DDFR: DEGRADATION-DISENTANGLED DIFFUSION-REGULARIZED FEATURE CORRECTION FOR BLURRED-QUERY FACE RETRIEVAL

Anonymous authors
Paper under double-blind review

ABSTRACT

Face retrieval across varying degrees of Gaussian blur is challenging because clean-gallery and blurred-query representations are inherently asymmetric: conventional systems assume sufficiently clear probe and gallery images, whereas real deployments often match blurred probes against clean stored identities. Gaussian blur suppresses discriminative cues and displaces the blurred-query embedding from its clean-gallery identity anchor, degrading cosine ranking. Existing recognizers compare degraded queries directly with clean-gallery embeddings, while blind face restoration (BFR) preprocessing pipelines optimize visual quality rather than retrieval alignment. We introduce DDFR, a degradation-aware and diffusion-regularized framework that corrects blur-induced displacement in feature space. A shared encoder produces identity and degradation representations; the latter estimates an identity-space correction vector and is regularized by diffusion-style latent denoising. Directional separation discourages identity-aligned correction, and the estimated vector is subtracted from the blurred-query representation before retrieval. We further construct a dedicated identity-disjoint Gaussian-blur face retrieval dataset covering $\sigma = 1, \ldots, 10$. On this benchmark, DDFR achieves $88.05 \pm 1.26\%$ average Rank-1 accuracy, while ArcFace and AdaFace trained on the same constructed dataset achieve 59.52% and 65.58%, respectively. At $\sigma = 10$, DDFR retains $60.76 \pm 3.74\%$ Rank-1 accuracy, compared with 21.37% and 28.06% for the same references. Component analyses support degradation modeling, latent denoising, and directional separation; the extended sweep identifies the limit imposed by severe information loss. These results suggest a path for lower-resolution camera inputs to approach the retrieval utility of clearer, higher-quality captures when identity cues remain recoverable.

1 INTRODUCTION

Face retrieval ranks a probe by cosine similarity to gallery identity representations. Conventional face retrieval assumes that probe and gallery images satisfy a minimum clarity requirement and therefore yield approximately symmetric representations; real deployments break this symmetry by matching a low-clarity, Gaussian-blurred probe against clean stored identities. In this asymmetric setting, the gallery is encoded from clean images and cached, whereas each query is encoded from a Gaussian-blurred observation. A recognizer may therefore fail even when identity cues remain, because the blurred-query embedding can move away from the corresponding clean identity region.

Existing approaches usually leave this asymmetry implicit. Direct recognition compares degraded-query and clean-gallery embeddings as if both came from the same domain. Blind face restoration (BFR) preprocessing sharpens the probe before recognition, but visual quality does not necessarily preserve the cosine-ranking structure of a fixed recognizer. The core issue is not blur alone, but blur-induced embedding misalignment.

This limitation is visible in distant crowded scenes, where a face occupies only a small fraction of the frame. Resizing a tighter crop to the recognizer's fixed input resolution enlarges both the target and the acquisition blur; it cannot recover identity evidence lost through under-sampling.

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: Distant-capture motivation for asymmetric blurred-query–clean-gallery retrieval. From left to right, progressively tighter crops centered on the lower-right subject are resized to the same display size. Stronger digital enlargement increases target occupancy but exposes the blur and under-sampling of the original observation; it cannot recreate identity evidence absent from the captured pixels. The yellow box marks the query subject. This practical example motivates DDFR's feature-space correction, while our controlled experiments isolate the degradation using Gaussian blur.

Figure 1 connects this practical failure mode to the controlled Gaussian-blur setting studied in this paper. The nested crops make the target progressively larger while preserving the limited evidence of the same distant observation. A blurred query therefore remains representationally asymmetric with a clean gallery even after both inputs are resized to the same network resolution. DDFR addresses this mismatch in feature space by estimating and subtracting blur-induced identity displacement rather than attempting to hallucinate missing pixels.

We model a blurred identity representation as a clean identity component plus a blur-induced displacement:

$$a_i^\sigma = a_i^0 + \Delta_i^\sigma + \epsilon_i^\sigma.$$

Here a_i^0 is the clean identity component, Δ_i^σ is the systematic identity-space displacement induced by Gaussian blur, and ϵ_i^σ collects residual nuisance. DDFR estimates a correction vector $\gamma_i^\sigma \approx \Delta_i^\sigma$ and forms $\mathrm{normalize}(a_i^\sigma - \gamma_i^\sigma)$ before clean-gallery retrieval.

DDFR implements this idea through a task-specific split between identity and degradation representations. The degradation representation estimates the correction term and is regularized by diffusion-style latent denoising. Directional separation discourages identity leakage into the subtracted vector, and inference produces one corrected embedding for retrieval against a fixed clean-gallery cache. We evaluate this design under a controlled identity-disjoint Gaussian-blur protocol with fixed preprocessing and no radius-specific test adaptation. DDFR achieves $88.05 \pm 1.26\%$ average Rank-1 accuracy over $\sigma = 1, \ldots, 10$; ablations, BFR preprocessing comparisons, and extended-blur experiments characterize its scope. The synthetic Gaussian queries are generated from public WIDER FACE/LFW source crops through a curated identity manifest; TinyFace only augments training, and splits are disjoint with respect to manifest identity keys Yang et al. (2016); Huang et al. (2007); Cheng et al. (2018).

Our contributions are threefold:

- We formulate blurred-query clean-gallery face retrieval as an asymmetric embedding-alignment problem and characterize Gaussian blur as a structured identity-space displacement.
- We introduce DDFR, a feature-correction framework that separates identity and degradation representations, regularizes degradation latents with diffusion-style denoising, and estimates a directionally constrained correction vector.
- We construct a dedicated identity-disjoint Gaussian-blur face retrieval dataset from public face-image sources, with clean galleries and paired queries spanning $\sigma = 1, \ldots, 10$. Controlled ablations, BFR comparisons, and extended-blur evaluations show gains over same-data ArcFace/AdaFace baselines and identify the method's limit under severe information loss.

2 Related Work

Margin-based and quality-aware face recognition.

FaceNet established metric embeddings for retrieval Schroff et al. (2015); SphereFace, CosFace, and ArcFace improved angular or cosine margins Liu et al. (2017); Wang et al. (2018); Deng et al. (2019); and MagFace/AdaFace account for sample quality Meng et al. (2021); Kim et al. (2022). These objectives produce strong identity representations, but they do not explicitly estimate a query-specific correction relative to a clean-gallery embedding space. DDFR keeps cosine retrieval and adds degradation-aware correction before scoring.

Low-quality and degraded face recognition.

Prior work on low-quality face recognition improves robustness through resolution matching, recognition-driven enhancement, quality estimation, and alignment-robust features Cheng et al. (2018); Shi et al. (2020); Chai et al. (2023); Saadabadi et al. (2024). Here, only the query branch is degraded at inference, and the gallery cache remains fixed. DDFR instead models blurred-query displacement directly in identity space.

Recognition after blind face restoration.

Blind face restoration (BFR) methods such as GFP-GAN, CodeFormer, DR2, and DiffBIR reconstruct plausible facial detail Wang et al. (2021); Zhou et al. (2022); Wang et al. (2023); Lin et al. (2024). Perceptual restoration does not necessarily preserve the embedding structure required by a fixed recognizer: a sharper probe can still produce a less aligned identity representation. We use BFR preprocessing pipelines as comparisons for the retrieval objective; DDFR corrects features and does not generate images.

Disentangled and diffusion-regularized representation learning.

Disentangled representation learning separates factors of variation for interpretability or control Higgins et al. (2017). DDFR uses a task-oriented separation: identity features are used for retrieval, while degradation features estimate the correction vector. Its diffusion component is limited to a training-time noise-prediction loss on degradation latents, not image generation or inference-time sampling. Directional separation further discourages the subtracted vector from following identity-discriminative directions.

3 Method

Problem Formulation

Let x_i^0 denote a clean image with identity label y_i, and let $x_i^\sigma = \mathcal{B}_\sigma(x_i^0)$ be the query generated by Gaussian blur operator $\mathcal{B}_\sigma$. The scalar σ is the evaluation blur radius. For losses or conditioning terms that require a scalar blur value, we use the normalized Gaussian-blur severity $\rho_i = \sigma_i/\sigma_{\max}$, with $\rho_i = 0$ for clean images. A clean-gallery retrieval set contains N cached embeddings $\{r_j^0\}_{j=1}^N$ in D dimensions. Throughout this paper, r denotes the final L2-normalized retrieval embedding after the DDFR correction path. Thus r_j^0 is computed from clean gallery image x_j^0, and r_i^σ is computed from blurred query x_i^σ. Retrieval ranks the gallery by

$$s_{ij}^\sigma = \langle r_i^\sigma, r_j^0 \rangle, \qquad \hat{y}_i^\sigma = \arg\max_j s_{ij}^\sigma. \tag{1}$$

The predicted identity is $\hat{y}_i^\sigma$. All retrieval scores use L2-normalized embeddings. In the reported DDFR evaluation, clean-gallery embeddings are computed once with the same DDFR feature-correction path and then cached; after caching, retrieval processes only the incoming query and does not recompute the gallery, adapt to a test blur radius, or change the index.

Unless stated otherwise, a superscript 0 denotes the clean/reference view and a superscript σ denotes the blurred-query view; paired training equations write the blurred view as σ_i.

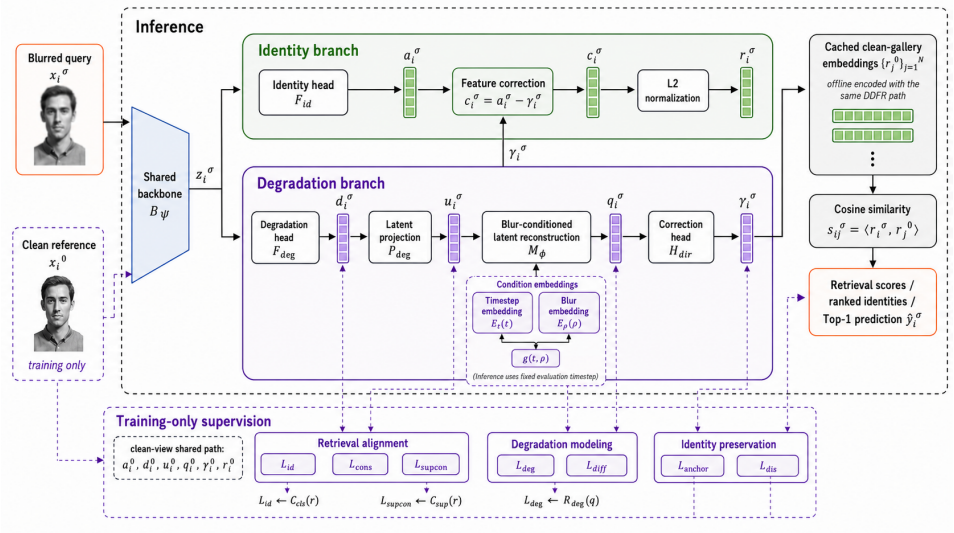

Figure 2: Overview of DDFR. The query x_i^σ is encoded into identity feature a_i^σ and degradation feature d_i^σ; subtracting γ_i^σ and L2-normalizing gives r_i^σ, which is compared with cached clean-gallery embeddings $\{r_j^0\}_{j=1}^N$. The degradation branch uses blur-conditioned latent reconstruction to estimate the correction vector, while the clean reference path and dashed arrows denote training-only supervision.

Blur-Induced Feature Displacement

The shared recognition backbone maps a clean or blurred image to a feature z_i^σ. The identity branch then produces a_i^σ, while a_i^0 denotes the corresponding clean identity representation. We view Gaussian blur as inducing a structured displacement in this identity space:

$$a_i^\sigma = a_i^0 + \Delta_i^\sigma + \epsilon_i^\sigma, \tag{2}$$

where Δ_i^σ is the systematic blur-induced shift and ϵ_i^σ contains residual nuisance. DDFR estimates a correction vector γ_i^σ from the degradation representation and forms

$$c_i^\sigma = a_i^\sigma - \gamma_i^\sigma, \qquad r_i^\sigma = \frac{c_i^\sigma}{\|c_i^\sigma\|_2}. \tag{3}$$

If γ_i^σ approximates the component of Δ_i^σ that harms clean-gallery ranking, the corrected embedding reduces the residual perturbation before cosine search. This motivates feature-space correction directly in the retrieval representation instead of pixel-space restoration.

Degradation-Disentangled Feature Correction

A shared backbone B_ψ and two lightweight branches produce identity and degradation views:

$$z_i^\sigma = B_\psi(x_i^\sigma), \quad a_i^\sigma = F_{\mathrm{id}}(z_i^\sigma), \quad d_i^\sigma = F_{\mathrm{deg}}(z_i^\sigma). \tag{4}$$

Here $z_i^\sigma \in \mathbb{R}^{512}$ is the shared visual feature, $a_i^\sigma \in \mathbb{R}^{512}$ is the identity feature to be corrected, and $d_i^\sigma \in \mathbb{R}^{512}$ is the degradation feature used to estimate the correction. The degradation branch is projected to a latent $u_i^\sigma \in \mathbb{R}^{256}$, processed by the blur-conditioned latent diffusion block $\mathcal{M}_\phi$, and mapped to the direct correction:

$$u_i^\sigma = P_{\mathrm{deg}}(d_i^\sigma), \quad q_i^\sigma = \mathcal{M}_\phi(u_i^\sigma, \rho_i), \quad \gamma_i^\sigma = H_{\mathrm{dir}}(q_i^\sigma). \tag{5}$$

Here P_{deg} is the degradation-latent projection, q_i^σ is the denoised degradation representation, and H_{dir} is the directional correction head that outputs $\gamma_i^\sigma \in \mathbb{R}^{512}$. The Gaussian-blur condition embedding used inside $\mathcal{M}_\phi$ is defined below. The reported main path uses the corrected embedding in Eq. 3. It does not use image restoration or radius-specific test-time tuning in the final retrieval path.

DIFFUSION-REGULARIZED DEGRADATION REPRESENTATION

The diffusion component regularizes the degradation latent; it is not an image generator. In the implementation, P_{deg} first produces the clean latent u_i^σ. During training, a timestep t is sampled uniformly from $\{0, \dots, T-1\}$ and Gaussian noise ϵ is added:

$$\tilde{u}_{i,t}^\sigma = \sqrt{\bar{\alpha}_t} u_i^\sigma + \sqrt{1-\bar{\alpha}_t}\epsilon, \qquad \epsilon \sim \mathcal{N}(0, I). \tag{6}$$

A denoiser predicts the injected noise using the noisy latent, a sinusoidal time embedding, and a learned Gaussian-blur severity embedding. Let $g(t, \rho_i) = E_t(t) + E_\rho(\rho_i)$ be this condition vector, where E_t is the timestep embedding and E_ρ maps the normalized blur severity to the condition embedding. The denoiser is

$$\hat{\epsilon}_{i,t}^\sigma = \epsilon_\phi(\tilde{u}_{i,t}^\sigma, g(t, \rho_i)), \qquad \mathcal{L}_{\text{diff}} = \mathbb{E}\|\epsilon - \hat{\epsilon}_{i,t}^\sigma\|_2^2. \tag{7}$$

The clean-latent estimate used by the correction head is

$$\hat{u}_{i,t}^\sigma = \frac{\tilde{u}_{i,t}^\sigma - \sqrt{1-\bar{\alpha}_t}\hat{\epsilon}_{i,t}^\sigma}{\sqrt{\bar{\alpha}_t}}, \qquad q_i^\sigma = O_\phi(\hat{u}_{i,t}^\sigma). \tag{8}$$

Here $\bar{\alpha}_t = \prod_{k=0}^{t}(1-\beta_k)$ is the cumulative schedule coefficient and O_ϕ is the output projection that maps the predicted clean latent back to the degradation-representation space. Therefore $\mathcal{M}_\phi$ in Eq. 5 consists of noise addition, Gaussian-blur-conditioned noise prediction, clean-latent reconstruction, and output projection. At inference, the same module is used once with the fixed evaluation timestep $t_{\text{eval}} = 99$ and zero injected noise under the $T = 100$ schedule. In the reported Gaussian-blur evaluation, all blur radii use the fixed severe-blur condition $\rho_\star$. Thus DDFR performs one deterministic latent reconstruction, not a reverse diffusion chain or stochastic sampler.

TRAINING OBJECTIVES

Training uses paired clean/blurred samples $(x_i^0, x_i^{\sigma_i}, y_i, \rho_i)$, where $\rho_i = \sigma_i/\sigma_{\max}$ for the blurred view and $\rho_i = 0$ for the clean view. Let $a_i^0, d_i^0, u_i^0, q_i^0, \gamma_i^0, r_i^0$ denote clean-view tensors and $a_i^{\sigma_i}, d_i^{\sigma_i}, u_i^{\sigma_i}, q_i^{\sigma_i}, \gamma_i^{\sigma_i}, r_i^{\sigma_i}$ denote blurred-view tensors defined by Eqs. 4–3. The classifier and contrastive heads receive $w = C_{\text{cls}}(r)$ and $v = C_{\text{sup}}(r)$ for each view.

The identity classification loss supervises the corrected embeddings:

$$\begin{aligned}\mathcal{L}_{\text{id}} = &\frac{1}{2}\mathbb{E}_i\big[-\log \text{softmax}(G(w_i^0))_{y_i} \\ &- \log \text{softmax}(G(w_i^{\sigma_i}))_{y_i}\big].\end{aligned} \tag{9}$$

The cosine consistency loss aligns clean and blurred corrected embeddings of the same identity:

$$\mathcal{L}_{\text{cons}} = \mathbb{E}_i\left[1 - \langle r_i^0, r_i^{\sigma_i}\rangle\right]. \tag{10}$$

Supervised contrastive learning is computed on the contrastive embeddings from the paired batch. For anchor index m, $\mathcal{P}(m)$ contains other mini-batch embeddings with the same identity label:

$$\begin{aligned}\mathcal{L}_{\text{supcon}} &= -\sum_m \frac{1}{|\mathcal{P}(m)|}\sum_{p\in\mathcal{P}(m)} \log\frac{\exp(\langle v_m, v_p\rangle/\tau)}{Z_m}, \\ Z_m &= \sum_{q\neq m}\exp(\langle v_m, v_q\rangle/\tau).\end{aligned} \tag{11}$$

The clean-anchor loss limits drift of the clean corrected embedding from the current clean identity feature:

$$\mathcal{L}_{\text{anchor}} = \mathbb{E}_i\left[1 - \left\langle \frac{a_i^0}{\|a_i^0\|_2}, r_i^0 \right\rangle\right]. \tag{12}$$

The reported configuration supervises the Gaussian-blur condition by regression from q_i^0 and $q_i^{\sigma_i}$:

$$\mathcal{L}_{\text{deg}} = \frac{1}{2}\mathbb{E}_i\left[R_{\text{deg}}(q_i^0)^2 + (R_{\text{deg}}(q_i^{\sigma_i}) - \rho_i)^2\right]. \tag{13}$$

Here R_{deg} is the scalar blur-condition regressor. The reported configuration does not use a degradation-classification objective.

Because the correction vector is subtracted from the identity feature, DDFR penalizes identity-aligned correction directions:

$$\mathcal{L}_{\text{dis}} = \frac{1}{2}\mathbb{E}_i \left[\left| \cos(a_i^0, \gamma_i^0) \right| + \left| \cos(a_i^{\sigma_i}, \gamma_i^{\sigma_i}) \right| \right]. \tag{14}$$

This term encourages directional separation; it should not be interpreted as a proof of statistical independence. In the implemented losses $\mathcal{L}_{\text{cons}}$, $\mathcal{L}_{\text{anchor}}$, and $\mathcal{L}_{\text{dis}}$ are evaluated without stop-gradient between the paired terms.

The full reported objective is

$$\begin{aligned} \mathcal{L} = & \lambda_{\text{id}}\mathcal{L}_{\text{id}} + \lambda_{\text{cons}}\mathcal{L}_{\text{cons}} + \lambda_{\text{supcon}}\mathcal{L}_{\text{supcon}} \\ & + \lambda_{\text{anchor}}\mathcal{L}_{\text{anchor}} + \lambda_{\text{deg}}\mathcal{L}_{\text{deg}} + \lambda_{\text{dis}}\mathcal{L}_{\text{dis}} + \lambda_{\text{diff}}\mathcal{L}_{\text{diff}}. \end{aligned} \tag{15}$$

Optional contrastive, counterfactual, teacher, prototype, queue, and adapter losses exist in the codebase but are disabled in the reported main configuration.

Inference and Complexity

The clean gallery is encoded offline and cached. At query time, DDFR evaluates one backbone pass, two branches, one blur-conditioned latent reconstruction, and one correction head. Exhaustive cosine retrieval costs $O(ND)$ for gallery size N and embedding dimension D, and the corrected embedding can be passed to the same nearest-neighbor backend used by a baseline recognizer. The final inference path contains no pixel-space restoration or radius-specific adaptation. The technical supplement gives the complete deterministic inference sequence.

4 Experiments

We evaluate Gaussian-blur retrieval, component ablations, BFR preprocessing comparisons, and extended-blur boundary cases under the same clean-gallery protocol unless stated otherwise.

Dedicated Gaussian-Blur Dataset and Protocol

We construct a dedicated Gaussian-blur face retrieval dataset from public WIDER FACE and LFW source crops through a curated identity manifest, with TinyFace used only for training augmentation. WIDER FACE is not treated as an official identity-labeled benchmark; its detections are included only after local preprocessing assigns stable manifest identity keys, which are treated as pseudo-identities when no public identity annotation is available. Exact image hashes are de-duplicated, identity keys are split before blur generation, and the protocol builder audits train/validation/test identity overlap. Thus "identity-disjoint" means disjoint manifest identity keys; unknown cross-dataset real-person duplicates under different keys remain a limitation. The gallery contains clean images from the held-out test split, and each paired query is generated with PIL GaussianBlur(radius=σ), for $\sigma \in \{1, \ldots, 10\}$. This range spans mild to severe blur at 112×112: even $\sigma = 1$ or 2 affects local facial contours, while $\sigma = 10$ strongly attenuates identity information. The technical supplement reports the complete benchmark settings. The exact test identities are not used during training, so evaluation measures generalization to unseen identities under a fixed Gaussian-blur range.

Rank-1 accuracy measures direct retrieval of the correct identity. Top-5 and Top-10 measure whether it remains in the candidate set after top-1 failure, which is informative under severe blur.

The test split contains 1715 manifest identities and 1750 gallery entries because 35 identities contribute a second clean view; each query is paired with one gallery image, and correctness is evaluated by the identity label of the top-ranked gallery image.

Figure 3 separates recognition-only trends from BFR preprocessing comparisons. Qualitative examples and extended analyses are provided in the Technical Supplement.

Baselines and Implementation

All methods are evaluated on the same gallery, queries, preprocessing resolution, and cosine-retrieval protocol. ArcFace and AdaFace Deng et al. (2019); Kim et al. (2022) are trained on the same

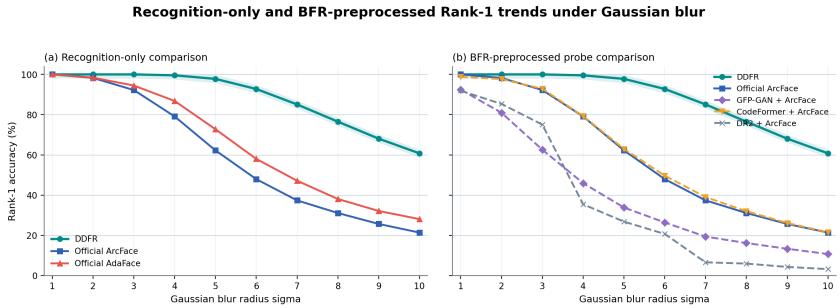

Figure 3: Rank-1 accuracy of DDFR and same-data trained ArcFace/AdaFace references across Gaussian-blur radii. Panel (a) compares recognition-only methods; panel (b) compares raw recognition, BFR preprocessing, and DDFR under the same clean-gallery protocol using stored retrieval logs.

Method	R1(σ=1)	R1(σ=5)	R1(σ=10)	Top-5(σ=10)	Avg. R1
ArcFace (same data)	100.00	62.23	21.37	27.09	59.52
AdaFace (same data)	100.00	72.74	28.06	33.66	65.58
GFP-GAN + ArcFace	92.46	33.83	10.69	14.63	40.17
CodeFormer + ArcFace	98.74	62.97	21.54	25.89	59.98
DR2 + ArcFace	91.89	26.80	3.20	8.97	35.51
Ours, DDFR	$\mathbf{100.00 \pm 0.00}$	$\mathbf{97.77 \pm 0.37}$	$\mathbf{60.76 \pm 3.74}$	$\mathbf{77.83 \pm 3.21}$	$\mathbf{88.05 \pm 1.26}$

Table 1: Main retrieval results under the fixed Gaussian-blur protocol. Columns report Rank-1 at selected radii, Top-5 at $\sigma = 10$, and average Rank-1 over $\sigma = 1, \ldots, 10$. DDFR reports three-seed mean and standard deviation. ArcFace, AdaFace, and DDFR are trained on the same constructed Gaussian-blur training split and evaluated under the same identity-disjoint clean-gallery retrieval protocol; BFR preprocessing comparisons use official restoration checkpoints.

constructed Gaussian-blur training split as DDFR, with the same identity-disjoint train/validation/test partitions and the same input resolution. This controls data exposure and evaluation protocol, so the comparison is fair in terms of training data and retrieval setting. BFR preprocessing comparisons first restore the blurred query with GFP-GAN, CodeFormer, or DR2 and then recognize it with ArcFace. DDFR evaluates the corrected feature directly. Component ablations further isolate the effect of the degradation branch, latent denoising, blur conditioning, directional separation, and correction path.

All inputs are resized to 112×112 pixels. DDFR uses a ResNet18 backbone He et al. (2016) and 512-dimensional identity/degradation representations. The latent diffusion branch uses 256 dimensions and a 100-step diffusion-style noise schedule during training. At inference, all blur radii use the fixed condition embedding $\rho_\star$. Complete optimization settings are reported in the technical supplement.

Main Results

Table 1 and Figure 3(a) report the main Gaussian-blur results. Across $\sigma = 1, \ldots, 10$, DDFR reaches 88.05 ± 1.26 average Rank-1 accuracy. ArcFace and AdaFace trained on the same constructed dataset reach 59.52 and 65.58, giving robustness gaps of $+28.53$ and $+22.47$ percentage points under this fair data-exposure protocol. At $\sigma = 10$, DDFR retains $60.76 \pm 3.74\%$ Rank-1 accuracy, while the same baselines reach 21.37% and 28.06%, respectively.

The recognition-only trend is consistent with explicit feature correction preserving alignment with clean-gallery embeddings under increasing blur. Figure 3(b) shows that BFR preprocessing does not consistently close the gap: GFP-GAN + ArcFace and DR2 + ArcFace can underperform raw recognition, while CodeFormer + ArcFace remains close to the same-data ArcFace baseline but below DDFR.

Component Ablation

Full DDFR reaches 88.05 ± 1.26 average Rank-1, versus 27.09 for the uncorrected identity representation. Removing the degradation branch, latent denoising, directional separation, or

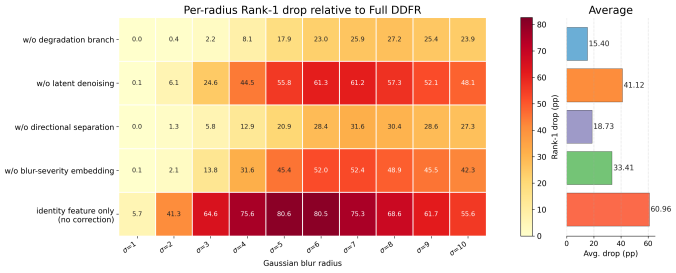

Figure 4: Component ablation analysis. The heatmap reports Rank-1 drop relative to Full DDFR by blur radius; the side bar reports the average drop.

blur-severity embedding reduces average Rank-1 to 72.65, 46.93, 69.32, and 54.64, respectively. These results support degradation-aware correction, diffusion-style regularization, identity-leakage control, and blur conditioning; Figure 4 shows that the losses grow under severe blur. Per-radius values are in the supplement.

Scope and Boundary Analyses

Figure 3(b) shows that BFR preprocessing yields inconsistent retrieval gains, whereas DDFR corrects features against the cached gallery directly. With the checkpoint and gallery fixed, DDFR remains above 50% Rank-1 at $\sigma = 11$ and reaches 49.05% at $\sigma = 12$, bounding the claim to recoverable Gaussian blur. Further BFR and extended-blur results are in the supplement.

5 Discussion

DDFR directly corrects identity-space displacement, whereas sharper BFR output need not preserve gallery ranking geometry. Its degradation context, latent denoising, and directional separation are complementary, but performance still declines once blur removes identity evidence. Pose, illumination, compression, sensor noise, alignment error, domain shift, and demographic robustness remain open. Because DDFR, ArcFace, and AdaFace use the same constructed training split, input resolution, gallery, and retrieval protocol, their comparison controls blur-data exposure; architecture and objective design remain the intended method-level differences.

6 Conclusion

We introduced DDFR for correcting blur-induced displacement before clean-gallery retrieval. It improves over same-data ArcFace/AdaFace and BFR preprocessing, with ablations supporting its components. When identity cues remain, this may help lower-resolution cameras approach the retrieval utility of clearer captures, subject to deployment safeguards. The accompanying identity-disjoint Gaussian-blur dataset provides a controlled benchmark for this asymmetric retrieval setting.

Ethics Statement and Limitations. Synthetic Gaussian-blur results may not generalize to pose, motion blur, compression, weather, misalignment, demographic variation, or domain shift. They do not validate unconstrained surveillance; deployment requires legal and privacy safeguards, robustness audits, calibrated uncertainty, and human review.

AI Use Statement. Generative AI assisted language, notation, figures, and code packaging. The authors reviewed the material, verified values against code and logs, tested inference, and take responsibility for the final content.

REFERENCES

Jacky Chen Long Chai, Tiong-Sik Ng, Cheng-Yaw Low, Jaewoo Park, and Andrew Beng Jin Teoh. Recognizability embedding enhancement for very low-resolution face recognition and quality estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9957–9967, 2023.

Zhiyi Cheng, Xiatian Zhu, and Shaogang Gong. Low-resolution face recognition. In *Proceedings of the Asian Conference on Computer Vision*, pp. 605–621, 2018.

Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4690–4699, 2019.

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

Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. beta-vae: Learning basic visual concepts with a constrained variational framework. In *International Conference on Learning Representations*, 2017.

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

Minchul Kim, Anil K. Jain, and Xiaoming Liu. Adaface: Quality adaptive margin for face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 18750–18759, 2022.

Xinqi Lin, Jingwen He, Ziyan Chen, Zhaoyang Lyu, Bo Dai, Fanghua Yu, Wanli Ouyang, Yu Qiao, and Chao Dong. DiffBIR: Towards blind image restoration with generative diffusion prior. In *Proceedings of the European Conference on Computer Vision*, 2024.

Weiyang Liu, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song. Sphereface: Deep hypersphere embedding for face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 212–220, 2017.

Qiang Meng, Shichao Zhao, Zhida Huang, and Feng Zhou. Magface: A universal representation for face recognition and quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14225–14234, 2021.

Mohammad Saeed Ebrahimi Saadabadi, Sahar Rahimi Malakshan, Ali Dabouei, and Nasser M. Nasrabadi. ARoFace: Alignment robustness to improve low-quality face recognition. In *Proceedings of the European Conference on Computer Vision*, 2024.

Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 815–823, 2015.

Yichun Shi, Xiang Yu, Kihyuk Sohn, Manmohan Chandraker, and Anil K. Jain. Towards universal representation learning for deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6817–6826, 2020.

Hao Wang, Yitong Wang, Zheng Zhou, Xing Ji, Dihong Gong, Jingchao Zhou, Zhifeng Li, and Wei Liu. Cosface: Large margin cosine loss for deep face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5265–5274, 2018.

Xintao Wang, Yu Li, Honglun Zhang, and Ying Shan. Towards real-world blind face restoration with generative facial prior. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9168–9178, 2021.

Zhixin Wang, Xiaoyun Zhang, Ziying Zhang, Huangjie Zheng, Mingyuan Zhou, Ya Zhang, and Yanfeng Wang. DR2: Diffusion-based robust degradation remover for blind face restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1704–1713, 2023.

Shuo Yang, Ping Luo, Chen Change Loy, and Xiaoou Tang. WIDER FACE: A face detection benchmark. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5525–5533, 2016.

Shangchen Zhou, Kelvin C. K. Chan, Chongyi Li, and Chen Change Loy. Towards robust blind face restoration with codebook lookup transformer. In *Advances in Neural Information Processing Systems*, volume 35, pp. 30599–30611, 2022.