*, 2020.

X. Yang, T. Taketomi, Y. Endo, and Y. Kanamori. FreeUV: Ground-truth-free realistic facial UV texture recovery via cross-assembly inference strategy. In *CVPR*, 2025.

R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018.

Y. Zheng et al. General facial representation learning in a visual-linguistic manner. In *CVPR*, 2022.

W. Zielonka, T. Bolkart, and J. Thies. Towards metrical reconstruction of human faces. In *ECCV*, 2022.