GeoFAME: Restoring Beautified Faces via Geometry Correction and Frequency-Aware Multi-Expert Learning

Anonymous authors
Paper under double-blind review

Abstract

With the widespread use of social media platforms and intelligent portrait editing tools, facial beautification has become a common form of image manipulation, raising potential concerns about the authenticity of face images and identity consistency. Existing studies mainly focus on detecting beautified face images, while how to accurately recover their corresponding natural faces remains underexplored. Operations such as face slimming, eye enlargement, skin smoothing, and skin whitening simultaneously alter facial geometry, local texture, and color distribution, making facial beautification restoration a challenging task. To address this problem, we propose GeoFAME, a Geometry-Corrected Frequency-Aware Multi-Expert Network for facial beautification restoration. GeoFAME first alleviates non-rigid deformations introduced by beautification through geometric correction, and then employs a Laplacian pyramid to decompose multi-scale frequency information, with structure, texture, and color experts performing targeted restoration. Meanwhile, a spatially adaptive router dynamically fuses the expert outputs according to the restoration requirements of different facial regions. Extensive experiments demonstrate that GeoFAME effectively restores structural, textural, and color changes caused by both single and composite beautification operations, while achieving favorable restoration performance and cross-dataset generalization under different beautification intensities.

1 Introduction

In recent years, rapid advances in deep generative models and facial editing techniques have enabled DeepFake-based face generation and manipulation systems to produce increasingly realistic visual content, raising growing concerns about the authenticity of digital media and the security of identity-related applications. Research on such manipulated facial content has primarily focused on authenticity detection (Ye et al., 2024; Ding et al., 2026b;a)and on improving the generalization of detectors to unseen generation methods (Zhou et al., 2025). Meanwhile, with the widespread use of social media platforms, mobile photography applications, and intelligent portrait editing tools, facial beautification has become an increasingly common form of face editing. Users can easily modify facial appearance through operations such as skin whitening, skin smoothing, blemish removal, face slimming, eye enlargement, and local facial reshaping to generate portraits that better match their aesthetic preferences. However, while improving visual appearance, these operations may alter facial texture, color distribution, local geometry, and overall contours, thereby reducing the fidelity of the image to an individual's natural appearance. Such alterations may be exploited for commercial deception (Ateq et al., 2024) and online fraud (Bharati et al., 2016), while also disrupting face recognition and undermining the reliability of identity verification(Mirabet Herranz et al., 2022; Hedman et al., 2022; Botezatu et al., 2023; Libourel et al., 2024).Unlike detection tasks that mainly determine whether an image has been generated or manipulated, facial beautification restoration aims to reverse the appearance changes introduced by beautification operations and recover the corresponding natural face as faithfully as possible. To address these issues, current efforts mainly focus on regulatory measures and automatic detection. Some countries have introduced disclosure or labeling requirements for retouched advertisements or facial images (United States Congress, 2014; Forbrukertilsynet, 2021), while researchers have developed dedicated datasets and detection mod-

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.

els to automatically identify retouching traces in face images (Chen et al., 2013; Dantcheva et al., 2012; Ying et al., 2023).Beyond conventional binary detection, recent studies have further explored the localization of manipulated facial regions and the attribution of their corresponding editing or generation sources (He et al., 2026).

Most existing studies focus on detecting whether face images have been retouched, while only a limited number of works investigate how to recover the original face from a beautified image, which is important for face identity tracing. Therefore, we focus on the task of facial beautification restoration. Previous studies have explored related tasks, such as makeup removal and image restoration (IR). Makeup removal methods typically recover a more natural facial appearance by removing or transferring makeup information, while image restoration aims to reconstruct high-quality images from degraded inputs to improve visual clarity. Although these methods are effective in recovering texture-related features, facial beautification restoration involves more than appearance and texture changes. It also requires correcting geometric modifications introduced by operations such as eye enlargement and face slimming. General-purpose image restoration methods (Zamir et al., 2022; Wang et al., 2021; Zhou et al., 2022) provide a theoretical foundation for this task. However, these methods primarily target image degradations and lack explicit modeling of the inverse process of facial beautification. Moreover, generative models may introduce visually plausible but identity-inconsistent details, making them less suitable for facial beautification restoration in identity-sensitive scenarios.

Facial beautification restoration aims to learn a mapping from beautified face images to their corresponding original, unretouched counterparts, and can be regarded as the inverse process of facial beautification. Unlike conventional image restoration tasks such as super-resolution, denoising, and deblurring, facial beautification not only alters low-frequency color and illumination distributions and suppresses high-frequency facial textures, but may also introduce significant geometric distortions through operations such as face slimming and eye enlargement. Recent studies have begun to investigate this inverse problem. Zhao et al. (2025) predicted retouching labels and used them as conditions to guide the reversal process, improving controllability across different retouching operations. Liu et al. (2025) assigned different retouching types to specialized experts, introduced a shared expert for common retouching traces, and combined low- and high-frequency branches to recover global structure and local details. More recently, Xing et al. (2026) combined diffusion models with Mamba-style state-space modeling to enhance long-range dependency modeling for face retouching reversal. These works demonstrate the feasibility of facial beautification restoration, yet challenges remain in handling mixed retouching operations, improving parameter efficiency, preserving realistic details, and maintaining identity consistency.

Therefore, this paper proposes a frequency-guided spatially adaptive multi-expert framework for face beauty restoration. The framework first performs preliminary correction for geometric distortions caused by face slimming, eye enlargement and other operations via the geometric correction branch. Afterwards, the corrected image is fed into a Laplacian pyramid to obtain multi-scale frequency components, which are further processed by the texture, structure and color experts. Specifically, high-frequency and mid-high-frequency components are utilized to recover texture detail losses induced by skin smoothing; mid-frequency components are adopted to remedy local structural deviations of facial contours and organ boundaries; and low-frequency components are responsible for restoring skin tone, brightness and overall hue. Meanwhile, the corrected image is fed into a U-Net to produce a base restoration result that preserves global content. On this basis, we design a spatially adaptive router, which predicts pixel-wise fusion weights for the three experts according to the remaining correction requirements at different positions within the base restored image. Finally, the weighted expert residuals are added to the base restored image to obtain the final natural face image. The main contributions of this work are summarized as follows:

- We propose a frequency-guided multi-expert framework for face beauty restoration. Combined with geometric correction and expert-based residual refinement, our framework jointly handles retouching artifacts including geometric distortions, texture loss and color shift.
- We explicitly separate multi-scale frequency information using the Laplacian pyramid, and construct three dedicated residual experts for texture, structure and color. Each expert performs targeted restoration for its corresponding type of beauty-retouching artifacts.
- We design a spatially adaptive expert routing mechanism, which dynamically assigns expert weights for each pixel based on the base restored image. It achieves region-adaptive fusion

of residuals from different experts and improves the model's capability to cope with mixed retouching operations.

2 Related Work

2.1 Makeup Transfer and Makeup Removal

Makeup transfer(Chen et al., 2017; Li et al., 2018) and makeup removal (Chang et al., 2018) aim to modify or remove cosmetic styles while preserving the underlying identity and basic facial structure. To address large pose variations, complex makeup patterns, and local region editing, Gu et al. (2019) disentangled identity-related and makeup-related information through local adversarial learning. Jiang et al. (2020) introduced spatial-aware makeup representations and attention-based alignment, improving robustness under pose and expression variations. Liu et al. (2021) further combined makeup distillation, identity distillation, and detail supervision, enabling detail-preserving makeup transfer and removal. Sun et al. (2022) modeled semantic correspondences between facial regions with a semantic-aware Transformer, improving makeup alignment under complex pose and local appearance variations. Yang et al. (2022) emphasized local editability and the preservation of high-frequency makeup details. Recently, diffusion models have also been applied to adversarial makeup transfer as well as text-conditioned makeup editing and removal (Sun et al., 2024; Ruan & Shuai, 2024).

2.2 Facial Retouching Detection

Facial retouching detection focuses on identifying whether a face image has undergone operations such as whitening, smoothing, local facial reshaping, or other retouching edits. Ying et al. (2023) trained supervised deep learning models to learn local retouching artifacts and distinguish retouched faces from original ones. Bharati et al. (2017) further studied facial retouching detection under demographic variations, proposed the MDRF multi-ethnic dataset (building upon their prior ND-IIITD Retouched Faces dataset), and revealed severe generalization gaps of retouch detectors across different ethnic and population groups. Rathgeb et al. (2020b) proposed a differential detection strategy that compares biometric discrepancies between a trusted reference image and a probe image to identify retouching. Their subsequent work analyzed the impact of retouching on photo-response non-uniformity traces, providing an imaging-forensics perspective for retouching detection Rathgeb et al. (2020a). These methods are useful for forensic detection, but they do not recover the unretouched reference face. These methods can determine whether an image has been modified, but they are unable to identify the specific retouching operations or their intensities. In addition, recent studies (Jain et al., 2018; Geng et al., 2020) have further explored model attribution by identifying the retouching networks used to process an image, as well as detecting the models employed for face occlusion, thereby providing new insights into digital image provenance and editing behavior analysis.

3 Method

3.1 Proposed Framework

We propose a frequency-aware spatially adaptive multi-expert framework for facial beautification restoration, whose overall pipeline is illustrated in Figure 1.

We decompose the facial beautification restoration process into several stages, including geometric correction, feature encoding and frequency decomposition, and expert-guided restoration. This design enables specialized modeling for different types of beautification operations and dynamically adjusts the contributions of individual experts according to the restoration requirements of different facial regions. First, the geometric correction branch predicts the non-rigid deformations caused by operations such as face slimming and eye enlargement, and performs preliminary geometric correction on the input image using the predicted displacement field:

$$\mathbf{D} = \mathcal{G}_{\theta_g}(\mathbf{x}), \qquad \mathbf{x}_r = \mathcal{W}(\mathbf{x}, \mathbf{D}), \tag{1}$$

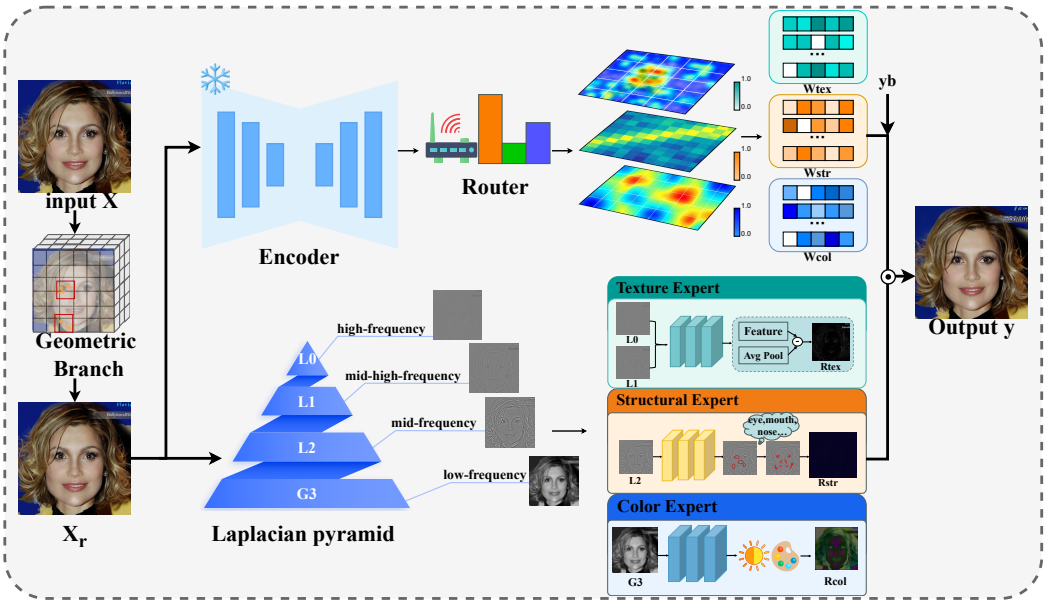

Figure 1: The proposed frequency-aware spatially adaptive multi-expert framework for facial beautification restoration first performs preliminary recovery through geometric correction. Subsequently, the encoder extracts shared restoration features, while the Laplacian pyramid decomposes the input into multi-scale frequency information to guide the structure, texture, and color experts in predicting restoration residuals. The spatially adaptive router generates pixel-wise expert weights based on the encoded features, and the weighted residuals are finally fused to produce the restored face.

where $\mathcal{G}_{\theta_g}$ denotes the geometric correction branch, $\mathbf{D}$ is the predicted displacement field, $\mathcal{W}(\cdot)$ represents the image warping operation based on the displacement field, and $\mathbf{x}_r$ denotes the geometrically corrected image.

Subsequently, the corrected image $\mathbf{x}_r$ is fed into the frequency-expert branch and the encoder branch, respectively. In the frequency-expert branch, a four-level Laplacian pyramid is adopted to perform multi-scale frequency decomposition on $\mathbf{x}_r$:

$$\{\mathbf{L}_0, \mathbf{L}_1, \mathbf{L}_2, \mathbf{G}_3\} = \mathcal{P}(\mathbf{x}_r), \tag{2}$$

where $\mathbf{L}_0$ and $\mathbf{L}_1$ denote the high-frequency and mid-high-frequency residual components, respectively, which mainly contain skin textures, pores, and fine edges. $\mathbf{L}_2$ denotes the mid-frequency residual component that mainly describes facial contours and local structures, while $\mathbf{G}_3$ represents the Gaussian low-frequency base at the lowest resolution, which mainly preserves skin tone, brightness, and overall appearance. These components are further fed into the texture, structure, and color experts, respectively.

In the encoder branch, the geometrically corrected image $\mathbf{x}_r$ is fed into an encoder to extract shared features containing global content and restoration-related information:

$$\mathbf{F}_e = \mathcal{E}_{\theta_e}(\mathbf{x}_r), \tag{3}$$

where $\mathcal{E}_{\theta_e}$ denotes the feature encoder, and $\mathbf{F}_e$ represents the shared restoration features extracted by the encoder.

Considering that different beautification operations affect facial regions to varying degrees, we further introduce a spatially adaptive router. Taking the shared feature $\mathbf{F}_e$ extracted by the encoder as input, the router predicts pixel-wise spatial weight maps for the three experts:

$$[\mathbf{W}_{\text{tex}}, \mathbf{W}_{\text{str}}, \mathbf{W}_{\text{col}}] = \text{Softmax}\left(\mathcal{R}_{\theta_r}(\mathbf{F}_e)\right), \tag{4}$$

where $\mathbf{W}_{\text{tex}}$, $\mathbf{W}_{\text{str}}$, and $\mathbf{W}_{\text{col}}$ denote the spatial routing weights of the texture, structure, and color experts, respectively.

Finally, the residuals predicted by the three experts are fused in a pixel-wise manner according to the spatial routing weights. The fused expert residuals are then combined with the geometrically corrected image $\mathbf{x}_r$ to obtain the final restored image:

$$\hat{\mathbf{y}} = \mathbf{x}_r + \mathbf{W}_{\text{tex}} \odot \mathbf{R}_{\text{tex}} + \mathbf{W}_{\text{str}} \odot \mathbf{R}_{\text{str}} + \mathbf{W}_{\text{col}} \odot \mathbf{R}_{\text{col}}, \tag{5}$$

where $\odot$ denotes element-wise multiplication.

Through frequency decomposition, expert-specific modeling, and pixel-wise spatial routing, the proposed framework adaptively coordinates texture, structure, and color restoration across different facial regions, thereby producing natural restoration results with consistent facial details.

3.2 Frequency-Guided Multi-Expert Refinement

Different types of beauty-retouching operations exert distinct impacts on various frequency components of face images. For example, skin smoothing usually weakens mid-to-high-frequency details such as pores, fine wrinkles and local edges; while whitening, brightness adjustment and skin tone modification are mostly manifested as low-frequency appearance variations. Therefore, we perform a four-level Laplacian pyramid decomposition on the geometrically corrected image $\mathbf{L}_0$, $\mathbf{L}_1$, $\mathbf{L}_2$, and $\mathbf{G}_3$. According to the information embedded in different frequency components, we construct three dedicated residual experts for texture, structure and color, so that each expert learns specific types of restoration within a constrained frequency range.

Texture Expert The texture expert takes the high-frequency component $\mathbf{L}_0$ and the mid-high-frequency component $\mathbf{L}_1$ as inputs. Specifically, $\mathbf{L}_0$ mainly contains fine-grained textures such as pores, fine wrinkles, and subtle edges, while $\mathbf{L}_1$ provides complementary local texture and edge information. Exploiting both frequency levels improves the completeness of texture restoration and reduces the noise and local artifacts caused by relying on a single high-frequency component. Furthermore, we apply local mean removal to the initial texture residual to suppress slowly varying low-frequency brightness and color components. This encourages the texture expert to focus more on high-frequency detail restoration while reducing its influence on global skin tone and illumination distribution.

Structure Expert The structure expert takes the mid-frequency component $\mathbf{L}_2$ as input and exploits the facial contour and local organ boundary information contained in this frequency band to refine structural deviations around the eyes, mouth, nose, and facial contours. Unlike the geometric correction branch, which mainly handles large-scale spatial misalignment, the structure expert further corrects residual contour discontinuities, local boundary offsets, and abnormal organ shapes after geometric alignment, thereby achieving a progressive restoration process from coarse-grained geometric correction to mid-scale structural refinement.

Color Expert The color expert takes the low-frequency base $\mathbf{G}_3$ as input and utilizes the global skin tone, brightness, and illumination information contained in this component to restore low-frequency appearance variations caused by whitening and skin-tone adjustment. The expert predicts color residuals in the low-resolution space and then upsamples them to the original resolution, thereby producing smoother and more stable color corrections while reducing local color spots and unnatural color variations.

Feature Encoding and Spatially Adaptive Routing. Considering the spatially non-uniform effects of different beauty-retouching operations across facial regions, we further design a spatially adaptive router that takes the shared restoration features $\mathbf{F}_e$ extracted by the encoder as input and predicts pixel-wise fusion weights for the three experts. For each spatial position $\mathbf{p}$, the routing weights are defined as

$$[W_{\text{str}}(\mathbf{p}), W_{\text{tex}}(\mathbf{p}), W_{\text{col}}(\mathbf{p})] = \text{Softmax}\left(\mathbf{A}(\mathbf{p})\right). \tag{6}$$

Here, $\mathbf{A}(\mathbf{p})$ denotes the routing responses generated from the encoded features. The Softmax operation ensures that all expert weights are non-negative and satisfy

$$W_{\text{str}}(\mathbf{p}) + W_{\text{tex}}(\mathbf{p}) + W_{\text{col}}(\mathbf{p}) = 1. \tag{7}$$

Therefore, the three experts form a normalized adaptive competition mechanism at each pixel location. According to the restoration requirements of different facial regions, the router dynamically

adjusts the relative contributions of the structure, texture, and color experts. Finally, the residuals predicted by the three experts are adaptively weighted according to the corresponding spatial routing weights and combined with the geometrically corrected image $\mathbf{x}_r$ to obtain the final restored image.

4 EXPERIMENTS

4.1 EXPERIMENTAL SETUP

Implementation Details. We train the proposed face beautification restoration network using the AdamW optimizer with an initial learning rate of 5×10^{-5}. The batch size is set to 2, and the model is trained for 15 epochs. The same training configuration is used throughout all experiments to ensure consistency across different experimental settings.

Dataset. We use the public paired face-retouching dataset RetouchFFHQ (Ying et al., 2023), which contains facial edits such as eye enlargement and skin whitening. To enrich the diversity of retouching operations, we apply Meitu Xiuxiu to perform beauty retouching on the CelebA dataset. Considering the model's restoration performance under different retouching intensities, we set two retouching levels of 30% and 60%, yielding 1000 and 500 retouched images, respectively. During training and testing, retouched images are used as model inputs, and their corresponding natural facial images are regarded as restoration targets. Figure 2 shows several examples from the dataset.

Comparing methods. To comprehensively evaluate the effectiveness of the proposed method, we compare it with several representative approaches, including PCGAN (Liu et al., 2021), SCGAN (Deng et al., 2021), RestoreFormer (Wang et al., 2022), DR2 (Wang et al., 2023), SSAT (Sun et al., 2022), MAD (Ruan & Shuai, 2025), OSDFace (Wang et al., 2025), MoFRR (Liu et al., 2025), and DiffClean (Gavas et al., 2026). SCGAN and SSAT focus on makeup transfer and removal, while MAD supports multiple makeup editing tasks through cross-domain diffusion. RestoreFormer, DR2, DiffClean, and OSDFace are designed for face restoration or generation, with DR2 and DiffClean adopting diffusion-based restoration strategies.

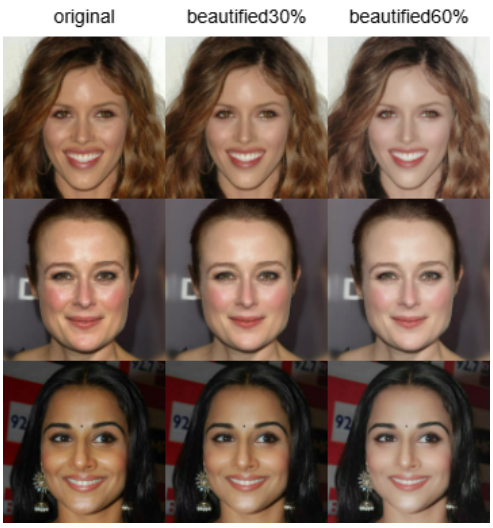

Figure 2: Examples from CelebA under different retouching intensities. Each row keeps the same retouching type.

Evaluation Metrics. We quantitatively evaluate restoration performance using PSNR, SSIM (Wang et al., 2004), LPIPS (Zhang et al., 2018), and FID (Lyu et al., 2019). PSNR measures pixel-level reconstruction fidelity between the restored and ground-truth images, while SSIM evaluates structural similarity. LPIPS measures perceptual discrepancy in deep feature space, and FID further evaluates the distributional difference between the restored and natural face images. Higher PSNR and SSIM values indicate better reconstruction quality, whereas lower LPIPS and FID values indicate better perceptual quality and distributional consistency.

4.2 RESTORATION PERFORMANCE EVALUATION

Table 1 presents the quantitative comparison on the CelebA dataset under two beautification strengths. Overall, PSGAN, SCGAN, RestoreFormer, DR2, OSDFace, and MAD can restore beautified faces to a certain extent, but still exhibit varying limitations in pixel-level reconstruction, structural preservation, or perceptual quality. In comparison, DiffClean, MoFRR, and our method achieve more competitive overall restoration performance.

Under the 30% beautification setting, our method achieves a PSNR of 32.1517 dB, outperforming DiffClean by approximately 1.31 dB, while also obtaining the best SSIM and FID scores. Under the

stronger 60% beautification setting, the PSNR further increases to 35.6723 dB, exceeding DiffClean by approximately 4.90 dB, while maintaining the best SSIM and FID. These results indicate that our method preserves favorable structural restoration capability and distributional consistency even under stronger beautification distortions.

Although MoFRR achieves slightly better LPIPS scores, our method performs better in terms of PSNR, SSIM, and FID. Overall, these results demonstrate that the proposed method achieves a favorable balance among pixel-level reconstruction accuracy, structural preservation, and distributional consistency, while maintaining stable restoration performance across different beautification strengths.

Table 1: Quantitative comparison of different methods on the CelebA dataset under 30% and 60% beautification strengths. The best results are highlighted in bold and the second-best results are underlined.

Method	30% Beau.				60% Beau.			
	PSNR↑	SSIM↑	LPIPS↓	FID↓	PSNR↑	SSIM↑	LPIPS↓	FID↓
PSGAN(CVPR2021)	20.6786	0.8375	0.0797	18.0144	20.7793	0.8628	0.0709	24.0114
SCGAN(CVPR2021)	25.8837	0.8256	0.1303	35.8089	24.1724	0.7982	0.1799	53.9759
RestoreFormer(CVPR2022)	24.0260	0.7228	0.0960	21.8759	22.6924	0.7274	0.0936	30.8826
DR2(CVPR2023)	24.2885	0.7180	0.1436	32.0144	21.5009	0.6358	0.2075	50.9507
OSDFace(CVPR2025)	20.9594	0.6143	0.2433	25.1049	20.0431	0.6170	0.2441	34.1121
MoFRR(ICCV2025)	29.4234	0.8927	**0.0271**	5.7299	28.2706	0.9094	**0.0153**	7.7546
MAD(CVPR2025)	19.9481	0.6911	0.2232	78.8955	19.5450	0.7081	0.2206	89.4291
DiffClean(CVPR2026)	30.8467	0.9092	0.0297	18.1537	30.7712	0.9451	0.0170	21.7769
Ours	**32.1517**	**0.9118**	0.0372	**5.0915**	**35.6723**	**0.9619**	0.0161	**4.1184**

Figure 3 presents the visual comparison between our method and other approaches under the 30% beautification setting. PSGAN can alleviate beautification-induced appearance changes to some extent, but it still suffers from color shifts and insufficient recovery of local textures. SCGAN also shows limited restoration capability for certain beautification effects, particularly around the eye makeup regions. RestoreFormer retains noticeable beautification traces, with some skin textures and local details insufficiently recovered, while DR2 tends to produce overly smooth or even blurry results in some cases, leading to the loss of local facial details. OSDFace is prone to introducing artificial artifacts and color shifts, whereas MAD exhibits more evident visual artifacts and facial structural distortions. In comparison, DiffClean and MoFRR produce relatively natural restoration results, but they still show limitations in recovering skin textures and fine-grained local details. Our method better restores natural skin tones, skin textures, and local details while maintaining stable facial structures, resulting in restored faces that are overall more consistent with the corresponding original natural faces.

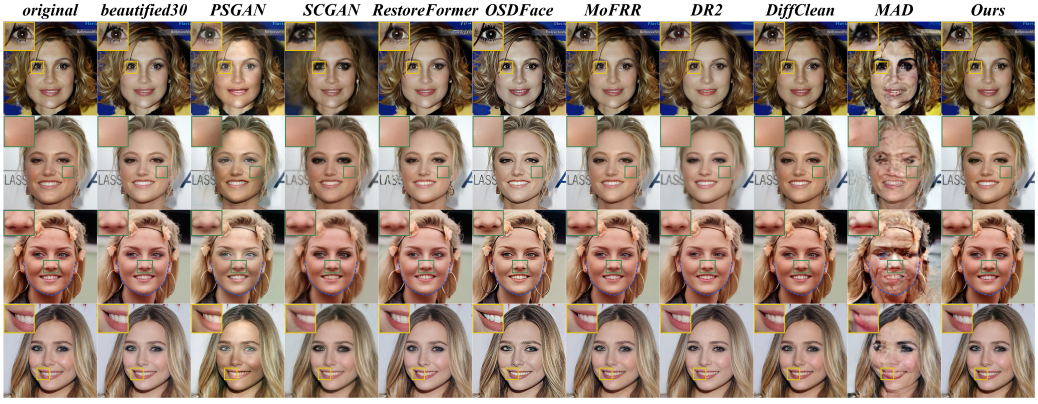

Figure 3: Qualitative comparison of restored faces obtained by different methods. The images are processed with multiple beauty-retouching operations. Our method achieves superior restoration quality with fewer artifacts. Zoom in for best view.

To further evaluate the cross-dataset generalization ability of our method, we conduct experiments on the RetouchFFHQ dataset, which contains 1,000 pairs of natural and beautified facial images covering eye enlargement, skin whitening, skin smoothing, face lifting, and composite beautification effects produced by multiple operations. In Table 2, 'Single" denotes the averaged results over the four individual beautification operations, while 'Multi" denotes the results under composite beautification. Despite the noticeable distribution gap between the training and test datasets, our method still demonstrates strong restoration performance. Under the single-operation setting, our method achieves an SSIM of 0.9478. Under the more challenging composite beautification setting, it still obtains a PSNR of 22.7003 dB and an SSIM of 0.8684. These results demonstrate that our method exhibits good cross-dataset generalization across different beautification types and data distributions.

Table 2: Cross-dataset comparison on the RetouchFFHQ dataset under single-type and multi-type beauty manipulation settings. "Single" reports the average performance over four individual beauty operations, including eye enlargement, skin whitening, skin smoothing, and face lifting, while "Multi" denotes the results under their combined manipulation. Best results are highlighted in bold, and second-best results are underlined.

Method	Single				Multi			
	PSNR↑	SSIM↑	LPIPS↓	FID↓	PSNR↑	SSIM↑	LPIPS↓	FID↓
PSGAN(CVPR2021)	20.8415	0.8840	0.0690	16.4033	18.9910	0.8105	0.0982	22.6978
SCGAN(CVPR2021)	18.9820	0.7371	0.2838	50.1551	18.4363	0.7030	0.2928	53.0952
RestoreFormer(CVPR2022)	24.4838	0.6932	0.1173	18.0350	19.7076	0.6476	0.1476	21.2140
DR2(CVPR2023)	23.3750	0.6513	0.2352	37.0947	19.3520	0.6194	0.2579	41.0987
OSDFace(CVPR2025)	22.8460	0.7026	0.1635	24.3044	18.2455	0.6488	0.1872	27.5115
MoFRR(ICCV2025)	<u>31.2315</u>	0.9138	**0.0132**	<u>3.7595</u>	<u>21.7005</u>	0.8286	<u>0.0528</u>	13.2175
MAD(CVPR2025)	19.8441	0.7038	0.2266	91.7426	17.6472	0.6545	0.2465	91.9270
DiffClean(CVPR2026)	29.8851	<u>0.9306</u>	0.0604	8.0812	20.2230	**0.9628**	**0.0335**	<u>3.8374</u>
Ours	**31.4257**	**0.9478**	<u>0.0323</u>	**3.2709**	**22.7003**	<u>0.8684</u>	0.0861	**3.2833**

4.3 User Study

To further evaluate the subjective perceptual quality of the restored results, we conduct a user study. We randomly select 100 beautified face images from the FFHQ test set and compare our method with PSGAN, SCGAN, RestoreFormer, DR2, DiffClean, OSDFace, MAD, and MoFRR. A total of 20 evaluators are invited to participate in the subjective assessment. During evaluation, the beautified input, the corresponding natural face reference, and the restoration results produced by different methods are presented simultaneously. To reduce subjective bias, all method names are hidden and the display order of the restoration results is randomized.

Evaluators rate each restored image on a 1–5 scale from three aspects: **Identity Fidelity**, which measures the consistency of identity characteristics and facial structures; **Restoration Fidelity**, which evaluates the recovery of beautification-induced changes in geometry, texture, and color; and **Visual Naturalness**, which assesses the overall naturalness of the restored results and the presence of obvious artifacts. Table 3 reports the average scores across all evaluators and test samples. Our method achieves the highest scores on all three metrics, demonstrating better subjective performance in identity preservation, beautification restoration, and visual naturalness.

4.4 Ablation Study

Ablation analysis of expert combinations. Table 4 presents the ablation results of different expert combinations under 30% and 60% beautification intensities. Activating only a single expert yields limited improvements on partial metrics, while dual-expert combinations also show performance fluctuations and may even underperform the setting without experts. This suggests that structure, texture, and color restoration are not independent processes and require effective collaboration among experts. When all three experts are jointly activated, the model achieves the best overall performance under both intensities. Compared with the configuration without experts, PSNR improves from 30.1727 dB to 32.1517 dB at 30%, and from 33.3789 dB to 35.6723 dB at 60%, with overall

Table 3: Quantitative results of the user study. All metrics are evaluated on a 1–5 scale, where higher scores denote better subjective performance.

Method	Identity Fidelity	Restoration Fidelity	Visual Naturalness
PSGAN(CVPR2021)	3.42	3.82	3.21
SCGAN(CVPR2021)	3.84	3.56	3.95
RestoreFormer(CVPR2022)	3.74	2.06	2.28
DR2(CVPR2023)	2.94	2.81	2.64
OSDFace(CVPR2025)	3.26	3.47	3.09
MoFRR(ICCV2025)	3.95	3.93	3.99
MAD(CVPR2025)	1.78	1.45	1.26
DiffClean(CVPR2026)	3.96	3.81	4.07
Ours	**4.22**	**4.01**	**4.18**

improvements in SSIM, LPIPS, and FID. These results demonstrate that the three experts provide complementary restoration cues, enabling more stable and comprehensive recovery of diverse beautification effects.

Table 4: Ablation study of the geometry correction branch and expert combinations under 30% and 60% beautification intensities. The geometry-branch results are obtained from an independently retrained control experiment and are compared only within the corresponding two settings. For expert combinations, "000" denotes using no expert module, while "111" denotes jointly using the structure, texture, and color experts. Best results within the expert-combination group are highlighted in bold.

Ablation Setting	30% Beautification				60% Beautification			
	PSNR↑	SSIM↑	LPIPS↓	FID↓	PSNR↑	SSIM↑	LPIPS↓	FID↓
w/o Geom. Branch	30.6204	0.8878	0.0608	13.3113	33.0801	0.9465	0.0179	7.6703
w/ Geom. Branch	31.2484	0.8912	0.0621	14.1485	33.2539	0.9446	0.0191	7.6694
000 (No expert)	30.1727	0.8872	0.0658	6.4413	33.3789	0.9222	0.0971	4.6314
100 (Structure only)	30.2541	0.8904	0.0627	5.4852	33.2684	0.9162	0.1002	4.9490
010 (Texture only)	30.0616	0.8825	0.0647	6.8835	33.2794	0.9221	0.0941	4.6576
001 (Color only)	30.0840	0.8915	0.0634	5.9701	33.1500	0.9205	0.0952	4.8321
110 (Structure+Texture)	29.4652	0.8678	0.0653	6.3082	32.9617	0.9140	0.0927	4.4370
101 (Structure+Color)	30.3894	0.8907	0.0631	5.6654	33.4379	0.9222	0.0964	5.5930
011 (Texture+Color)	29.8993	0.8858	0.0653	6.3267	33.4928	0.9225	0.0938	4.6583
111 (All experts)	**32.1517**	**0.9118**	**0.0372**	**5.0915**	**35.6723**	**0.9619**	**0.0161**	**4.1184**

Ablation analysis of the geometry branch. We further conduct an independent controlled experiment to evaluate the effectiveness of the geometry correction branch, as shown in Table 4. At the 30% beautification intensity, introducing the geometry branch improves PSNR from 30.6204 dB to 31.2484 dB and SSIM from 0.8878 to 0.8912. At the 60% intensity, PSNR also increases from 33.0801 dB to 33.2539 dB. Despite slight fluctuations in some perceptual metrics, the overall results show that geometry correction improves reconstruction accuracy and provides more noticeable structural gains under the lower beautification intensity.

5 Conclusion

we proposes GeoFAME, a Geometry-Corrected Frequency-Aware Multi-Expert Network for facial beautification restoration, to address geometric deformation, texture loss, and color changes caused by beautification. GeoFAME first alleviates non-rigid deformations through geometric correction, and then employs multi-scale frequency decomposition to guide structure, texture, and color experts for targeted restoration. A spatially adaptive routing mechanism dynamically fuses the expert outputs. Extensive experiments demonstrate that GeoFAME effectively handles both single and composite beautification operations across different intensities, achieving favorable restoration performance and cross-dataset generalization.

6 AI USE STATEMENT

Generative AI tools were used only for language polishing, grammar checking, and readability improvement. They were not involved in methodology development, experimentation, or result interpretation. All AI-assisted content was reviewed by the authors, who take full responsibility for the final manuscript.

REFERENCES

Khadijah Ateq, Mohammed Alhajji, and Noara Alhusseini. The association between use of social media and the development of body dysmorphic disorder and attitudes toward cosmetic surgeries: a national survey. *Frontiers in Public Health*, 12:1324092, 2024. doi: 10.3389/fpubh.2024.1324092.

Aparna Bharati, Richa Singh, Mayank Vatsa, and Kevin W. Bowyer. Detecting facial retouching using supervised deep learning. *IEEE Transactions on Information Forensics and Security*, 11(9): 1903–1913, 09 2016. doi: 10.1109/TIF.2016.2561898.

Aparna Bharati, Mayank Vatsa, Richa Singh, Kevin W Bowyer, and Xin Tong. Demography-based facial retouching using subclass supervised sparse autoencoder. In *2017 IEEE International Joint Conference on Biometrics (IJCB)*, pp. 474–482. IEEE, 2017.

Cristian Botezatu, Mathias Ibsen, Christian Rathgeb, and Christoph Busch. Fun selfie filters in face recognition: Impact assessment and removal. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 5(1), 2023. doi: 10.1109/TBIOM.2022.3185884.

Huiwen Chang, Jingwan Lu, Fisher Yu, and Adam Finkelstein. Pairedcyclegan: Asymmetric style transfer for applying and removing makeup. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3036–3045, 2018.

Cunjian Chen, Antitza Dantcheva, and Arun Ross. Automatic facial makeup detection with application in face recognition. In *2013 International Conference on Biometrics (ICB)*, pp. 1–8. IEEE, 2013.

Ying-Cong Chen, Xiaoyong Shen, and Jiaya Jia. Makeup-go: Blind reversion of portrait edit. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 1–8, 2017.

Antitza Dantcheva, Cunjian Chen, and Arun Ross. Can facial cosmetics affect the matching accuracy of face recognition systems? In *2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pp. 391–398. IEEE, 2012.

Han Deng, Chu Han, Hongmin Cai, Guoqiang Han, and Shengfeng He. Spatially-invariant style-codes controlled makeup transfer. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6545–6553. IEEE, 2021.

Feng Ding, Wenhui Yi, Xinan He, Mengyao Xiao, Jianfeng Xu, and Jianqiang Du. Diffface-edit: A diffusion-based facial dataset for forgery-semantic driven deepfake detection analysis. In *ICASSP 2026-2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 14117–14121. IEEE, 2026a.

Feng Ding, Wenhui Yi, Yunpeng Zhou, Xinan He, Hong Rao, and Shu Hu. Decoupling bias, aligning distributions: Synergistic fairness optimization for deepfake detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 32704–32713, 2026b.

Forbrukertilsynet. The marketing control act, 2021. URL `https://www.forbrukertilsynet.no/english/the-marketing-control-act`.

Ekta Gavas, Sudipta Banerjee, Chinmay Hegde, and Nasir Memon. Diffclean: Diffusion-based makeup removal for accurate age estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 1516–1526, 2026.

Mengyue Geng, Peixi Peng, Yangru Huang, and Yonghong Tian. Masked face recognition with generative data augmentation and domain constrained ranking. In *Proceedings of the 28th ACM International Conference on Multimedia*, pp. 2246–2254, 2020. doi: 10.1145/3394171.3414271.

Qiao Gu, Guanzhi Wang, Mang Tik Chiu, Yu-Wing Tai, and Chi-Keung Tang. Ladn: Local adversarial disentangling network for facial makeup and de-makeup. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 1042–1051, 2019.

Xinan He, Yue Zhou, Bing Fan, Bin Li, Guopu Zhu, and Feng Ding. Vlforgery face triad: Detection, localization and attribution via multimodal large language models. *Advances in Neural Information Processing Systems*, 38:163010–163044, 2026.

Pontus Hedman, Vasilios Skepetzis, Kevin Hernandez-Diaz, Josef Bigun, and Fernando Alonso-Fernandez. On the effect of selfie beautification filters on face detection and recognition. *Pattern Recognition Letters*, 163:114–121, 2022. doi: 10.1016/j.patrec.2022.09.018.

Anubhav Jain, Richa Singh, and Mayank Vatsa. On detecting gans and retouching based synthetic alterations. In *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pp. 1–8. IEEE, 2018. doi: 10.1109/BTAS.2018.8533720.

Wentao Jiang, Si Liu, Chen Gao, Jie Cao, Ran He, Jiashi Feng, and Shuicheng Yan. Psgan: Pose and expression robust spatial-aware gan for customizable makeup transfer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5194–5202, 2020.

Tingting Li, Ruihe Qian, Chao Dong, Si Liu, Qiong Yan, Wenwu Zhu, and Liang Lin. Beautygan: Instance-level facial makeup transfer with deep generative adversarial network. In *Proceedings of the 26th ACM International Conference on Multimedia (MM '18)*, pp. –, Seoul, Republic of Korea, October 2018. ACM. ISBN 978-1-4503-5665-7. doi: 10.1145/3240508.3240618.

Alexandre Libourel, Sahar Husseini, Nelida Mirabet-Herranz, and Jean-Luc Dugelay. A case study on how beautification filters can fool deepfake detectors. In *2024 12th International Workshop on Biometrics and Forensics (IWBF)*, pp. 1–6. IEEE, 2024. doi: 10.1109/IWBF62628.2024.10593932.

Jiaxin Liu, Qichao Ying, Zhenxing Qian, Sheng Li, Runqi Zhang, Jian Liu, and Xinpeng Zhang. Mofrr: Mixture of diffusion models for face retouching restoration. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 12842–12850. IEEE, 2025. doi: 10.1109/ICCV51701.2025.01193.

Si Liu, Wentao Jiang, Chen Gao, Ran He, Jiashi Feng, Bo Li, and Shuicheng Yan. Psgan++: Robust detail-preserving makeup transfer and removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(10):3538–3555, 2021. doi: 10.1109/TPAMI.2021.3088684.

He Lyu, Ningyu Sha, Shuyang Qin, Ming Yan, Yuying Xie, and Rongrong Wang. Advances in neural information processing systems. *Advances in neural information processing systems*, 32, 2019.

Nelida Mirabet Herranz, Chiara Galdi, and Jean-Luc Dugelay. Impact of digital face beautification in biometrics. In *2022 10th European Workshop on Visual Information Processing (EUVIP)*, pp. 1–6, Lisbon, Portugal, Sep 2022. IEEE. ISBN 978-1-6654-6623-3. doi: 10.1109/EUVIP53989.2022.9922802.

Christian Rathgeb, Angelika Botaljov, Fabian Stockhardt, Sergey Isadskiy, Luca Debiasi, Andreas Uhl, and Christoph Busch. Prnu-based detection of facial retouching. *IET Biometrics*, 9(4):154–164, 2020a. doi: 10.1049/iet-bmt.2019.0196.

Christian Rathgeb, Cezar-Iustin Satnoianu, Nanda E. Haryanto, Kevin Bernardo, and Christoph Busch. Differential detection of facial retouching: A multi-biometric approach. *IEEE Access*, 8:106373–106385, 2020b. doi: 10.1109/ACCESS.2020.3000254.

Bo-Kai Ruan and Hong-Han Shuai. Mad: Makeup all-in-one with cross-domain diffusion model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 749–758, 2024.

Bo-Kai Ruan and Hong-Han Shuai. Mad: makeup all-in-one with cross-domain diffusion model. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 740–749. IEEE, 2025.

Yuhao Sun, Lingyun Yu, Hongtao Xie, Jiaming Li, and Yongdong Zhang. Diffam: Diffusion-based adversarial makeup transfer for facial privacy protection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 24584–24594, 2024.

Zhaoyang Sun, Yaxiong Chen, and Shengwu Xiong. Ssat: A symmetric semantic-aware transformer network for makeup transfer and removal. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pp. 2325–2333, 2022.

United States Congress. Truth in advertising act of 2014. H.R. 4341, 113th Congress, 2014. URL `https://www.congress.gov/bill/113th-congress/house-bill/4341`.

Jingkai Wang, Jue Gong, Lin Zhang, Zheng Chen, Xing Liu, Hong Gu, Yutong Liu, Yulun Zhang, and Xiaokang Yang. Osdface: One-step diffusion model for face restoration. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12626–12636. IEEE, 2025.

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 (CVPR)*, 2021.

Zhixin Wang, Ziying Zhang, Xiaoyun Zhang, Huangjie Zheng, Mingyuan Zhou, Ya Zhang, and Yanfeng Wang. Dr2: Diffusion-based robust degradation remover for blind face restoration. In *Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1704–1713, 2023. doi: 10.1109/CVPR52729.2023.00170.

Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

Zhouxia Wang, Jiawei Zhang, Runjian Chen, Wenping Wang, and Ping Luo. Restoreformer: High-quality blind face restoration from undegraded key-value pairs. In *Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17491–17500, 2022. doi: 10.1109/CVPR52688.2022.01699.

Fengchuang Xing, Xiaowen Shi, Chao Zhou, and Hongwei Yue. A diffusion-mamba synergetic framework for face retouching reversal. *Engineering Applications of Artificial Intelligence*, 179: 115170, 2026. doi: 10.1016/j.engappai.2026.115170.

Chenyu Yang, Wanrong He, Yingqing Xu, and Yang Gao. Elegant: Exquisite and locally editable gan for makeup transfer. In *European Conference on Computer Vision (ECCV)*, volume 13676, pp. 737–754. Springer Nature Switzerland AG, 2022. doi: 10.1007/978-3-031-19787-1_42.

Wei Ye, Xinan He, and Feng Ding. Decoupling forgery semantics for generalizable deepfake detection. *arXiv preprint arXiv:2406.09739*, 2024.

Qichao Ying, Jiaxin Liu, Sheng Li, Haisheng Xu, Zhenxing Qian, and Xinpeng Zhang. Retouchingffhq: A large-scale dataset for fine-grained face retouching detection. In *Proceedings of the 31st ACM International Conference on Multimedia (MM '23)*, pp. 737–746. ACM, 10 2023. ISBN 979-8-4007-0108-5. doi: 10.1145/3581783.3611843.

Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5718–5728. IEEE, 2022. doi: 10.1109/CVPR52688.2022.00564.

Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *2018 IEEE/CVF conference on computer vision and pattern recognition*, pp. 586–595. IEEE, 2018.

Guanhua Zhao, Yu Gu, Xuhan Sheng, Yujie Hu, and Jian Zhang. Label-guided facial retouching reversion. In *2025 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6. IEEE, 2025. doi: 10.1109/ICME59968.2025.11209321.

Shangchen Zhou, Kelvin C.K. Chan, Chongyi Li, and Chen Change Loy. Towards robust blind face restoration with codebook lookup transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. –, 2022. URL https://shangchenzhou.com/projects/CodeFormer.

Yue Zhou, Xinan He, Kaiqing Lin, Bing Fan, Feng Ding, and Bin Li. Breaking latent prior bias in detectors for generalizable aigc image detection. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 30649–30679. Curran Associates, Inc., 2025. doi: 10.52202/085713-1030. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/2c2be72635379ec8600eea13df4743e3-Paper-Conference.pdf.