

Supplementary Material for IConFace: Identity-Structure Asymmetric Conditioning for Unified Reference-Aware Face Restoration

Axi Niu*, Jinyang Zhang*, Senyan Qing

Northwestern Polytechnical University
nax@nwpu.edu.cn zhangjinyang@mail.nwpu.edu.cn qingsenyan@nwpu.edu.cn
Homepage: <https://cosmicrealm.github.io/IConFace/>

Supplementary Overview

This supplement provides protocol details, implementation specifications, and diagnostic evidence supporting the main paper: (i) dataset, training, inference, comparison-method, and metric protocols; (ii) reference-GT gap statistics, target-state distortion/structure checks, and GT-AdaFace analysis; and (iii) extended reference-aware, no-reference, and ablation visual comparisons. The main paper reports the compact method and benchmark summary, while this supplement documents the implementation choices and additional quantitative/qualitative evidence behind those results.

Protocol and Implementation Details

Training data. IConFace is trained on FFHQ-Ref, a reference-aware extension of FFHQ. FFHQ contains 70,000 aligned 1024×1024 face images with broad variation in age, ethnicity, background, accessories, and image conditions. FFHQ-Ref groups FFHQ images by ArcFace-predicted identity and provides reference mappings for images that have same-identity partners. The released FFHQ-Ref reference graph contains 20,405 usable high-quality images; our training split is the trainval mapping, which contains 6,373 identity groups and 19,548 usable images with at least one same-identity reference. During training, the target image is degraded online with mixed blind degradations, while same-identity images from the mapping serve as optional protocol references.

Evaluation data. Reference-aware evaluation uses CelebA-Test-Ref, FFHQ-Ref Moderate, FFHQ-Ref Severe, and CelebHQRef100. CelebA-Test-Ref contains 2,533 test samples with paired LQ/GT images and same-identity references from CelebA-HQ/CelebAMask-HQ. FFHQ-Ref Moderate and FFHQ-Ref Severe each contain 857 FFHQ-Ref test targets with fixed same-identity references and two synthetic degradation levels. CelebHQRef100 is a compact diagnostic benchmark built from 100 high-quality identity groups. The identities are selected deterministically from the sorted source identity folders, the first sorted image is used as target/GT, up to three remaining same-identity images are used as protocol references, and the target is degraded with mixed BSRGAN and Real-ESRGAN style

degradations using downsampling scales from 4 to 10. The degradation seed is fixed as $42 + i$ for identity index i to make the generated LQ images deterministic. CelebHQRef100 is not used for training, checkpoint selection, or prompt tuning. We verified that no released CelebHQRef100 image path or protocol identity folder appears in the FFHQ-Ref trainval mapping; the split is used only as a compact diagnostic benchmark. No-reference evaluation uses CelebA-Test (3,000 paired synthetic samples), LFW (1,711), CelebChild (360), WebPhoto (407), and Wider-Test (970); LFW, CelebChild, WebPhoto, and Wider-Test are treated as real-world sets without paired ground truth.

Backbone, training, and inference. IConFace uses the FLUX.2-klein-base-4B restoration backbone and adds rank 16 LoRA adapters together with the identity and degraded-structure modules described in the main paper. The final optimization setting uses 512×512 crops, fixed learning rate $1e-5$, batch size 1, and gradient accumulation 4. Degradation strength is sampled from 0–16 with bucket probabilities 0.5, 0.3, and 0.2 for ranges 0–3, 4–8, and 9–16. All reported results use 12 sampling steps, guidance scale 4.0, base seed 42, and 512 resolution.

Reference sampling and conditioning. Training mixes reference availability so the same checkpoint can operate with or without references: 30% of samples use no reference, 30% use one, 20% use two, and 20% use three. When fewer than three references are available, we use the actual number without duplication. The global identity pathway uses AdaFace IR50 embeddings with norm-only multi-reference aggregation and temperature 1.0. The degraded structure pathway uses degraded strength 1.0 and low-rank degraded cross-attention rank 16. No-reference inference removes reference tokens and uses the degraded-image AdaFace fallback only as weak forward conditioning.

The implemented objective uses the same losses as the main paper with explicit scalar multipliers. Eq. (8) uses uniform flow-matching timestep weight, $w(\sigma) = 1$ (`loss_weighting=none`), and scalar multiplier $\alpha_{\text{fm}} = 0.75$. The AdaFace identity multiplier is $\lambda_{\text{id}} = 0.30$, the hard target stabilizer uses $\lambda_h = 0.25$, and the sigma floor is $\omega_{\text{min}} = 0.25$:

$$\mathcal{L} = \alpha_{\text{fm}} \mathcal{L}_{\text{fm}} + \lambda_{\text{id}} \omega(\sigma) [(1 - \lambda_h^*) \mathcal{L}_{\text{ref-id}} + \lambda_h^* \mathcal{L}_{\text{hard}}], \quad (1)$$

*These authors contributed equally.

where $\lambda_h^* = \lambda_h(1 - \cos(e_{\text{ref}}, e_{\text{gt}}))$ and $\omega(\sigma) = \max(1 - \sigma, \omega_{\text{min}})^2$. Legacy sigma-threshold launcher fields are ignored by the final identity loss; the only active sigma control is the floor $\omega_{\text{min}} = 0.25$ above.

Token packing and RoPE. All image and text conditions use the four-axis FLUX.2 rotary position layout (t, h, w, l) . Scene latents are packed from (B, C, H, W) to (B, HW, C) with ids $(0, h, w, 0)$, while text tokens use ids $(0, 0, 0, l)$. The degraded image condition is assigned a separate temporal group with $t = 2$. Reference images are packed as grouped visual tokens with $t = 10 + r$ for reference index r , which lets the transformer distinguish the generated scene, degraded observation, and each reference without changing the spatial (h, w) axes. For a token i with position id $p_i = (t_i, h_i, w_i, l_i)$, the query and key vectors are split over axes $a \in \{t, h, w, l\}$ with dimensions $d_a = [32, 32, 32, 32]$. For each two-channel pair m on axis a , the RoPE computation is

$$\omega_{a,m} = \theta^{-2m/d_a}, \quad \theta = 2000, \quad (2a)$$

$$R(\rho, \omega) = \begin{bmatrix} \cos(\rho\omega) & -\sin(\rho\omega) \\ \sin(\rho\omega) & \cos(\rho\omega) \end{bmatrix}, \quad (2b)$$

$$\tilde{q}_{i,a,m} = R(p_i^a, \omega_{a,m})q_{i,a,m}, \quad (2c)$$

$$\tilde{k}_{j,a,m} = R(p_j^a, \omega_{a,m})k_{j,a,m}, \quad (2d)$$

$$\text{Attn}_{ij} \propto \exp\left(\frac{\tilde{q}_i^\top \tilde{k}_j}{\sqrt{d}}\right). \quad (2e)$$

This construction preserves the ordinary spatial axes while using the temporal axis to mark scene, degraded, and reference token groups.

Comparison methods. Reference-aware baselines are DMDNet, ReF-LDM, RestorerID, InstantRestore, and FaceMe; they are evaluated with the same protocol references whenever supported. Blind baselines are GFP-GAN, VQFR, CodeFormer, RestoreFormer++, and DAEFR, evaluated in the empty-reference setting. DMDNet learns dual memory dictionaries; ReF-LDM uses latent diffusion with high-quality references; RestorerID injects single-reference identity through an ID-preserving adapter; InstantRestore uses shared-image attention for single-step personalized restoration; and FaceMe extracts identity prompts from references. For blind baselines, GFP-GAN uses a GAN facial prior, VQFR uses a vector-quantized dictionary, CodeFormer predicts codebook entries with a transformer, RestoreFormer++ uses reconstruction-oriented priors, and DAEFR couples LQ evidence with high-quality codebook priors. References for these methods are provided in the main paper.

Metrics. For each reference-aware sample, let \hat{I}_i be the restored image and R_i^1 be the first protocol reference. Ref-ArcFace and Ref-AdaFace are dataset averages of cosine similarities

$$\text{RefMetric} = \frac{1}{N} \sum_{i=1}^N \frac{E(\hat{I}_i)^\top E(R_i^1)}{\|E(\hat{I}_i)\|_2 \|E(R_i^1)\|_2}, \quad (3)$$

where E is the corresponding frozen ArcFace or AdaFace encoder. We also report GT-AdaFace in the analysis section as the cosine similarity between \hat{I}_i and the paired target image G_i . PSNR, SSIM, and LPIPS are computed only on paired synthetic benchmarks and are treated as target-state distortion/structure checks rather than primary real-world quality metrics. MUSIQ, CLIP-IQA, and MANIQA are learned no-reference perceptual quality metrics; each restored image is scored independently and the dataset mean is reported. We use these metrics together so that reference-aligned identity consistency, target-state distortion, and perceptual quality can be read separately.

Table 1: Paired no-reference CelebA-Test distortion metrics. Main-paper no-reference results emphasize learned perceptual metrics; this table reports the standard paired distortion metrics for completeness.

Method	PSNR	SSIM	LPIPS
CodeFormer	25.146	0.685	0.227
GFP-GAN	25.105	0.696	0.241
VQFR	23.233	0.658	0.244
RestoreFormer++	25.313	0.685	0.226
DAEFR	22.591	0.628	0.250
IConFace	22.291	0.635	0.288

Reproducibility details. For reproducibility, we keep the restoration prompt, random seed, sampling steps, guidance scale, dataset split files, metric scripts, and checkpoint identifiers fixed across all reported evaluations. The released code package is organized around the inference and evaluation paths used for the main-paper tables and qualitative panels.

Reference-Aligned Identity Metric Analysis

Reference-aware restoration has two identity-related observations: the degraded target image, whose clean counterpart defines the target state, and one or more same-identity references, which provide external identity evidence. These observations are not spatially aligned and may differ in pose, expression, illumination, age, makeup, occlusion, or local facial state. Therefore, a GT-aligned identity score and a reference-aligned identity score do not always measure the same property. GT-aligned identity measures resemblance to the target-state image, while Ref-ArcFace and Ref-AdaFace measure consistency with the fixed reference evidence supplied by the protocol.

This distinction is important for interpreting reference-aware results. If the reference and GT are nearly identical, GT-based and reference-based identity scores tend to agree. When the reference and GT are only moderately aligned, however, a GT-only score can penalize a restoration that correctly uses reference identity evidence but does not reproduce the exact target-state appearance. We therefore treat Ref-AdaFace as the primary reference-aware identity metric, report perceptual quality metrics alongside it, and use qualitative comparisons to verify that the restored face still follows the degraded-image structure rather than copying the reference pose or expression.

Table 2: Identity similarity between the first protocol reference and the GT target over the full reference-aware evaluation splits. The last three columns report the percentage of samples whose Ref₁-GT AdaFace similarity falls below each threshold.

Dataset	N	Ref ₁ -GT ArcFace	Ref ₁ -GT AdaFace	AdaFace < 0.5	AdaFace < 0.6	AdaFace < 0.7
CelebA-Test-Ref	2533	0.616±0.093	0.605±0.104	15.52%	49.11%	80.77%
FFHQ-Ref Moderate	857	0.687±0.084	0.663±0.099	4.55%	23.45%	63.36%
FFHQ-Ref Severe	857	0.687±0.084	0.663±0.099	4.55%	23.45%	63.36%
CelebHQRef100	100	0.653±0.107	0.640±0.113	16.00%	34.00%	73.00%

Table 3: Paired target-state checks on the reference-aware benchmarks. SSIM is the direct structure check; PSNR and LPIPS give distortion context.

Dataset	Method	PSNR	SSIM	LPIPS
CelebA-Test-Ref	DMDNet	25.535	0.696	0.253
	ReF-LDM	23.901	0.638	0.268
	RestorerID	24.744	0.662	0.279
	InstantRestore	25.400	0.702	0.219
	FaceMe	25.947	0.698	0.253
	IConFace	22.327	0.632	0.277
FFHQ-Ref Moderate	DMDNet	25.844	0.726	0.234
	ReF-LDM	23.971	0.664	0.232
	RestorerID	24.924	0.691	0.240
	InstantRestore	25.452	0.727	0.213
	FaceMe	26.335	0.733	0.173
	IConFace	22.755	0.671	0.218
FFHQ-Ref Severe	DMDNet	20.111	0.570	0.449
	ReF-LDM	19.563	0.566	0.347
	RestorerID	19.412	0.559	0.415
	InstantRestore	21.372	0.647	0.323
	FaceMe	20.606	0.616	0.338
	IConFace	18.376	0.570	0.357
CelebHQRef100	DMDNet	22.857	0.648	0.304
	ReF-LDM	21.701	0.615	0.289
	RestorerID	21.857	0.612	0.340
	InstantRestore	22.801	0.674	0.244
	FaceMe	22.598	0.659	0.307
	IConFace	20.479	0.627	0.283

Table 2 shows that reference-GT mismatch is common. The FFHQ-Ref rows are identical because the Moderate and Severe splits share the same GT/reference protocol and differ only in degraded inputs. CelebA-Test-Ref is clearest: 49.11% and 80.77% of Ref₁-GT pairs fall below AdaFace 0.6 and 0.7. This is not a protocol error; reference and GT images are same-identity portraits captured under different pose, expression, lighting, age, makeup, occlusion, or local facial states. A reference-aware result can therefore become more consistent with supplied identity evidence without exactly reproducing the paired target state. The diagnostic panels in Figures 1-4 verify the complementary requirement: IConFace improves identity cues such as facial geometry, eye spacing, nose/mouth shape, and local identity marks, while pose, expression, and layout remain anchored to the degraded target rather than copied from the reference.

Table 3 provides paired target-state checks. IConFace has the lowest PSNR on these benchmarks, consistent with the perception-distortion tradeoff of Blau and Michaeli (2018): a flow/diffusion-style restorer that recovers plausible identity-consistent detail is not optimized for pixel-wise regression to the paired GT. The tradeoff is metric-dependent.

Table 4: GT-AdaFace on the full reference-aware benchmarks. Scores are computed against paired GT targets rather than protocol references.

Method	CelebA	FFHQ-M	FFHQ-S	CHQ100
DMDNet	0.768	0.835	0.170	0.509
ReF-LDM	0.806	0.862	0.661	0.683
RestorerID	0.782	0.828	0.412	0.543
InstantRestore	0.797	0.841	0.551	0.664
FaceMe	0.826	0.894	0.546	0.582
IConFace	0.736	0.814	0.717	0.712

On FFHQ-Ref Severe, IConFace has lower PSNR but better LPIPS than DMDNet and RestorerID and close LPIPS to ReF-LDM. SSIM is competitive on FFHQ-Ref Moderate/Severe and CelebHQRef100, while the CelebA-Test-Ref gap reflects the high Ref₁-GT divergence rate on that benchmark (Table 2), where reference-guided restoration can legitimately deviate from the exact target-state pixel structure. Table 1 gives the analogous paired CelebA-Test blind metrics. Together with the main-paper MUSIQ/CLIP-IQA/MANIQA results, these checks show an identity-perception/distortion tradeoff rather than a PSNR-oriented optimization.

Table 4 makes the GT-based identity picture explicit. The reversal is the key observation: IConFace is not highest on GT-AdaFace in the easier CelebA-Test-Ref and FFHQ-Ref Moderate splits, but becomes strongest on FFHQ-Ref Severe and CelebHQRef100. This directly validates the asymmetric design intent. When the degraded input retains sufficient identity evidence, the reference anchor can introduce mild divergence from the specific target state if Ref₁ and GT differ; conservative baselines may therefore stay closer to the paired GT. When severe degradation destroys that evidence, the reference pathway provides the reliable identity signal, and both Ref-AdaFace and GT-AdaFace improve. Thus high Ref-AdaFace is meaningful only when read together with GT identity, structure checks, and visual evidence that the output keeps the degraded pose/expression rather than copying the reference.

Qualitative material. The following pages expand the qualitative evidence for each benchmark. Reference-aware figures report AdaFace similarity to the first protocol reference image, while no-reference pages omit identity scores because no same-identity reference is supplied. The selected reference-GT diagnostic cases are drawn from the 20 lowest IConFace GT-AdaFace cases in each split and kept when Ref₁ and GT show a visible facial-state gap with clear inter-method differences; the Ref₁ and GT columns expose the actual Ref₁-GT AdaFace values.

CelebA-Test-Ref selected reference-GT gap cases





































Ref ₁	LQ	DMD Net	ReF- LDM	RestorerID	Instant Restore	FaceMe	Ours	GT
								
GT 0.459 R1 1.000	GT 0.454 R1 0.213	GT 0.450 R1 0.228	GT 0.509 R1 0.447	GT 0.568 R1 0.291	GT 0.617 R1 0.215	GT 0.534 R1 0.225	GT 0.351 R1 0.590	GT 1.000 R1 0.459
								
GT 0.512 R1 1.000	GT 0.730 R1 0.360	GT 0.740 R1 0.314	GT 0.716 R1 0.482	GT 0.636 R1 0.473	GT 0.795 R1 0.400	GT 0.794 R1 0.503	GT 0.478 R1 0.651	GT 1.000 R1 0.512
								
GT 0.573 R1 1.000	GT 0.711 R1 0.476	GT 0.688 R1 0.469	GT 0.696 R1 0.640	GT 0.651 R1 0.438	GT 0.727 R1 0.498	GT 0.775 R1 0.555	GT 0.498 R1 0.813	GT 1.000 R1 0.573
								
GT 0.351 R1 1.000	GT 0.778 R1 0.247	GT 0.733 R1 0.271	GT 0.705 R1 0.295	GT 0.602 R1 0.299	GT 0.541 R1 0.251	GT 0.790 R1 0.359	GT 0.436 R1 0.559	GT 1.000 R1 0.351

Figure 1: CelebA-Test-Ref reference-GT gap cases. Scores are AdaFace to GT/R1; the final row is an added low Ref₁-GT case. IDs: 18646, 20257, 26060, 02060.

FFHQ-Ref-Moderate selected reference-GT gap cases





































Ref ₁	LQ	DMD Net	ReF- LDM	RestorerID	Instant Restore	FaceMe	Ours	GT
								
GT 0.509 R1 1.000	GT 0.834 R1 0.495	GT 0.865 R1 0.442	GT 0.739 R1 0.432	GT 0.818 R1 0.486	GT 0.780 R1 0.427	GT 0.843 R1 0.464	GT 0.592 R1 0.663	GT 1.000 R1 0.509
								
GT 0.469 R1 1.000	GT 0.548 R1 0.250	GT 0.530 R1 0.200	GT 0.683 R1 0.372	GT 0.607 R1 0.341	GT 0.759 R1 0.313	GT 0.644 R1 0.355	GT 0.607 R1 0.536	GT 1.000 R1 0.469
								
GT 0.479 R1 1.000	GT 0.942 R1 0.515	GT 0.946 R1 0.487	GT 0.873 R1 0.450	GT 0.917 R1 0.485	GT 0.885 R1 0.414	GT 0.942 R1 0.557	GT 0.642 R1 0.729	GT 1.000 R1 0.479
								
GT 0.296 R1 1.000	GT 0.613 R1 0.148	GT 0.678 R1 0.113	GT 0.637 R1 0.221	GT 0.640 R1 0.139	GT 0.634 R1 0.238	GT 0.765 R1 0.197	GT 0.658 R1 0.365	GT 1.000 R1 0.296

Figure 2: FFHQ-Ref-Moderate reference-GT gap cases. Scores are AdaFace to GT/R1; the final row is an added low Ref₁-GT case. IDs: 57244, 31223, 02293, 37245.

FFHQ-Ref-Severe selected reference-GT gap cases

Ref ₁	LQ	DMD Net	ReF- LDM	RestorerID	Instant Restore	FaceMe	Ours	GT
GT 0.369 R1 1.000	GT -0.003 R1 -0.039	GT 0.179 R1 0.126	GT 0.469 R1 0.321	GT 0.335 R1 0.412	GT 0.331 R1 0.195	GT 0.455 R1 0.313	GT 0.450 R1 0.468	GT 1.000 R1 0.369
GT 0.513 R1 1.000	GT 0.060 R1 0.113	GT 0.018 R1 0.099	GT 0.492 R1 0.379	GT 0.456 R1 0.368	GT 0.336 R1 0.222	GT 0.462 R1 0.359	GT 0.473 R1 0.705	GT 1.000 R1 0.513
GT 0.449 R1 1.000	GT 0.357 R1 0.160	GT 0.155 R1 0.093	GT 0.538 R1 0.415	GT 0.398 R1 0.217	GT 0.604 R1 0.456	GT 0.645 R1 0.424	GT 0.483 R1 0.724	GT 1.000 R1 0.449
GT 0.416 R1 1.000	GT 0.131 R1 0.054	GT 0.048 R1 0.028	GT 0.491 R1 0.457	GT 0.292 R1 0.344	GT 0.402 R1 0.315	GT 0.325 R1 0.231	GT 0.484 R1 0.584	GT 1.000 R1 0.416

Figure 3: FFHQ-Ref-Severe reference-GT gap cases. Scores are AdaFace to GT/R1; the final row is an added low Ref₁-GT case. IDs: 48700, 04716, 12051, 08081.

CelebHQRef100 selected reference-GT gap cases

Ref ₁	LQ	DMD Net	ReF- LDM	RestorerID	Instant Restore	FaceMe	Ours	GT
GT 0.395 R1 1.000	GT -0.006 R1 -0.040	GT 0.032 R1 -0.001	GT 0.335 R1 0.348	GT -0.008 R1 -0.003	GT 0.358 R1 0.279	GT 0.201 R1 0.174	GT 0.406 R1 0.642	GT 1.000 R1 0.395
GT 0.554 R1 1.000	GT 0.380 R1 0.212	GT 0.461 R1 0.352	GT 0.574 R1 0.586	GT 0.547 R1 0.426	GT 0.575 R1 0.519	GT 0.571 R1 0.482	GT 0.489 R1 0.618	GT 1.000 R1 0.554
GT 0.683 R1 1.000	GT 0.027 R1 0.024	GT 0.489 R1 0.535	GT 0.458 R1 0.293	GT 0.151 R1 0.108	GT 0.349 R1 0.286	GT 0.180 R1 0.120	GT 0.515 R1 0.680	GT 1.000 R1 0.683
GT 0.414 R1 1.000	GT -0.051 R1 -0.015	GT 0.082 R1 -0.029	GT 0.516 R1 0.540	GT 0.193 R1 0.144	GT 0.482 R1 0.267	GT 0.291 R1 0.214	GT 0.502 R1 0.683	GT 1.000 R1 0.414

Figure 4: CelebHQRef100 reference-GT gap cases. Scores are AdaFace to GT/R1; the final row is an added low Ref₁-GT case. IDs: 00051_0, 00074_0, 00090_0, 00072_0.

CelebA-Test-Ref

Ref	LQ	DMDNet	ReF-LDM	RestorerID	InstantRestore	FaceMe	Ours	GT
CelebA-Test-Ref: case 18646								
1.000	0.213	0.228	0.447	0.291	0.215	0.225	0.590	0.459
CelebA-Test-Ref: case 24093								
1.000	0.293	0.306	0.338	0.255	0.307	0.309	0.422	0.380
CelebA-Test-Ref: case 17353								
1.000	0.195	0.173	0.272	0.394	0.222	0.343	0.567	0.389
CelebA-Test-Ref: case 27315								
1.000	0.235	0.269	0.513	0.479	0.494	0.488	0.577	0.448
CelebA-Test-Ref: case 02060								
1.000	0.247	0.271	0.295	0.299	0.251	0.359	0.559	0.351
CelebA-Test-Ref: case 18375								
1.000	0.273	0.319	0.395	0.461	0.514	0.526	0.569	0.465
CelebA-Test-Ref: case 19681								
1.000	0.212	0.126	0.323	0.248	0.182	0.257	0.535	0.388
CelebA-Test-Ref: case 08806								
1.000	0.296	0.232	0.413	0.430	0.316	0.401	0.506	0.515
CelebA-Test-Ref: case 20257								
1.000	0.360	0.314	0.482	0.473	0.400	0.503	0.651	0.512
CelebA-Test-Ref: case 00218								
1.000	0.323	0.346	0.361	0.419	0.366	0.386	0.430	0.469

Figure 5: Additional reference-aware qualitative comparisons on CelebA-Test-Ref. Scores under images report AdaFace similarity to the first protocol reference.

FFHQ-Ref-Moderate

















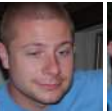
























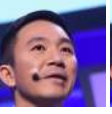





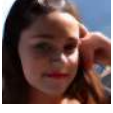

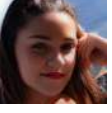
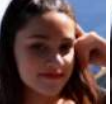
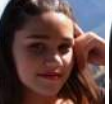
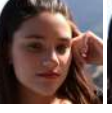






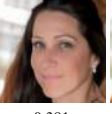
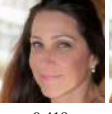

























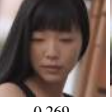



Ref	LQ	DMDNet	ReF-LDM	RestorerID	InstantRestore	FaceMe	Ours	GT
FFHQ-Ref-Moderate: case 57244								
								
1.000	0.495	0.442	0.432	0.486	0.427	0.464	0.663	0.509
FFHQ-Ref-Moderate: case 44570								
								
1.000	0.374	0.224	0.356	0.366	0.478	0.470	0.593	0.533
FFHQ-Ref-Moderate: case 55156								
								
1.000	0.315	0.322	0.436	0.363	0.356	0.447	0.671	0.541
FFHQ-Ref-Moderate: case 48700								
								
1.000	0.262	0.242	0.400	0.318	0.306	0.360	0.439	0.369
FFHQ-Ref-Moderate: case 54575								
								
1.000	0.521	0.523	0.587	0.455	0.536	0.586	0.628	0.610
FFHQ-Ref-Moderate: case 31223								
								
1.000	0.250	0.200	0.372	0.341	0.313	0.355	0.536	0.469
FFHQ-Ref-Moderate: case 08081								
								
1.000	0.386	0.429	0.471	0.475	0.381	0.418	0.530	0.416
FFHQ-Ref-Moderate: case 27262								
								
1.000	0.346	0.177	0.507	0.461	0.479	0.530	0.595	0.519
FFHQ-Ref-Moderate: case 15716								
								
1.000	0.376	0.280	0.536	0.381	0.428	0.488	0.590	0.488
FFHQ-Ref-Moderate: case 02647								
								
1.000	0.300	0.292	0.362	0.285	0.269	0.456	0.499	0.400

Figure 6: Additional reference-aware qualitative comparisons on FFHQ-Ref Moderate. Scores under images report AdaFace similarity to the first protocol reference.

FFHQ-Ref-Severe

Ref	LQ	DMDNet	ReF-LDM	RestorerID	InstantRestore	FaceMe	Ours	GT
FFHQ-Ref-Severe: case 02647								
1.000	0.206	0.205	0.491	0.323	0.407	0.373	0.520	0.400
FFHQ-Ref-Severe: case 17083								
1.000	0.309	0.153	0.410	0.381	0.409	0.407	0.515	0.513
FFHQ-Ref-Severe: case 04716								
1.000	0.113	0.099	0.379	0.368	0.222	0.359	0.705	0.513
FFHQ-Ref-Severe: case 12051								
1.000	0.160	0.093	0.415	0.217	0.456	0.424	0.724	0.449
FFHQ-Ref-Severe: case 48700								
1.000	-0.039	0.126	0.321	0.412	0.195	0.313	0.468	0.369
FFHQ-Ref-Severe: case 14836								
1.000	0.050	0.094	0.555	0.239	0.493	0.326	0.739	0.596
FFHQ-Ref-Severe: case 27211								
1.000	0.082	0.136	0.538	0.465	0.353	0.491	0.464	0.416
FFHQ-Ref-Severe: case 01829								
1.000	0.028	0.042	0.637	0.441	0.485	0.359	0.653	0.518
FFHQ-Ref-Severe: case 02293								
1.000	0.061	0.160	0.556	0.487	0.324	0.492	0.843	0.479
FFHQ-Ref-Severe: case 69793								
1.000	0.028	0.104	0.508	0.289	0.316	0.253	0.724	0.521

Figure 7: Additional reference-aware qualitative comparisons on FFHQ-Ref Severe. Scores under images report AdaFace similarity to the first protocol reference.

CelebHQRef100


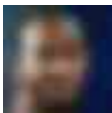





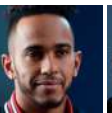




















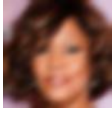








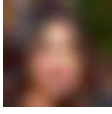



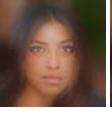
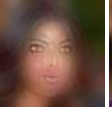
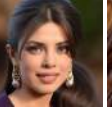


























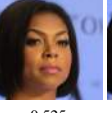


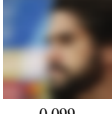



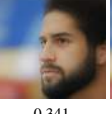

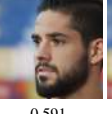


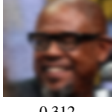







Ref	LQ	DMDNet	ReF-LDM	RestorerID	InstantRestore	FaceMe	Ours	GT
CelebHQRef100: case 00026_1								
								
1.000	0.022	0.134	0.235	0.304	0.214	0.244	0.492	0.426
CelebHQRef100: case 00098_0								
								
1.000	0.161	0.211	0.156	0.357	0.300	0.267	0.558	0.434
CelebHQRef100: case 00074_0								
								
1.000	0.212	0.352	0.586	0.426	0.519	0.482	0.618	0.554
CelebHQRef100: case 00046_1								
								
1.000	0.147	0.098	0.509	0.584	0.518	0.341	0.672	0.453
CelebHQRef100: case 00090_0								
								
1.000	0.024	0.535	0.293	0.108	0.286	0.120	0.680	0.683
CelebHQRef100: case 00067_0								
								
1.000	0.043	0.079	0.261	0.305	0.329	0.251	0.547	0.593
CelebHQRef100: case 00032_0								
								
1.000	0.007	0.137	0.447	0.428	0.426	0.216	0.620	0.488
CelebHQRef100: case 00052_0								
								
1.000	0.369	0.357	0.423	0.482	0.496	0.406	0.525	0.483
CelebHQRef100: case 00006_1								
								
1.000	0.099	0.171	0.400	0.247	0.341	0.169	0.591	0.582
CelebHQRef100: case 00049_1								
								
1.000	0.312	0.300	0.359	0.273	0.368	0.396	0.492	0.440

Figure 8: Additional reference-aware qualitative comparisons on CelebHQRef100. Scores under images report AdaFace similarity to the first protocol reference.

CelebA-Test

LQ CodeFormer GFP-GAN VQFR RF++ DAEFR Ours

CelebA-Test: case 0000216



CelebA-Test: case 00002335



CelebA-Test: case 00000893



CelebA-Test: case 00000894



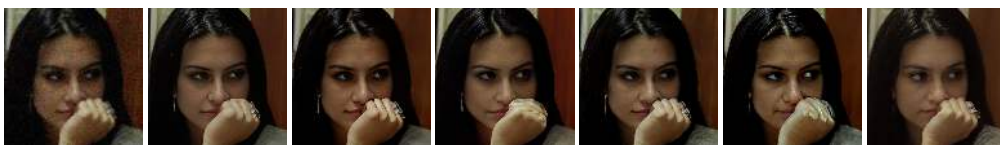
CelebA-Test: case 00001737



CelebA-Test: case 00000032



CelebA-Test: case 00000060



CelebA-Test: case 00002531



CelebA-Test: case 00002191

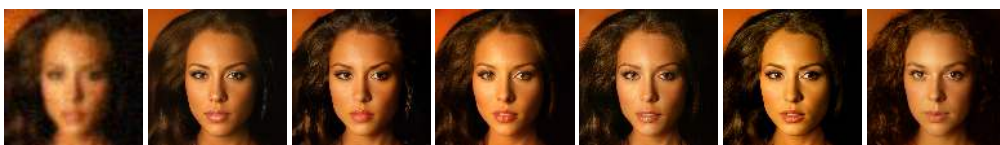


Figure 9: Additional no-reference qualitative comparisons on CelebA-Test.



Figure 10: Additional no-reference qualitative comparisons on LFW.

CelebChild



Figure 11: Additional no-reference qualitative comparisons on CelebChild.



Figure 12: Additional no-reference qualitative comparisons on WebPhoto.

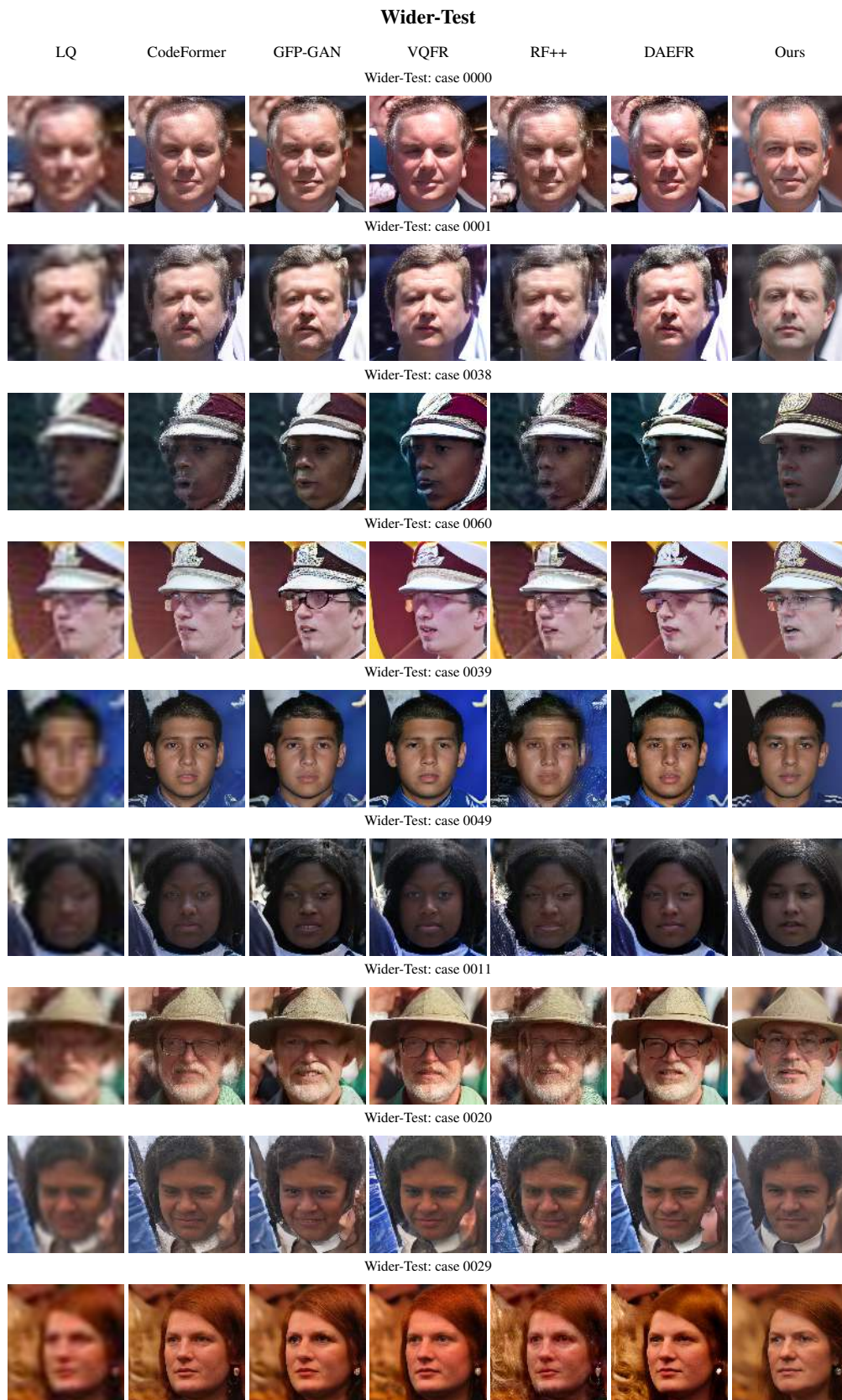


Figure 13: Additional no-reference qualitative comparisons on Wider-Test.

Reference-aware ablation (CelebHQRef100)

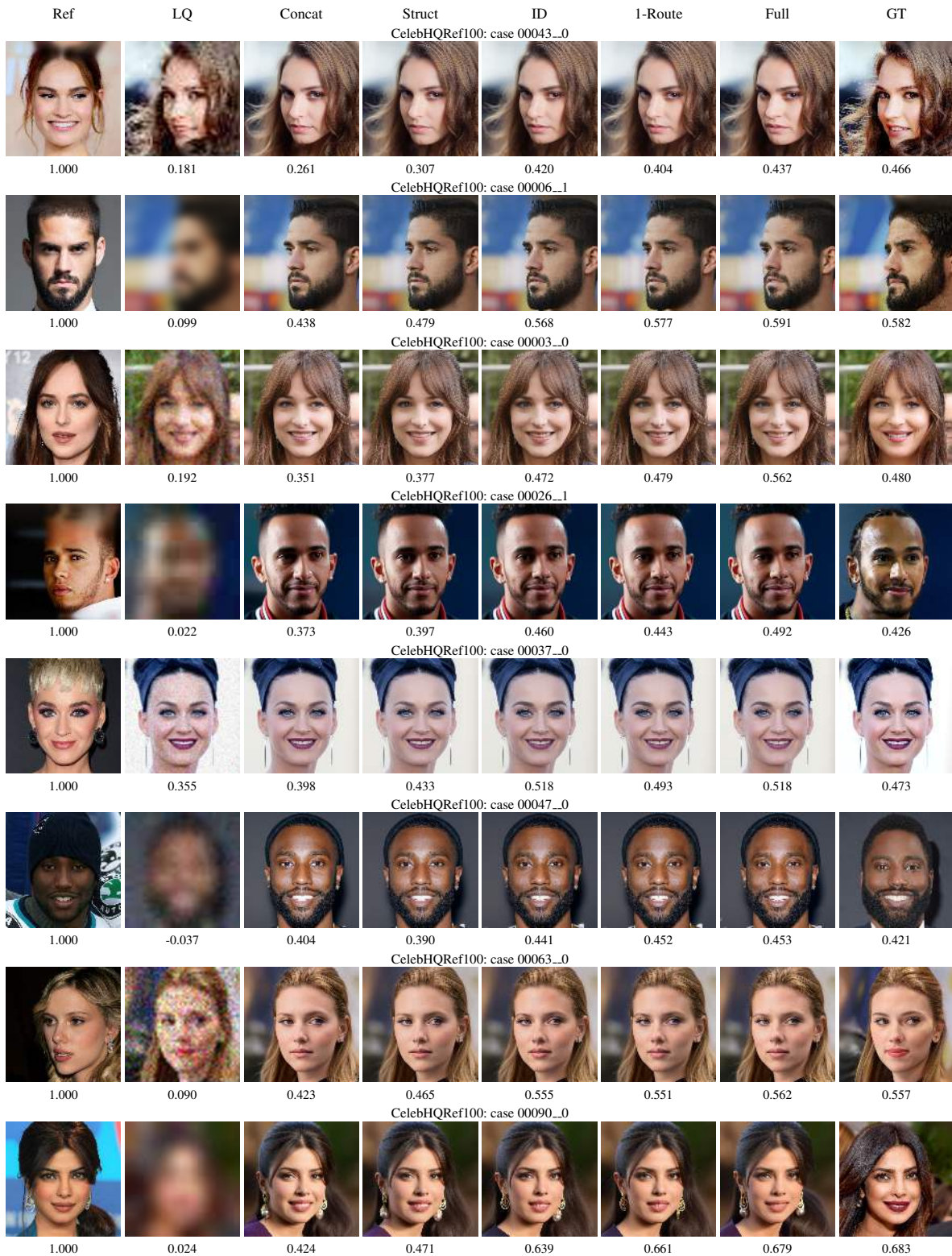


Figure 14: Reference-aware qualitative ablation on CelebHQRef100. Scores under images report AdaFace similarity to the first protocol reference.

Reference-aware ablation (FFHQ-Ref-Moderate)

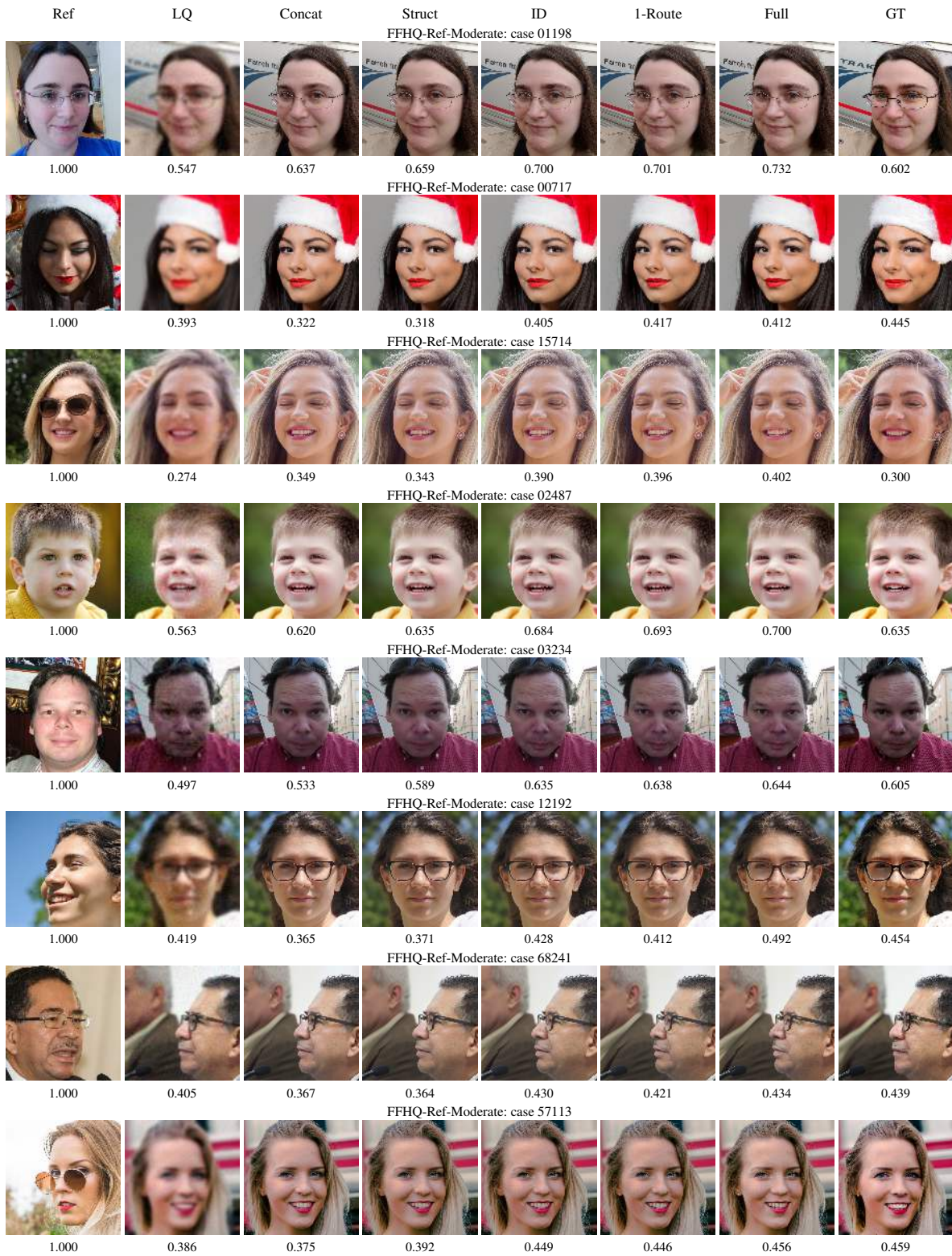


Figure 15: Reference-aware qualitative ablation on FFHQ-Ref Moderate. Scores under images report AdaFace similarity to the first protocol reference.

Reference-aware ablation (FFHQ-Ref-Severe)

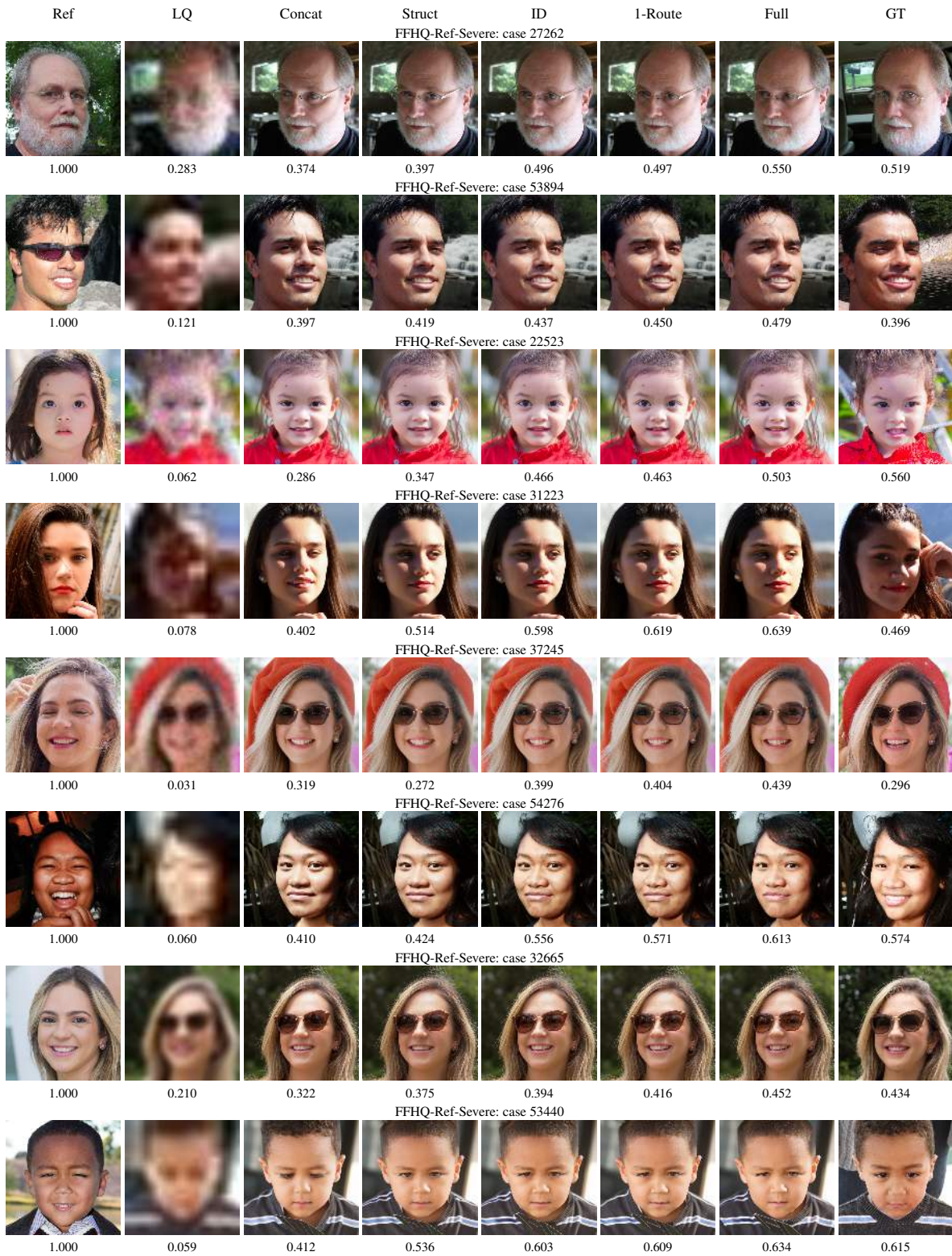


Figure 16: Reference-aware qualitative ablation on FFHQ-Ref Severe. Scores under images report AdaFace similarity to the first protocol reference.