Textile & Leather Review

ISSN 2623-6281| www.tlr-journal.com | 10.31881/TLR

Real-time Face Retouching for High Resolution Images

Xinghong Hu, Zhenyu Xiao, Wuyao Shen, Sichuang Xu

How to cite: Hu X, Xiao Z, Shen W, Xu S. Real-time Face Retouching for High Resolution Images. Textile & Leather Review. 2026; 9:1826-1841. https://doi.org/10.31881/TLR.2026.1826

How to link: https://doi.org/10.31881/TLR.2026.1826

Published: 25 April 2026

Real-time Face Retouching for High Resolution Images

Xinghong Hu[1†], Zhenyu Xiao[2†], Wuyao Shen[3*], Sichuang Xu[2]

[1]Department of Artificial Intelligence, Shenzhen Polytechnic University, Shenzhen 518055, Guangdong, China

[2]ulingjishi (Shenzhen) Technology Co., Ltd., Shenzhen 518110, Guangdong, China

[3]BinoViz Limitied, Hong Kong 999077, China

*ceo@binoviztechnology.com

[†]Xinghong HU and Zhenyu XIAO contributed equally to this work, they are both first authors.

Article
https://doi.org/10.31881/TLR.2026.1826
Published 25 April 2026

ABSTRACT

In this paper, we propose a real-time face retouching framework based on a neutral gray layer representation. Instead of directly predicting the retouched image, our method learns a neutral gray layer using image-to-image translation models derived from CycleGAN and pix2pix. During inference, the input image is processed at a reduced resolution for efficiency, and the predicted gray layer is upsampled and blended with the original image via a soft-light composition strategy. This design effectively decouples appearance enhancement from texture modeling, enabling natural-looking retouching while preserving fine skin details. The proposed method provides an efficient solution for enhancing portrait realism in fashion imaging and virtual try-on systems, where skin texture and garment details must coexist harmoniously. We train a face retouching model on the FFHQR dataset, a large-scale professionally retouched face dataset derived from FFHQ, and further demonstrate the flexibility of the proposed framework by training additional models for wrinkle reduction and skin smoothing on separate datasets. Quantitative evaluations using SSIM and PSNR demonstrate that the proposed method achieves favorable reconstruction fidelity compared to baseline approaches. In addition, GPU-based inference benchmarks show that our framework supports real-time performance, making it suitable for practical applications such as live video beautification, interactive fashion displays, and digital garment exhibitions.

KEYWORDS

face retouching, generative adversarial networks, neutral gray layer, real-time Image processing, virtual fashion display

INTRODUCTION

Face retouching and beautification play a crucial role in modern visual applications, including mobile photography, social media, and video conferencing. In the context of the digital textile and fashion industry, high-quality face retouching is also increasingly vital for virtual garment exhibition and e-commerce, where the aesthetic harmony between the model's skin and the textile textures significantly impacts user experience. The objective of face retouching is to reduce visual imperfections—such as acne, wrinkles, and uneven skin tone—while maintaining facial identity and realistic skin texture. Achieving this balance under real-time constraints remains a challenging problem.

Conventional face beautification methods are primarily based on handcrafted image processing techniques, such as Gaussian smoothing, bilateral filtering, and frequency separation [1-5]. Although computationally efficient, these methods often excessively suppress high-frequency details, leading to unnatural and over-smoothed skin appearances. The loss of fine textures, such as pores and subtle skin patterns, significantly degrades perceptual realism, especially when multiple beautification operations are applied.

Deep learning–based approaches, particularly generative adversarial networks (GANs), have demonstrated strong capability in modeling complex appearance transformations for face retouching. Image-to-image translation frameworks such as pix2pix [6] and CycleGAN [7] can learn data-driven mappings between unretouched and retouched faces, producing visually appealing results. However, most existing methods require high-resolution inference to preserve detail, which results in high computational cost and limits their applicability in real-time scenarios.

To address these limitations, we propose a real-time face retouching framework that introduces a neutral gray layer representation, allowing appearance adjustment to be handled independently from texture preservation. Rather than generating the retouched image directly, the network predicts a neutral gray layer that modulates the original image through soft-light blending. This representation allows the model to focus on perceptually relevant skin appearance changes while retaining original high-frequency texture details.

To further improve efficiency, inference is performed at a lower spatial resolution, followed by upsampling of the predicted gray layer to the original resolution. The final retouched output is obtained through a soft-light blending operation, which ensures smooth transitions and artifact-free reconstruction.

We evaluate the proposed method on the FFHQR dataset, using SSIM and PSNR as quantitative metrics. Experimental results demonstrate that our approach achieves high reconstruction fidelity while preserving fine skin textures. Moreover, GPU-based inference benchmarks confirm that the proposed framework supports real-time performance, making it suitable for deployment in practical face beautification systems.

In summary, the main contributions of this work are as follows:

Gray-layer–guided face retouching: We introduce a neutral gray layer representation for face beautification, allowing the model to enhance skin appearance while preserving fine texture details and facial realism.

Real-time GAN-based framework: By combining low-resolution inference with high-resolution reconstruction via upsampling and soft-light blending, the proposed approach significantly improves runtime efficiency without sacrificing visual quality.

Flexible retouching models: The proposed framework supports diverse face retouching scenarios under a unified architecture and training strategy, offering a robust and efficient tool for real-time human-centric visualization in digital fashion platforms.

This work demonstrates that integrating gray-layer representations with GAN-based image translation provides an effective and practical solution for real-time, high-quality face retouching, bridging the gap between visual realism and computational efficiency.

RELATED WORK

Traditional Face Beautification and Smoothing Methods

Early face beautification and retouching methods are mainly based on handcrafted image processing techniques, such as Gaussian filtering, bilateral filtering, guided filtering, and frequency separation [1-4]. These approaches aim to smooth skin regions while preserving strong edges by exploiting local intensity statistics.

Despite their efficiency, traditional methods lack semantic awareness of facial structures and skin imperfections. As a result, they often blur fine-grained textures such as pores and subtle wrinkles, leading to over-smoothed and artificial appearances [5]. Moreover, their performance heavily depends on manually designed parameters and skin masks, limiting robustness across diverse lighting conditions and facial variations.

Deep Generative Models for Face Retouching and Beautification

The advent of deep generative models, particularly GANs, has significantly advanced face retouching and beautification. A representative method, pix2pix [6], introduces conditional GANs for supervised image-to-image translation, while CycleGAN [7] enables unpaired translation via cycle-consistency constraints. These frameworks have been widely adopted for face-related enhancement tasks.

Building upon these foundations, several works explore GAN-based face beautification and retouching. AutoRetouch [8] proposes an automatic professional face retouching pipeline trained on paired data. With the assistance of the discriminator, the generator can distinguish between facial blemishes and normal skin features to generate more natural beautified faces. HQRetouch [9] introduces masked feature fusion and semantic-aware modulation to enhance retouching quality. ISFB-GAN [10] focuses on interpretable semantic face beautification by disentangling facial attributes.

Other studies investigate generalized portrait retouching using generative priors. StyleRetoucher [11] leverages GAN priors to enable style-consistent retouching across different portrait domains. These methods demonstrate strong visual quality but often require high-resolution inference and complex architectures, making real-time deployment challenging.

Face Retouching with Texture Preservation, Temporal Consistency, and Real-Time Constraints

Preserving fine skin texture while removing imperfections has become a central research focus in recent years. FabSoften [12] introduces dynamic skin smoothing with explicit wavelet-based texture restoration. BPFRe [13] proposes a blemish-aware progressive framework that separates coarse imperfection removal from fine detail refinement. ABPN [14] predicts the parts that need beautification in low-resolution images before processing the high-resolution original image, making it efficient enough to run on mobile devices.

Transformer-based approaches further improve retouching fidelity. RetouchFormer [15] employs prior-based selective self-attention for high-quality face retouching under limited supervision. For video scenarios, VRetouchEr [16] models cross-frame feature interdependence using imperfection flow to track the temporal motion of defects, achieving temporally consistent face retouching.

Parallel to retouching research, blind face restoration methods aim to recover realistic facial details from degraded images. GFP-GAN [17], GPEN [18], RestoreFormer, and RestoreFormer++ [19] leverage pretrained face GAN priors or key–value memory mechanisms to enhance texture realism.

In contrast to existing methods, our approach introduces a neutral gray layer representation that decouples appearance enhancement from texture preservation. Combined with low-resolution inference and high-resolution reconstruction via soft-light blending, the proposed framework achieves real-time performance while maintaining natural and realistic skin texture.

METHOD

Overview

We propose a real-time face retouching framework that enhances facial skin appearance while preserving fine-grained texture details. The overall pipeline is illustrated in Figure 1.

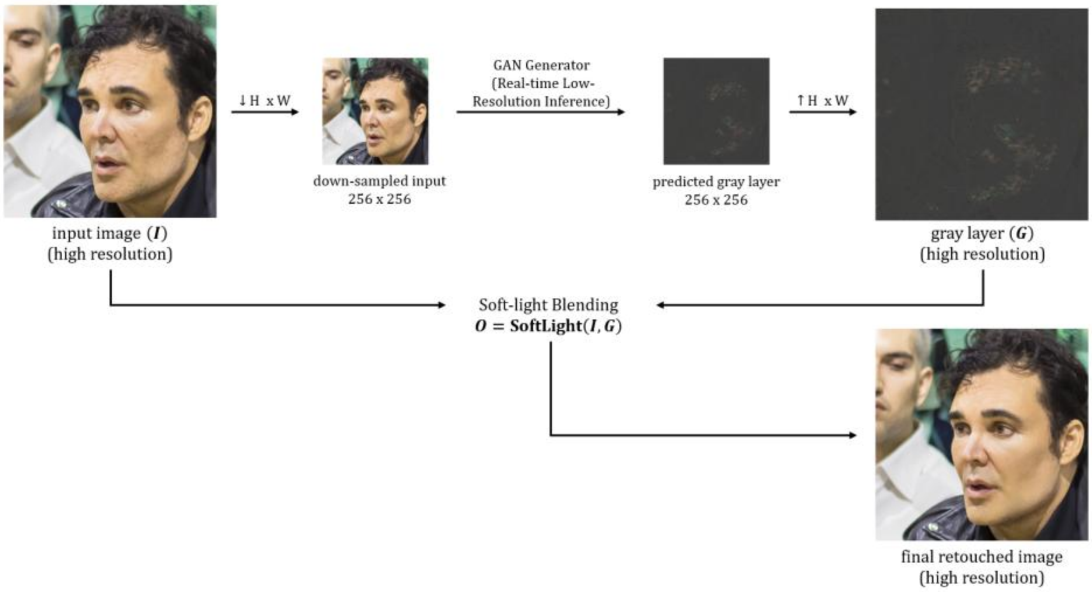

Figure 1. Overall framework of the proposed real-time face retouching system

The input image is first downsampled for efficient GAN inference. The network predicts a neutral gray layer, which is then upsampled and blended with the original high-resolution image using soft-light blending to produce the final retouched output.

Instead of directly predicting a fully retouched high-resolution image, our method introduces a neutral gray layer representation that shifts appearance adjustment away from explicit texture reconstruction. This design allows efficient low-resolution inference while maintaining natural and realistic high-resolution outputs.

Specifically, given an input facial image $\boldsymbol{I} \in \mathrm{R}^{H \times W \times 3}$, we first downsample it to a lower resolution $\boldsymbol{I}^{lr}$. A generator network predicts a neutral gray layer $\boldsymbol{G}^{lr}$, which is subsequently upsampled to the original resolution $\boldsymbol{G}$. The final retouched image is obtained by blending $\boldsymbol{G}$ with the original image using a soft-light blending strategy.

Neutral Gray Layer Representation

Traditional GAN-based retouching methods aim to learn a direct mapping from an input image to a retouched image. Such approaches require the network to jointly model color changes, illumination variations, and high-frequency texture details, making them computationally expensive and prone to texture degradation.

To alleviate this issue, we introduce a neutral gray layer representation inspired by professional photo retouching workflows. Instead of predicting the retouched image $\hat{\boldsymbol{I}}$, the network predicts a gray-scale modulation map $\boldsymbol{G} \in \mathrm{R}^{H \times W}$, which encodes local appearance adjustments while preserving the original image structure.

The neutral gray layer is defined as a three-channel image whose pixel values are centered around a mid-gray reference value g_0 (typically normalized to 0.5):

$$\boldsymbol{G}_{\mathrm{c}}(x,y) \in [0,1], \boldsymbol{G}_c(x,y) \approx g_0, \quad c \in \{\mathrm{R, G, B}\} \tag{1}$$

Intuitively, regions where $\boldsymbol{G}_{\mathrm{c}}(x,y) > g_0$ correspond to local brightening or smoothing effects, while regions where $\boldsymbol{G}_{\mathrm{c}}(x,y) < g_0$ indicate local darkening or detail suppression. Functionally, this representation operates as a luminance-conditioned residual modulation map. It is closely related to residual learning and spatial attention mechanisms, with the distinction that it is explicitly constrained to a neutral gray reference and applied through a fixed photometric blending operator. Since the gray layer varies smoothly across spatial locations, it can be reliably predicted at a lower resolution and later upsampled without introducing artifacts.

Soft-Light Blending for High-Resolution Reconstruction

To reconstruct the final retouched image, we blend the predicted gray layer with the original high-resolution image using a soft-light blending operation. Soft-light blending is widely used in image editing due to its ability to adjust local contrast while preserving underlying textures.

Let $\boldsymbol{I} \in [0,1]^{H \times W \times 3}$ denote the original image and $\boldsymbol{G} \in [0,1]^{H \times W \times 3}$ denote the upsampled gray layer. The output image $\boldsymbol{O}$ is computed pixel-wise as:

$$\boldsymbol{O} = \mathrm{SoftLight}(\boldsymbol{I}, \boldsymbol{G}) \tag{2}$$

Following the standard soft-light formulation, the blending operation is defined as:

$$\mathrm{SoftLight}(\boldsymbol{I}, \boldsymbol{G}) = \begin{cases} \boldsymbol{I} - (1-2\boldsymbol{G})\boldsymbol{I}(1-\boldsymbol{I}), \boldsymbol{G} < 0.5 \\ \boldsymbol{I} + (2\boldsymbol{G}-1)(\mathrm{f}(\boldsymbol{I})-\boldsymbol{I}), \boldsymbol{G} \geq 0.5 \end{cases} \tag{3}$$

where $f(\boldsymbol{I})=\sqrt{\boldsymbol{I}}$ is a nonlinear luminance adjustment function applied channel-wise.

This blending strategy ensures that:

1. High-frequency skin textures (e.g., pores) from the original image are preserved.
2. Retouching effects are applied in a perceptually smooth and spatially adaptive manner.
3. Upsampling artifacts from low-resolution inference are effectively suppressed.

From left to right: the original input image, the predicted neutral gray layer (contrast enhanced for visualization), and the final retouched result (Figure 2). The gray layer clearly captures the imperfection locations while leaving other regions neutral. The predicted three-channel gray layer is blended with the RGB image in a channel-wise manner, enabling independent modulation of each color channel. Since the gray layer is explicitly designed to model only low-frequency, spatially smooth appearance modulation (e.g., global tone and color balance), it can be safely predicted at a lower resolution and upsampled. High-frequency details and pixel-level structures are preserved by the original RGB image rather than encoded in the gray layer, ensuring spatial precision even under a 4× or higher upsampling factor.

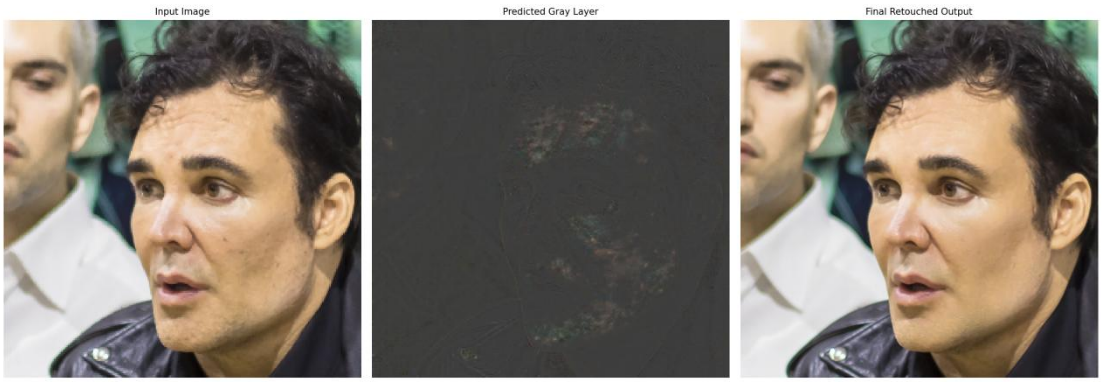

Figure 2. Visualization of the neutral gray layer and soft-light blending

Network Architecture

Our generator network is based on an image-to-image translation architecture derived from CycleGAN / pix2pix, implemented as a fully convolutional encoder–decoder with skip connections.

Given a low-resolution input image $\boldsymbol{I}^{lr} \in \mathrm{R}^{H\times W\times 3}$, the generator produces a gray layer prediction:

$$\boldsymbol{G}^{lr}=G\left(\boldsymbol{I}^{lr};\theta\right) \tag{4}$$

where G denotes the generator network parameterized by ϑ.

The discriminator D adopts a PatchGAN design, focusing on local structural consistency rather than global appearance. This encourages the generated gray layers to be locally smooth and visually coherent.

Training Objective

We formulate the training objective as a combination of adversarial loss and reconstruction loss.

Adversarial Loss

The adversarial loss encourages the generated gray layer to be indistinguishable from the ground-truth gray layer:

$$\mathrm{L}_{adv}=E_{\boldsymbol{G}_{gt}}\left[\log D\left(\boldsymbol{G}_{gt}\right)\right]+E_{\boldsymbol{I}^{lr}}\left[\log\left(1\text{-}D\left(G\left(\boldsymbol{I}^{lr}\right)\right)\right)\right] \tag{5}$$

Reconstruction Loss

To ensure pixel-wise consistency, we adopt an L1 loss between the predicted and ground-truth gray layers:

$$\mathrm{L}_{rec}=|\boldsymbol{G}^{lr}\text{-}\boldsymbol{G}^{lr}_{gt}|_1 \tag{6}$$

Full Objective

The final training objective is:

$$\mathrm{L}_{total}=\mathrm{L}_{adv}+\lambda\mathrm{L}_{rec} \tag{7}$$

where λ balances realism and accuracy.

EXPERIMENTS

Dataset and Experimental Setup

We conduct experiments on multiple paired face retouching datasets.

FFHQR is used as the primary dataset for quantitative evaluation and comparison with existing methods. It contains paired original and professionally retouched facial images and reflects holistic retouching effects.

In addition, separate paired datasets are used to train models focusing on specific retouching effects, including wrinkle reduction and skin smoothing. These datasets are used to demonstrate the applicability of the proposed framework under different retouching scenarios.

All datasets are split into training, validation, and test sets with a ratio of 8:1:1. Since FFHQR is derived from FFHQ, which contains only one image per unique identity, this split structure inherently guarantees no identity overlap across subsets. Issues regarding data leakage from varying poses or lighting of the same subject do not apply. During training, all images are resized to 256 × 256 for efficient low-resolution inference. During evaluation, the predicted gray layer is upsampled and blended with the original high-resolution image to produce the final retouched output, which is used for quantitative and qualitative assessment.

Evaluation Metrics

We quantitatively evaluate the retouching performance using the Peak Signal-to-Noise Ratio (PSNR) and the Structural Similarity Index Measure (SSIM).The PSNR is defined as:

$$\mathrm{PSNR}=10\log_{10}\left(\frac{\mathrm{MAX}^2}{\mathrm{MSE}}\right) \tag{8}$$

where MAX denotes the maximum possible pixel value of the image and MSE is the mean squared error between the retouched image and the ground-truth image.

SSIM evaluates perceptual similarity in terms of luminance, contrast, and structure.

Implementation Details

All models are implemented in PyTorch and trained on a single NVIDIA RTX 4090 GPU. Multiple independent models are trained under the same network architecture and optimization strategy, with different models obtained by training on different paired datasets.

Dataset and Preprocessing

All images are aligned and center-cropped following the FFHQ protocol. Input images are downsampled to 256×256 before being fed into the network. Ground-truth neutral gray layers are computed from paired original and retouched images.

Network Architecture

The generator adopts a U-Net–style encoder–decoder architecture derived from pix2pix, consisting of convolutional downsampling layers, residual blocks at the bottleneck, and symmetric upsampling layers with skip connections. The discriminator follows a PatchGAN design that focuses on local structural consistency.

Optimization Details

We use the Adam optimizer with $\beta_1 = 0.5$ and $\beta_2 = 0.999$. The initial learning rate is set to 2×10^{-4} for both the generator and discriminator. Models are trained for 200 epochs, with a constant learning rate for the first 100 epochs followed by linear decay. The batch size is set to 1.

Data Augmentation

Standard data augmentation techniques, including random horizontal flipping, random cropping, and mild color jittering, are applied during training.

Inference and Runtime Setup

During inference, only the generator is used. The FPS is measured for the core neural network inference stage, excluding image I/O and resolution restoration. The high-resolution soft-light blending is a lightweight pixel-wise operation that can be efficiently implemented on GPU.

Quantitative Results

Note that while RetouchFormer and VRetouchEr emphasize high-resolution attention modeling and video temporal consistency respectively, they share the common goal of face retouching. We include them to contextualize the trade-offs between computational efficiency and reconstruction fidelity. Our results highlight that the proposed method achieves competitive visual quality while satisfying the strict latency requirements of real-time applications (Table 1). All baseline results are obtained using the officially released implementations with default settings. Models trained for wrinkle reduction and skin smoothing are reported for reference only and are not directly compared with existing methods due to dataset differences.

Table 1. Quantitative results on FFHQR dataset

Method	PSNR ↑	SSIM ↑
Cross-method comparison		
Ours	38.49	0.9849
RetouchFormer [15]	36.08	0.9435
VRetouchEr [16]	30.77	0.9826
Ours with the same architecture and training strategy		
Ours (wrinkle reduction)	37.86	0.9844
Ours (skin smoothing)	36.74	0.9836

We compare the inference speed of our method with representative face retouching approaches, including the transformer-based RetouchFormer [15] and the video-based VRetouchEr [16]. All experiments are conducted on a single RTX 4090 GPU with batch size 1. FPS is measured for the core neural network inference stage, excluding image I/O and resolution restoration post-processing.

Our method achieves an inference speed of 185.69 FPS, which is significantly faster than RetouchFormer (4.46 FPS) and VRetouchEr (6.01 FPS), resulting in more than 10× speedup over existing transformer-based and video-based retouching methods. The high-resolution soft-light fusion used in our framework is a lightweight pixel-wise post-processing operation and can be efficiently implemented on GPU in real-world systems. These results demonstrate that our method is well suited for real-time face retouching applications.

Qualitative Results

From left to right: input image, RetouchFormer, VRetouchEr, our result, and the predicted neutral gray layer (Figure 3). Our method effectively removes skin imperfections while preserving natural skin texture and fine facial details. The gray layer visualizes the spatial distribution of retouching strength.

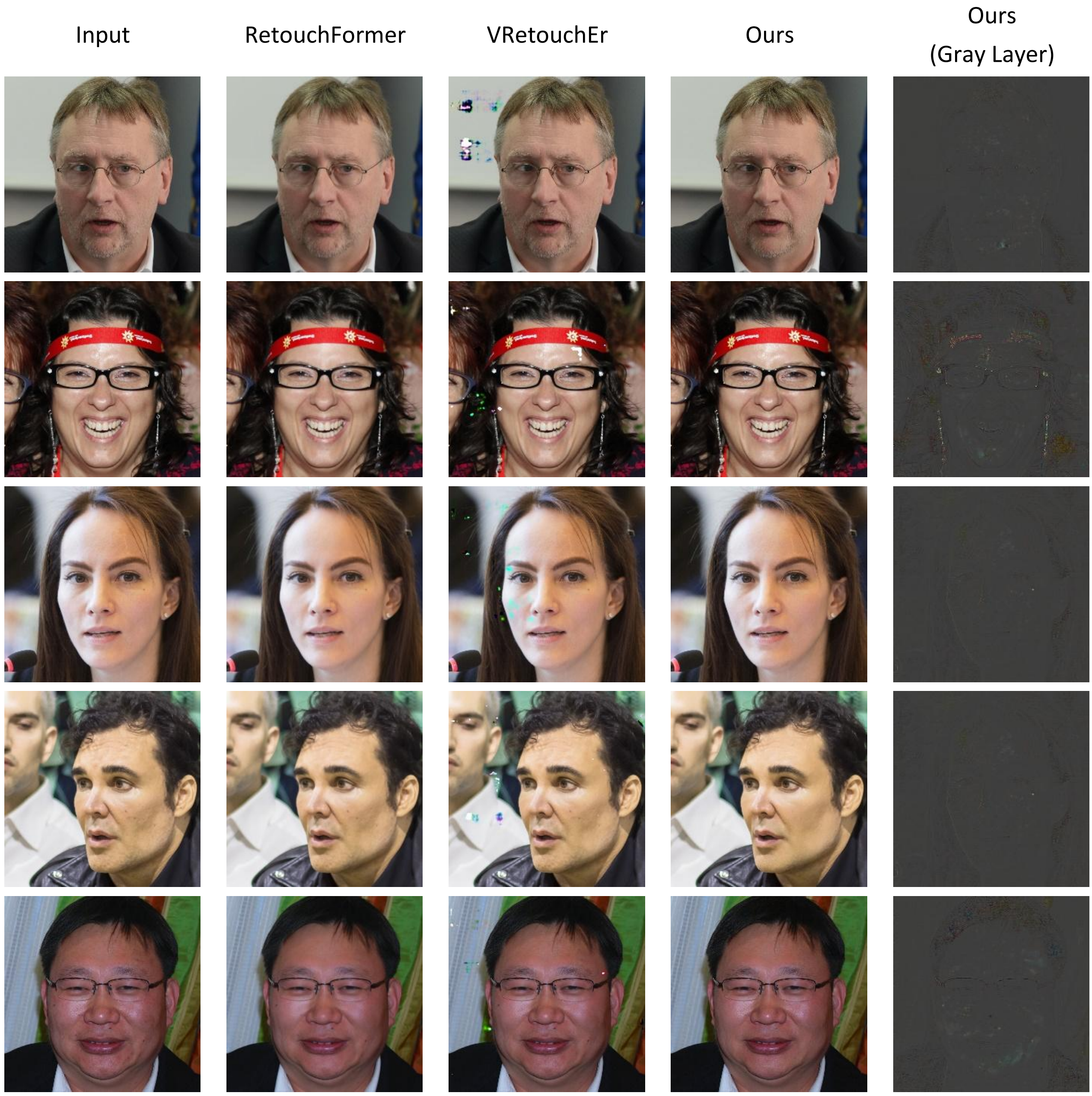

Figure 3. Qualitative comparison of face retouching results

DISCUSSION

The experimental results confirm that predicting a neutral gray layer instead of a full-color image allows appearance adjustment to be performed without explicitly modifying the underlying texture. Compared to traditional smoothing-based methods, our approach avoids over-smoothing artifacts and maintains realistic skin details. This capability is particularly relevant for the textile sector, as it ensures that facial aesthetics are enhanced without losing the authentic details required for high-fidelity virtual try-on displays. Furthermore, the low-resolution inference strategy significantly reduces computational cost, making the framework an ideal candidate for integration into intelligent fashion retailing and virtual reality (VR) textile environments.

CONCLUSION

We presented a real-time face retouching framework that integrates neutral gray layer prediction with GAN-based image translation. By performing low-resolution inference and reconstructing high-resolution outputs via soft-light blending, the proposed method achieves high visual quality while maintaining computational efficiency. Experiments on the FFHQR dataset demonstrate that our approach effectively enhances facial appearance without sacrificing texture realism, making it well-suited for real-world face beautification systems.

Author Contributions

Conceptualization – X.H. and Z.X; methodology – X.H. and Z.X; formal analysis – W.S.; investigation – S.X.; resources – W.S.; writing-original draft preparation – X.H. and Z.X; writing-review and editing – W.S. and S.X.; visualization – S.X.; supervision – W.S. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

Funding

This research was supported by the Shenzhen Science and Technology Program (Grant No. RCBS20221008093249079), the Shenzhen Polytechnic University-level Scientific Research Project (No. 6022310013K), and the Shenzhen Polytechnic High-level Talent Research Start-up Project (No. 6022310050k).

REFERENCES

[1] Tomasi C, Manduchi R. Bilateral filtering for gray and color images. Proceedings of the IEEE International Conference on Computer Vision; 23-25 June 1998; Santa Barnara, CA, USA. 1998; 1:839–846. doi: 10.1109/ICCV.1998.710815

[2] He K, Sun J, Tang X. Guided image filtering. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2013; 35(6):1397–1409. doi: 10.1109/TPAMI.2012.213

[3] Paris S, Kornprobst P, Tumblin J, Durand F. Bilateral filtering: Theory and applications. Foundations and Trends in Computer Graphics and Vision. 2009; 4(1):1–73. doi: 10.1561/0600000020

[4] Raskar R, Agrawal A, Tumblin J. Computational photography: Mastering new techniques. Computer. 2006; 39(8):30–38. doi: 10.1109/MC.2006.263

[5] Buades A, Coll B, Morel JM. A non-local algorithm for image denoising. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 20-25 June 2005; San Diego, CA, USA. 2005; 2:60–65. doi: 10.1109/CVPR.2005.38

[6] Isola P, Zhu JY, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 21-26 July 2017; Honolulu, HI, USA. 2017; 1:5967–5976. doi: 10.1109/CVPR.2017.632

[7] Zhu JY, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. Proceedings of the IEEE International Conference on Computer Vision (ICCV); 22-29 October 2017; Venice, Italy. 2017; 1:2223–2232. doi: 10.1109/ICCV.2017.244

[8] Shafaei A, Malik J, Ommer B. AutoRetouch: Automatic professional face retouching. Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV); 3-8 January 2021; Waikoloa, HI, USA. 2021; 1:990–998. doi: 10.1109/WACV48630.2021.00103

[9] Hong G, Wang F, Tian S, Lu M, Liu J, Zhang S. HQRetouch: Learning professional face retouching via masked feature fusion and semantic-aware modulation. Proceedings of the IEEE International Conference on Image Processing (ICIP); 8-11 October 2023; Kuala Lumpur, Malaysia. 2023; 1:2560–2564. doi: 10.1109/ICIP49359.2023.10222306

[10]Peng T, Li M, Chen F, Xu Y, Xie Y, Sun Y, et al. ISFB-GAN: Interpretable semantic face beautification with generative adversarial network. Expert Systems with Applications. 2024; 236:121131. doi: 10.1016/j.eswa.2023.121131

[11]Su W, Wang C, Liu C, Hanzhou, Fu H, Liao J. StyleRetoucher: Generalized portrait image retouching with GAN priors. IEEE Transactions on Visualization and Computer Graphics. 2024; 31(9):5089–5100. doi: 10.1109/TVCG.2024.3432910

[12]Xie L, Xue W, Xu Z, Wu S, Yu Z, Wong HS. Blemish-aware and progressive face retouching with limited paired data. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 17-24 June 2023; Vancouver, BC, Canada. 2023; 1:5599–5608. doi: 10.1109/CVPR52729.2023.00542

[13]Wang Z, Zhang J, Chen R, Wang W, Luo P. RestoreFormer: High-quality blind face restoration from undegraded key-value pairs. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 18-24 June 2022; New Orleans, LA, USA. 2022; 1:17512–17521. doi: 10.1109/CVPR52688.2022.01699

[14]Lei B, Guo X, Yang H, Cui M, Xie X, Huang D. ABPN: Adaptive blend pyramid network for real-time local retouching of ultra high-resolution photo. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 18-24 June 2022; New Orleans, LA, USA. 2022; 1:2108–2117. doi: 10.1109/CVPR52688.2022.00215

[15]Wen X, Xie L, Jiang L, Chen T, Wu S, Liu C, Wong H-S. RetouchFormer: Semi-supervised high-quality face retouching transformer with prior-based selective self-attention. Proceedings of the AAAI Conference on Artificial Intelligence; 20-27 February 2024; Vancouver, Canada. 2024; 38(6):5903–5911. doi: 10.1609/aaai.v38i6.28404

[16]Xue W, Jiang L, Xie L, Wu S, Xu Y, Wong HS. VRetouchEr: Learning cross-frame feature interdependence with imperfection flow for face retouching in videos. Proceedings of the IEEE/CVF Conference on

Computer Vision and Pattern Recognition; 16-22 June 2024; Seattle, WA, USA. 2024; 1:9141–9150. doi: 10.1109/CVPR52733.2024.00873

[17]Wang X, Li Y, Zhang H, Shan Y. Towards real-world blind face restoration with generative facial prior. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 19-25 June 2021; Nashville, TN, USA. 2021; 1:9168–9178. doi: 10.1109/CVPR46437.2021.00905

[18]Yang T, Liu Z, Chen Y, Zhao X, Li W, Guo Y. GAN prior embedded network for blind face restoration in the wild. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 19-25 June 2021; Nashville, TN, USA. 2021; 1:672–681. doi: 10.1109/CVPR46437.2021.00073

[19]Wang Z, Zhang J, Chen T, Wang W, Luo P. RestoreFormer++: Towards real-world blind face restoration from undegraded key-value pairs. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2023; 45(12):15462–15476. doi: 10.1109/TPAMI.2023.3315753