Sistemasi: Jurnal Sistem Informasi ISSN:2302-8149
Volume 15, Nomor 11, 2026: 4522-4537 e-ISSN:2540-9719

A Hybrid Method for Cartoonizing Faces with Pre-GAN Conditioning and Whale Optimization

[1]**Iman Subhi Mohammed***
[1]Computer Science Department, College of Computer Science and Mathematics, University of Mosul, Mosul, Iraq
*e-mail: imansubhi@uomosul.edu.iq

(***received*:** 11 October 2026, ***revised*:** 7 November 2026, ***accepted*:** 8 November 2026)

Abstract

While AI-powered facial cartoonization creates visually expressive portraits, standard generative models can sometimes render the faces with a variety of unappealing facial textures, inconsistencies in shading across facial regions, and a lack of fidelity in defining key facial features, which diminishes the overall quality and aesthetic appeal of cartoonized images. This paper proposes an adaptive facial beautification system for smoothing out the distracting details of the image in its original image, generating a more uniform shading for the cartoon, and preserving the necessary boundaries of the face. This framework involves bilateral spatial filtering, AnimeGAN2 stylization, skin-aware blending, adaptive luminance optimization with One-Dimensional Whale Optimization Algorithm (1D-WOA), and Secondary Edge Fusion (SEF). The reported evaluation is based on 1081 facial portrait images with 512×512 pixels. The first beautification step improved the skin-smoothness score from 0.6966 to 0.7031 and the normalized-edge contrast from 0.2796 to 0.1435. Edge contrast was further improved and edges were refined to 0.2667 (85.85% better than beautification), with the final skin-smoothness score still at 0.6975. The Peak Signal-to-Noise Ratio (PSNR) was 19.62 dB, whereas the AnimeGAN2 baseline was 18.86 dB. In the comparative evaluation, the proposed framework had the highest Spatial Frequency (SF = 101.31 ± 41.51) and Root Mean Square (RMS) Contrast (0.3515 ± 0.0711), which means that the spatial intensity variation and tonal contrast were the highest. The results show that adaptive skin refinement with structural enhancement can be a potential solution to enhance the cartoon portrait aesthetics. The framework can help digital artists and content creators in social media and digital communication applications.

Keywords: Cartoonization, Facial Portrait Stylization, Generative Adversarial Networks (GANs), Image Quality Assessment, Lightweight Metaheuristics, Whale Optimization Algorithm (WOA).

1 Introduction

Generative facial portrait stylization is a fundamental problem in computer graphics, computational photography and interactive digital media, and is the basis of synthesizing non-linear artistic transformations while retaining the biometric structures of the faces [1-4]. Thus, early non-parametric approaches were mainly based on spatial domain filtering and determinative color quantization [1,5] with no semantic learning from the data and no consideration of subtle artistic aspects [6]. With the recent progress in deep generative learning, Generative Adversarial Networks (GANs) [7] like CartoonGAN and AnimeGAN2 have been developed to learn high-level stylistic distributions, and Latent Diffusion Models (LDMs) like DiverseGANs and InstantID have also been developed. In spite of all these achievements, state-of-the-art generative pipelines face a perception-distortion trade-off [2]. Weighting stylistic saliency, as in [8-10] is often accompanied by non-linear checkerboarding artifacts and undesirable propagation of shadows along the boundaries of the faces [11,12]. On the other hand, it is difficult to obtain realistic results with reconstruction constraints applied at the pixel level [6].

Although there have been great developments in AI facial cartoonization, it is difficult to get the facial skin to be visually refined while keeping the facial features intact. Conventional smoothing methods can minimize the unwanted texture variations and distracting lines in the skin while obscuring important structural features and boundaries of the facial surface, whereas generative stylization can add or highlight these imperfections while preserving these boundaries and features of the skin. Many of the current enhancement methods do not offer an adaptive balance between facial beautification, smoothness, preservation of the edges, and low computation time. This restriction spurs the need to develop a lightweight refinement framework to enhance the quality of the cartoonized faces without further training

 ISSN:2302-8149
e-ISSN:2540-9719

of the generative model. The present study thus contributes to this research gap in three major ways. First, the adaptive facial beautification framework is presented, which combines bilateral spatial filtering, skin-aware refinement, One-Dimensional Whale Optimization Algorithm (1D-WOA)-based luminance adjustment and SEF. The components are intended to minimize visually objectionable changes in the skin texture and restore the important boundaries within the face that can be compromised during smoothing [13-15]. Second, a spatially constrained blending strategy is used to focus the refinement on the area of interest, to ensure the preservation of facial natural characteristics and the avoidance of unnecessary alterations in the neighboring areas and background. Thirdly, the proposed framework is compared with the quantitative and qualitative criteria of skin smoothness, structural detail, tonal contrast, perceptual quality, and computational performance. Note that the cosmetic skin refinement is structural preservation rather than assuming simultaneous improvement across all evaluation criteria.

The scientific achievements of this work are the following: Architecture-Adaptive Hybrid Topology: A post-refinement framework that is independent of generative backbone topology and is robust to the non-linear structural-perceptual trade-off. Localized Dual-Mask Blending Engine: A localized geometric masking protocol which removes boundary artifacts by limiting parametric changes to critical facial planes. Dynamic Continuous Optimization Engine (1D-WOA): A lightweight and continuous 1D-WOA post-optimiser for optimizing downstream exposure, with the aim of robust and artifact-free contrast rendering. Empirical Validation and Edge Feasibility: Tested on dataset of 1,081 portrait images, achieving an operational balance (PSNR = 19.62 dB, LPIPS = 0.2541) with validated deployment feasibility on resource-constrained edge hardware (NVIDIA Jetson Orin Nano at 11.45 W).

This paper is structured as follows: Section 2 summarizes the related works on generative facial stylization and metaheuristics; Section 3 covers the mathematical formulation and architecture of the proposed pipeline. Section 4 discusses extensive empirical benchmark results and ablation studies, as well as hardware profiling; and Section 5 concludes with remarks and future research directions.

2 Literature Review

Operational constraints that inspire this work are explored by discussing facial stylization and image enhancement paradigms in the context of deep generative networks, spatial pre-conditioning filters, and meta-heuristic parameter optimization Pipelines.

2.1 Generative Adversarial Networks for Facial Stylization

Early non-parametric approaches used anisotropic diffusion or structure-adaptive filtering to obtain visually flat images without compromising geometric discontinuities [1,16], but without any semantic knowledge. To include semantic intent, Chen et al. [17] introduced CartoonGAN with an edge-promoting loss, and Kim et al. [18] introduced U-GAT-IT with adaptive layer-instance normalization. But an optimal balance of abstraction and identity preservation is not an easy task [2]. The latest appraisals have established that generative models have semantic drift, causing micro-structural distortions [6] and loss of geometric fidelity [19]. To stabilize features, CBA-GAN was introduced in [8], localized neural attention mechanisms were introduced in [10] and privacy-preserving Expressive Cartoon Cards (ECC) [4] were introduced. Technical Limitation: Fine-tuning multi-million-parameter weights of deep generative models is prohibitively costly due to the need for re-training, and they also cause micro-structural distortions and shifts in identity.

2.2 Spatial Pre-Conditioning in Generative Contexts

In addition to neural weight changes, Tomasi and Manduchi were the first to apply bilateral filtering to structure-aware abstraction [16]. Long et al. [20] continued this in terms of sparse gradient enhancement to reduce noise and neural divergence. In addition, Alhamuy et al. [21] made use of localized Gaussian filtering on frequency bands to mask regions of interest to avoid global over-filtering that would destroy identity traits. Technical Limitation: Standard spatial pre-filtering is uniform or static smoothing, which causes over-smoothing and cannot preserve sharp edge contrast in the structure under non-uniform illumination.

2.3 Meta-Heuristic Parameter Optimization

Mirjalili and Lewis [22] proposed the Whale Optimization Algorithm (WOA) for automatic parameter selection, on which Braik [13] made an improvement by using color entropy and contrast maximization.

 ISSN:2302-8149
e-ISSN:2540-9719

Han et al. [23] proposed an advanced Whale Optimization Algorithm (WOA) for the adaptive optimization of image-enhancement parameters, demonstrating its effectiveness in improving objective image-quality metrics, and Wei et al. [24] used Retinex decomposition to decouple illumination and reflectance to show how meta-heuristic tuning can act after generative baselines. The Traditional Meta-Heuristic Swarms (MHS) are computationally complex, prone to local minimum and experience too much runtime latency while searching in a high-dimensional vector space.

3. Methodology

This section offers some insight into the design of the proposed hybrid optimization framework and the resultant mathematical formulation.

3.1 System Overview and Input Representation

We propose a framework that balances identity preservation and aesthetic enhancement in cartoonizing facial portraits (Fig. 1). The input dataset consists of 1081 high-resolution facial images across ethnicities, ages, and lighting conditions. To avoid semantic drift and structural distortions due to differences in lighting, input portraits Iin are aligned, quality-filtered, and normalized to 512x512 pixels. To ensure experimental reproducibility, we summarise the system execution parameters in **Table 1**. The active pre-filtering parameters and loss weights adaptively move based on the generative core architecture GAN [25].

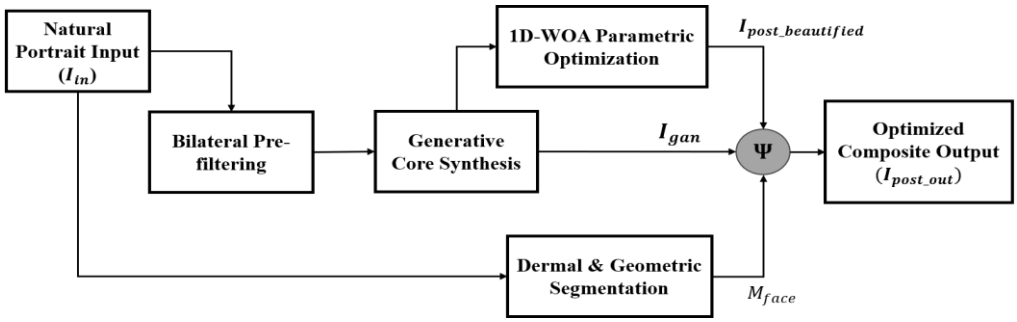


Fig. 1. Schematic architecture of the proposed hybrid framework.

3.2 Bilateral Spatial Pre-filtering

An anisotropic bilateral pre-filter is used to suppress high-frequency micro-artifacts [26] and to stabilize the downstream latent representations of I_{in}. Spatial luminance of the dermal region is smoothed, and high-frequency boundary information is preserved. The output state, at pixel coordinate *x*, of the pre-filter, $I_{pre_filtered}(x)$, is given by [27]:

$$I_{pre_filtered}(x) = \frac{1}{W_p} \sum_{y \in \Omega} g_s(y-x).\, g_r\left(\mathrm{I_{in}}(y) - \mathrm{I_{in}}(x)\right).\, I_{in}(y) \tag{1}$$

$$W_p = \sum_{y \in \Omega} g_s(y-x).\, g_r(\mathrm{I_{in}}(y) - \mathrm{I_{in}}(x)) \tag{2}$$

where *x* is the location of the pixel being filtered, y represents the neighbors in Ω, and W_p is the spatial normalization factor that normalises the sum of the spatial and intensity-dependent Gaussian weights at the output of the normalized filter at *x*. The indicated 30×30 domain defines the filter's local neighborhood. $I_{pre_filtered}(x)$ is the $I_{pre_filtered}$ pixel value at *x*, *I*_in is the raw input image tensor, $y \in \Omega$ represents neighboring coordinates in the spatial domain (Ω = 30×30), g_s and g_r are the spatial and range Gaussian kernels [13] defined as:

$$I_{pre_out} = I_{in} \odot \Lambda + Ipre_filtered \odot (1 - \Lambda) \tag{3}$$

where I_{pre_out} denotes the spatially blended pre-conditioning output. Λ is a binary logical boundary clipping matrix derived from localized spatial scale-space constraints to prevent global over-smoothing.

 ISSN:2302-8149
e-ISSN:2540-9719

Eq. (3) retains the original input wherever Λ = 1 and uses the bilateral-filtered output wherever Λ = 0; the complementary factor (1 − Λ) selects the latter region. The $pre_filtered$ tensor is represented by $Ipre_filtered$ and the Hadamard (element-wise) product is denoted by ⊙ [20].

3.3 Generative Core Synthesis

The pre-conditioned structural matrix I_{pre_out} is fed into the generative core to generate artistic stylization, which yields flat texture blocks and the stylized boundaries [17, 18, 28]:

$$I_{gan} = \mathcal{G}_{GAN}\left(I_{pre_out}\right) \tag{4}$$

where I_{gan} is the synthesized output image tensor, $\mathcal{G}_{GAN}$ denotes the generative core synthesis mapping operator, and I_{pre_out} is the pre-conditioned structural input matrix.

3.4 Dermal and Geometric Segmentation

To avoid cross-channel chromatic bleeding in sensitive facial regions, spatial sub-masks are extracted from the input portrait. Specifically, to isolate dental structures without interference from skin chrominance, the synthesized RGB image tensor I_{gan} is temporarily mapped into the HSV (Hue-Saturation-Value) color space, yielding decoupled luminance (V_{gan}) and saturation (S_{gan}) channels. Using these channels, a binary structural sub-mask M_{teeth} is derived via HSV space thresholding [29]. The numbers 1 and 0 in Eq. (5) indicate the inclusion or exclusion of a pixel in the dental sub-mask; the thresholds τV and τS are for HSV value and HSV saturation, respectively:

$$M_{teeth} = \begin{cases} 1 & if\ V_{gan}(x) > \tau_V \wedge S_{gan}(x) < \tau_S \\ 0 & otherwise \end{cases} \tag{5}$$

The image is synthesized by the XOR operation of M_{teeth} (binary structural sub-mask isolating dental regions) and V_{gan} (Value (luminance) component) and the Saturation component (S_{gan}) of the synthesized image in the HSV color space at the pixel coordinate x. $C_r(x)$ and $C_b(x)$ are the chrominance parts of pixels in position *x* in Eq. (6). A value of 1 indicates that a pixel is in both of the given chrominance ranges, and a value of *0* indicates that a pixel is not in both of the given chrominance ranges.

$$M_{skin}(x) = \begin{cases} 1, & if\ C_r(x) \in [133, 173] \wedge C_b(x) \in [77, 127] \\ 0, & otherwise \end{cases} \tag{6}$$

In Eq. (7), (*xc*, *yc*) denotes the centre of the ellipse, and (*x*, *y*) is the point that is being tested against the ellipse. Mask value "1" means a location within or on the ellipse; "0" means a location outside of the ellipse. Simultaneously, to construct the automated dermal mask M_{skin} [30, 31], the input portrait is converted into the illumination-invariant YCrCb color space. Dermal pixels are segmented by constraining chrominance components within empirical skin color bounds ($C_r \in$ [133, 173] and $C_b \in$ [77, 127]) as detailed in **Table 1**. The resulting mask M_{skin} is then combined with an elliptical facial geometry mask $M_{ellipse}$ [32] centred at centroid coordinates (*xc*, *yc*):

$$M_{ellipse}(x, y) = \begin{cases} 1, & if\ \dfrac{(x - x_c)^2}{a_{ellipse}^2} + \dfrac{(y - y_c)^2}{b_{ellipse}^2} \leq 1 \\ 0, & otherwise \end{cases} \tag{7}$$

With $a_{ellipse}$ and $b_{ellipse}$ being the semi-major axis and semi-minor axis, respectively. The composite facial mask M_{face} is generated via logical matrix intersection under illumination-invariant constraints, where M_{face} isolates the facial region of interest (ROI) while suppressing background content [33]:

 ISSN:2302-8149
e-ISSN:2540-9719

$$M_{face} = M_{skin} \odot M_{ellipse} \quad (8)$$

3.5 Pipeline-Adaptive 1D-WOA Parametric Optimization

A meta-heuristic optimization loop is placed downstream from the generative core following the hybrid modelling principles [34] to refine the features without the need for re-training the entire network [35]. Before parametric adjustment, I_{gan} is transformed from the native RGB grid into the perceptually uniform CIELAB color space, separating spatial luminance (L_{gan}) from chromatic components (a_{gan}, b_{gan}). Luminance scaling is applied exclusively to the L_{gan}channel via a continuous scalar gamma parameter:

$$I_{post_beautified} = 255 \cdot \left(\frac{I_{gan}}{255}\right)^{\gamma} \quad (9)$$

where $I_{post_beautified}$ is the gamma-adjusted output image tensor, I_{gan} is the generative output, and γ is the adaptive gamma scaling parameter optimized within [0.1, 2.5].

3.5.1 Dimensionality Reduction and Agent Mechanics

The search space is projected onto a 1-dimensional continuous manifold to reduce the computational latency [14] and local optima stagnation [36]. Based on the Whale Optimization Algorithm (WOA) [21], the computational complexity is reduced from O(D · N · T) to O(N · T) with N = 5 and T = 20, which is known as deterministic. Encircling and helical spiral mechanisms [21,37] are used to develop the candidate positions for search agent *i* on iteration *t+1*. In Eqs. (10)-(12), $\Gamma_i(t)$ represents the scalar parameter value that is being evaluated by the search agent *i* at the t-th iteration; *Γi (t+1)* represents the new value of the parameter; and *T* represents the maximum number of iterations for the optimization (20, as mentioned in the experimental settings).

$$\Gamma_i(t+1) = \begin{cases} \gamma^*(t) - A \cdot D & if\ p < 0.5 \\ D' \cdot e^{b \cdot l} \cdot cos(2\pi\, l) + \gamma^*(t) & if p \geq 0.5 \end{cases} \quad (10)$$

$$D = |C \cdot \gamma^*(t) - \Gamma_i(t)|\,, D' = |\gamma^*(t) - \Gamma_i(t)| \quad (11)$$

$$A = 2a \cdot r - a, C = 2 \cdot r, \qquad a = 2 - t \cdot \left(\frac{2}{T}\right) \quad (12)$$

Where $\Gamma_i(t+1)$is the updated parameter candidate of the *i-th* search agent at the *t+1-th* iteration, *γ*(t)* is the current global optimum parameter, *D* and *D'* are spatial distance measures for the exploration and helical exploitation, respectively, *A* and *C* are coefficient scalars, b is the constant of the logarithmic spiral shape (*b* = 1.0), ***l*** is a random value in the range of [-1, 1], and ***p*** is a random probability in the range of [0, 1] that is used for choosing between encircling and spiral [21]. *A* and *C* are chosen uniformly from the interval [0, 1], and ***a*** is linearly decayed.

3.5.2 Continuous Tri-Objective Fitness Function

The candidate parameter γ is evaluated by minimizing a tri-objective fitness function that balances contrast enhancement, edge preservation, and chromatic clipping constraints [38,39]:

$$f_{fitness}(\gamma) = \frac{1}{w_1 \cdot S_{contrast}(\gamma) + w_2 \cdot E_{edge}(\gamma) + \epsilon} w_3 + A_{chroma}(\gamma) \quad (13)$$

$$S_{contrast}(\gamma) = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(I_{post_beautified,i}(\gamma) - \mu\right)^2} \quad (13a)$$

 ISSN:2302-8149
e-ISSN:2540-9719

$$E_{edge}(\gamma) = \frac{\sum_i |\nabla I_{post_beautified,i}(\gamma)|}{\sum_i |\nabla I_{in,i}| + \epsilon} \quad (13b)$$

$$A_{chroma}(\gamma) = \frac{1}{N}\sum_{i=1}^{N} max\left(0, I_{post_beautified,i}(\gamma) - 255\right) \quad (13c)$$

where $f_{fitness}(\gamma)$is the raw fitness value for candidate parameter γ; $S_{contrast}$, E_{edge}, and A_{chroma} denote the contrast variance, edge preservation ratio, and saturation penalty metrics; $I_{post_beautified,i}$ and $I_{in,i}$ are enhanced and input pixel values at index *i*; *N* is total pixel count; μ is mean luminance; ∇ is the Sobel operator; *w*1 = 0.50, *w*2 = 0.30, and *w*3 = 0.20 are empirical weighting factors; and ϵ is a small stabilization constant. To enable standardized cross-evaluation, raw fitness scores are mapped to a bounded range [0, 1] via linear scaling:

$$f_{norm}(\gamma) = \frac{f_{fitness}(\gamma)}{F_{max-global}} \quad (13d)$$

where $f_{norm}(\gamma)$denotes the rescaled normalized objective fitness score mapped to the bounded interval [0, 1], $f_{fitness}(\gamma)$represents the raw unscaled tri-objective fitness value for candidate gamma parameter γ, and $F_{max-global}$is the empirical maximum raw fitness score ($F_{max-global}$= 48.8596), which was determined during initial global parameter profiling by evaluating peak raw fitness across all *N* = 1,081 dataset samples to establish a fixed anchor and maintain relative gradient sensitivity during evaluation.

3.6 Multi-Stream Spatial Fusion (Ψ) and Final Composite Output

The enhanced luminance layer $I_{post_beautified}$, original generative output I_{gan}, and composite facial mask M_{face} are integrated using a multi-stream spatial blending Ψ [30,40]:

$$I_{post_{out}}(x) = M_{face}(\mathrm{x})[1 - \lambda E_{SEF}(\mathrm{x})]I_{ref}(\mathrm{x}) + [1 - \lambda M_{face}(\mathrm{x})E_{SEF}(\mathrm{x})]\ I_{gan}(x) \quad (14a)$$

$$I_{ref}(\mathrm{x}) = \Psi'\left(\left(I_{post_beautified}, a_{gan}, b_{gan}\right) + \left(1 - M_{face}(x)\right)\right) \quad (14b)$$

In this case, I_{post_out} is the RGB output after compositing. $I_{ref}(\mathrm{x})$ is the refined cartoon image (after adaptive luminance and contrast adjustment) subjected to $M_{face}(x)$ to obtain $\left(1 - M_{face}(x)\right)$. $I_{gan}(x)$ is the original cartoon image created using AnimeGAN2, $M_{face}(x)$ is the spatial mask provided by Eq. (8), $E_{SEF}(\mathrm{x}) \in [0,1]$ is the normalized structural edge-weighting map to emphasize selected facial boundaries and $\lambda \in [0,1]$ is the edge-fusion coefficient to determine the degree of contribution from the generative output to the structural information. Ψ' denotes the inverse color space transformation operator mapping the recombined CIELAB components ($I_{post_beautified}, a_{gan}, b_{gan}$) back into the native RGB color space. This formulation regulates the influence of the initial generative output at specific structural boundaries while preserving the cartoony representation in softer facial areas. Thus, the fusion is designed to promote the refining of the skin and the maintenance of valuable facial contours while avoiding an indiscriminate emphasis on unwanted variations in texture. [40].

3.7 Experimental Setup and Reproducibility Protocol

The pretrained AnimeGAN2 model [41, 42] was used to cartoonize faces, followed by the proposed One-Dimensional Whale Optimization Algorithm (1D-WOA) based refinement framework, which improves facial aesthetics while maintaining significant facial structure details through adaptive luminance optimization and skin refinement. Further, the Enhanced Pixel Integration (EPI) method [43] was applied as a baseline technique for comparing the proposed work with existing enhancement techniques. The parameters in digital form as detailed in Table 1. The provided methodology is for generative inference and not network retraining. PyTorch inputs were scaled to [−1, 1]. The processing procedure in this paper includes bilateral spatial pre-filtering, generative synthesis, facial mask construction, luminance adjustment, and multi-stream fusion, as shown in Sections 3.1 – 3.5. The 1D-WOA outperformed a single continuous luminance parameter γ in the range [0.1, 2.5], employing five search agents and a maximum of

 ISSN:2302-8149
e-ISSN:2540-9719

20 iterations. The coefficient of the reported log-spiral is b = 1.0. The fitness function used for the assessment of candidate solutions is the tri-objective fitness function (Eq. (13)) with w1 = 0.50, w2 = 0.30, and w3 = 0.20. The fitness terms and normalization procedure are described in Section 3.5.2, which also includes the empirical normalization reference of 48.8596. The spatial pre-filtering configuration gives a neighbourhood size of 30 × 30 and a colour and spatial sigma of 75.0. The reported intervals for Cr of YCrCb-based dermal masking were [133, 173], and for *Cb* were [77, 127]. The matching module settings are summarized in Table 1, including the EPI baseline setting, which is not a 1D-WOA setting. The reported computational environment consisted of an NVIDIA T4 GPU (16 GB GDDR6) and an Intel Xeon CPU. Both NumPy and PyTorch random seeds were initialized with 42, and they both used the same random seed to ensure reproducible stochastic initialization under similar execution conditions, but note that a fixed seed does not guarantee the same results across software or hardware implementations. The model comparisons were made for the given computational conditions and also rely on the image-selection protocol used for the original software environment, which should be described separately from the fixed numerical conditions.

Table 1 Complete numerical specification of implementation parameters across processing stages and baselines

Processing Module	Theoretical Symbol	Core Parameter Description	Numerical Settings / Value
Adaptive 1D-WOA	N	Population Size	5
	t_{max}	Maximum Iterations	20
	Γ	Search Bound Range	[0.1, 2.5]
	w1, w2, w3	Fitness Function Weights	(Contrast-w1=0.50), (Edges-w2=0.30), (Exposure- w3= 0.20)
	$Seed_{opt}$	Optimization Random Seed	42
Spatial Pre-Filtering	Ω	Local Neighborhood Window	$30 \times 30\ (GAN)$
	σ_{color}	Bilateral Color Space Sigma	75.0
	σ_{space}	Bilateral Coordinate Space Sigma	75.0
Skin Segmentation	τ_S, τ_V, Λ	Skin Color Bounds (tableYCrCb)	$Cr \in [133, 173], Cb \in [77, 127]$
Pipeline-Adaptive Blending	$base_weight$	Original Image Retention Weight	0.25 (AnimeGAN2)
EPI Post-Processing	T	Color Difference Threshold	50
	$Wcomposite$	Global Composite Blending Ratio	[0.5, 0.7]
	Loss Weights	Loss Component Scaling (adv, percept, pixel)	α = 1.0, β = 0.5, γ = 0.2

3.8 Quantitative Evaluation Metrics

Framework performance is evaluated using standard structural fidelity, perceptual similarity, edge clarity, dermal smoothness, and color metrics:

1) **Peak Signal-to-Noise Ratio (PSNR):** Measures spatial structure preservation relative to the input I_{in} [44], where *M*, *N*, *C* denote image height, width, and color channels, and MAX*I* = 255. In Eqs. (15) and (16), I_{in} (*i*,*j*,*k*) and I_{post_out} (*i*,*j*,*k*) are the input and final output intensities for each of the $M \times N \times C$ samples of the image, respectively; MSE is the mean squared error over all $M \times N \times C$ image samples.

$$MSE = \frac{1}{M \times N \times C} \sum_{i=1}^{M} \sum_{j=1}^{N} \sum_{k=1}^{C} \left(I_{in}(i,j,k) - I_{post_out}(i,j,k) \right)^2 \quad (15)$$

$$PSNR = 10 \cdot log_{10} \left(\frac{MAX_I^2}{MSE} \right) \quad (16)$$

 ISSN:2302-8149
e-ISSN:2540-9719

2) Learned Perceptual Image Patch Similarity (LPIPS): Evaluates high-level feature divergence using a VGG-16 backbone [45], where $\hat{y}^l_{hw}$ and y^l_{hw} are normalized activations of the channels at layer l and W_l are the scales with the layer weights. The mean scores reported are the mean scores of the dataset. In Eq. (17), the perceptual feature layers are indexed by *l*, the position of a feature in the spatial feature map is indexed by *h* and *w*, the dimensions of the spatial feature map for the ith perceptual feature layer are H_l and W_l, and w_l is the channel weighting for the perceptual feature layer. The two feature activations are for the source image and processed output.

$$d(I_{in}, I_{post_out}) = \sum_l \frac{1}{H_l . W_l} \sum_{h,w} ||w_l \odot (\hat{y}^l_{hw} - y^l_{hw})||_2^2 \quad (17)$$

3) Normalized Edge Contrast Index (ECnorm): Quantifies boundary sharpness via spatial Sobel gradients *Gx*, *Gy* [14], where $EC_{norm} \in [0, 1]$. Comparative scores evaluate $M_{hybrid} = EC_{norm}(I_{post_out})$ against the baseline score $M_{bench} = EC_{norm}(I_{base})$. The Sobel-gradient components along the horizontal and vertical directions are G_x and G_y, respectively, and Eqs. (18)-(19) are the gradient magnitude at position *(i,j)*. The dimensions *M* and *N* are the same as their meanings in Eq. (15), image height and image width, respectively.

$$G(i,j) = \sqrt{(G_x(i,j)^2 + G_y(i,j)^2)} \quad (18)$$

$$EC_{norm}(I) = \frac{1}{M.N.255} \sum_{i=1}^{M} \sum_{j=1}^{N} G(i,j) \quad (19)$$

4) Skin Smoothness Index ($S_{smoothness}$): Evaluates dermal noise reduction within the mask M_{skin} [46, 47], The sum in Eq. (20) is over the pixels in the specified 5 × 5 neighborhood that are in N_{skin}, the number of such pixels, and $\sigma^2_{local}(i,j)$ is the local intensity variance calculated in the indicated 5 × 5 neighborhood, and ϵ is a term to prevent division by zero.

$$S_{smoothness} = \frac{1}{\left(\frac{1}{N_{skin}} \sum_{(i,j) \in M_{skin}} \sigma^2_{local}(i,j)\right) + \epsilon} \quad (20)$$

5) Colorfulness Index (C_{index}): Measures color vibrancy in opponent color space ($rg = R - G$, $yb = 0.5(R + G) - B$) [48], Eqs. (21)-(23): The image color channels are labelled R, G, and B; the two opponent-color components are labeled the standard deviations: σ_{rg} and σ_{yb} and the means: μ_{rg} and μ_{yb}. The result is the combination of these measures, as shown in Eq. (23), termed the C_{index}.

$$\sigma_{rgyb} = \sqrt{(\sigma^2_{rg} + \sigma^2_{yb})} \quad (21)$$

$$\mu_{rgyb} = \sqrt{(\mu^2_{rg} + \mu^2_{yb})} \quad (22)$$

$$C_{index} = \sigma_{rgyb} + 0.3 \cdot \mu_{rgyb} \quad (23)$$

4. Results and Discussion

1,081 portraits are evaluated quantitatively, qualitatively, and statistically using the proposed framework and the baselines.

4.1 Qualitative Visual Synthesis and Perception-Distortion Dynamics

The qualitative comparison between the baseline generative model (AnimeGAN2) and the proposed hybrid framework is presented in Fig. 2. The baseline output exhibits undesirable facial texture variations,

 ISSN:2302-8149
e-ISSN:2540-9719

uneven shading in skin regions, and reduced definition of important facial boundaries, which may affect the aesthetic quality of the cartoonized portrait. In contrast, the proposed framework aims to improve facial appearance by smoothing visually distracting skin details and producing more uniform cartoon-style shading while maintaining the clarity of essential facial features and structural boundaries. These visual differences illustrate the intended balance between facial beautification, skin refinement, and structural detail preservation.



(a) (b) (c)

Fig. 2. Qualitative stylization comparison:(a) Original input image with region of interest (ROI), (b) Baseline AnimeGAN2 output, and (c) Proposed hybrid framework output.

4.2 Quantitative End-to-End System Performance and Edge Preservation

Fig. 3 shows the individual necessity of each module of the framework by showing the ablations. The post-optimization is not spatial-level pre-conditioned, which is not effective in enhancing chromatic vibrancy. Isolated spatial pre-filtering, on the other hand, causes global smoothing, which reduces the sharpness of the boundaries. The perception-distortion dilemma is solved by integrating adaptive spatial pre-conditioning, post-generation 1D-WOA gamma optimization and SEF [49]. This single pipeline dynamically balances pre-conditioning and post-generation refinement, which meet perceptual needs and minimize distortion in the structure.



Fig. 3. Visual ablation comparison against baseline.

4.2.1 Perception-Distortion Trade-Offs, Evaluation Protocols, and Latency Profiling

The quantitative outcome of the experiments is shown in Table 2, in which a compromise between facial beautification, smoothing of skin texture and preservation of structural detail can be seen. The PSNR was increased from 18.86 dB to 24.12 dB with the bilateral pre-filtering stage used at the beginning (Beautified AnimeGAN2), and the smoothness of skin was improved from 0.6966 to 0.7031. The bilateral pre-filtering stage was used at the beginning, which improved the PSNR value from 18.86 dB to 24.12 dB and the smoothness of skin from 0.6966 to 0.7031, and reduced the LPIPS value from 0.2555 to 0.2420. This smoothing came at a price, however, of a significant loss in normalized edge contrast from 0.2796 to

 ISSN:2302-8149
e-ISSN:2540-9719

0.1435, meaning that some of the facial boundary definition was lost. To find a balance between these two demands, the proposed end-to-end framework integrates adaptive 1D-WOA optimization and Secondary Edge Fusion (SEF) to regain edge contrast of 0.2667, which is 85.85% better than the pre-filtered stage. The final results preserved a skin smoothness measure of 0.6975, which is just a little bit worse than the skin smoothness measure of the AnimeGAN2 baseline, and an RMS-independent colorfulness measure of 0.4099, with a PSNR of 19.62 dB. These are not the best values on each of the metrics, but reflect a more balanced result when it comes to cosmetic smoothing versus structural enhancement. This is the same behavior that is desired when using the framework to improve the appearance of a face by minimizing loss of important visual information.

Table 2 Quantitative evaluation of full end-to-end processing stages across the 1,081-image dataset.

Stage	PSNR (dB) ↑	LPIPS ↓	Edge Contrast ↑	Skin Smoothness ↑	Colorfulness ↑
AnimeGAN2	18.86	0.2555	0.2796	0.6966	0.4152
Beautified AnimeGAN2	24.12	0.2420	0.1435	0.7031	0.4014
End-to-end Proposed Hybrid (T = 20)	19.62	0.2541	0.2667	0.6975	0.4099

Note: All Edge Contrast values are normalized to the [0, 1] dynamic range. Results reflect the full end-to-end pipeline including bilateral pre-filtering, dynamic 1D-WOA gamma optimization (T = 20), and SEF.

Mathematically, relative to the edge contrast attenuated during pre-filtering (0.2796 - 0.1435 = 0.1361), this stage recovers 90.52% of the lost boundary contrast, effectively retaining 95.39% of the original baseline structural contrast while remaining 4.61% below it to maintain smooth dermal rendering (0.6975). Box plots in Fig. 4 validate low statistical variance across the 1,081 evaluated samples, confirming system robustness across diverse illumination and facial distributions.

To ensure analytical transparency across processing paradigms, two standardized evaluation protocols are established: **Stage-wise Ablation Analyses (Tables 2 and 3):** Metrics are measured relative to the AnimeGAN2 model output to isolate incremental performance gains contributed specifically by the 1D-WOA Pipeline-Adaptive and the SEF module. **Exhaustive Benchmarking (Table 4):** Full-reference metrics (PSNR, LPIPS, Entropy) are computed directly against raw input portraits (N=1,081) for rigorous comparison against non-generative filtering algorithms (e.g., MSR, CLAHE, Retinex).

To operationalize global normalization via Eq. (13d), empirical distribution scanning across N=1,081 test samples established global raw fitness bounds of Fmin-global = 13.2729, Fmean-global = 29.4913, and Fmax-global = 48.8596. Substituting Fmax-global = 48.8596 into Eq. (13d) rescales fitness scores in Tables 3 and 4 to the bounded [0, 1] domain. Finally, execution latency is decoupled into Core Optimization Time (259.6 ms for 1D-WOA mathematical convergence at T=20) and Total Module Overhead (969.1 ms for full end-to-end processing, encompassing matrix I/O, bilateral pre-filtering, dynamic fitness evaluation, and multi-stream spatial edge fusion).

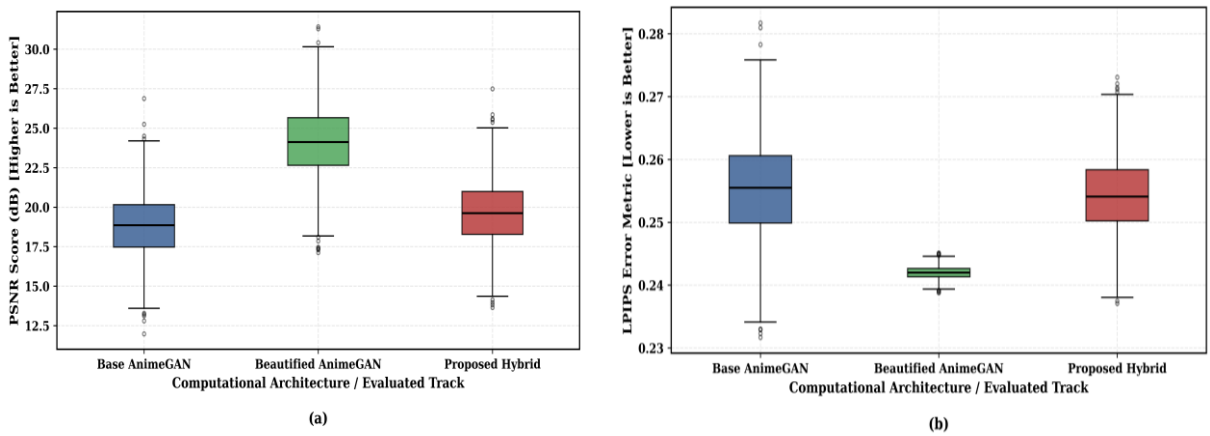


Fig. 4. Statistical metric profiles across processing pipelines:
(a) PSNR score distributions, and (b) LPIPS error metric distributions.

 ISSN:2302-8149
e-ISSN:2540-9719

4.2.2 Convergence Kinetics and Parametric Sensitivity Analysis

Convergence kinetics and parametric sensitivity analyses of the 1D-WOA metaheuristic across maximum iteration thresholds are presented in Table 3 and Fig. 5. During early exploration (T = 2 to T = 5), rapid convergence is observed, with mean normalized fitness increasing from 0.9618 to 0.9883 and PSNR improving from 19.54 dB to 19.60 dB. Beyond T = 10, the framework exhibits marginal convergence, maintaining metric fluctuations within tight numerical variance tolerances (e.g., LPIPS stabilizes at 0.2541 as fitness approaches 1.0000). Practical global convergence is achieved at T = 20, yielding a normalized fitness of 1.0000, PSNR of 19.62 dB, and peak colorfulness of 0.4099.

Table 3 Ablation and dynamic convergence analysis of 1D-WOA across iterations.

T	Mean γ* ± SD	Mean Normalized Fitness(F_{norm}) ↑	Mean PSNR (dB) ↑	Mean LPIPS ↓	Mean Edge Contrast ↑	Mean Skin Smoothness ↑	Mean Colorfulness ↑	Core Optimization Time (ms) ↓
2	1.616 ± 0.408	0.9618	19.54	0.2542	0.2760	0.6973	0.4072	39.2
5	1.347 ± 0.366	0.9883	19.60	0.2541	0.2692	0.6976	0.4092	85.8
10	1.269 ± 0.305	0.9983	19.62	0.2541	0.2671	0.6975	0.4098	150.5
20	1.253 ± 0.298	1.0000	19.62	0.2541	0.2667	0.6975	0.4099	259.6

Note 1: Core Optimization Time measures exclusively the pure 1D-WOA algorithm runtime at T = 20, excluding image matrix operations, pre-filtering overhead, and objective function evaluations. **Note 2:** Values for γ* are presented as Mean ± SD [Min, Max].

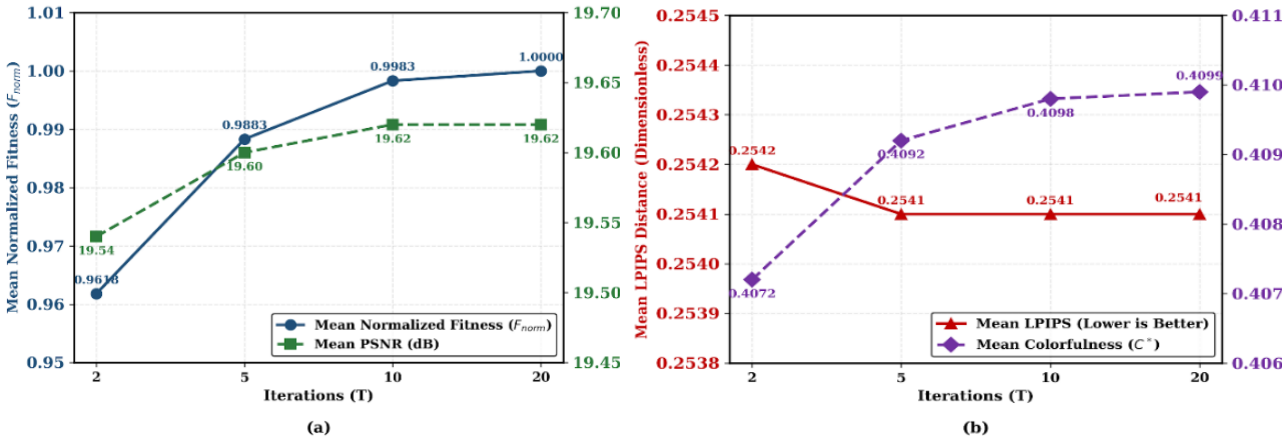


Fig. 5. Convergence kinetics of the 1D-WOA pipeline-adaptive: (a) Fitness and PSNR; (b) LPIPS and Colorfulness.

4.2.3 Comparative Analysis with Post-Processing Baselines

Table 4 shows a quantitative ablation analysis of each one of the post-processing methods that were used for the 1,081 raw GAN-generated images, which allows for a direct comparison of the ablation results of each method without the second step of edge fusion. The results show that the proposed 1D-WOA optimization is the best overall perceptual balance of the methods tested. It has the highest PSNR (18.88 dB), lowest LPIPS (0.3314), highest Fitness score (0.5742), highest Edge Contrast (0.1722), highest Skin Smoothness (0.8101) and highest Colorfulness (0.4140), and also achieves a score of 0.7189 for Identity, which is similar to the best PSO configuration. Conventional enhancement methods, such as CLAHE and histogram equalization, however, have significantly lower computational overheads (9.2 and 7.1 ms, respectively), but lower PSNR and LPIPS and lower identity and fitness scores, meaning that they are computationally efficient at the cost of overall perceptual quality. The skin-smoothness score for MSR is quite high, but the score is not good enough to counter its poorer PSNR (7.94 dB), LPIPS (0.5648), and lowest Identity (0.4171). The population-based optimizations alternatives are GWO and PSO, which have the results closest to the proposed 1D-WOA in terms of image quality, however they have a significantly higher number of computational overheads: 1167.9ms and 1164.1ms respectively. This means that the total computing time of the proposed 1D-WOA is 969.1 ms, which is lower than that of GWO and PSO, while perceptual metrics remain slightly better compared to GWO and PSO, and only the pre- and post-processing overhead of 1D-WOA is 259.6 ms. All alternative methods showed differences at the ($p<0.001$)

 ISSN:2302-8149
e-ISSN:2540-9719

level in comparison with the proposed 1D-WOA method (The designation 'Ref.' indicates the proposed 1D-WOA method is used as the statistical reference baseline for all other competing alternatives to be compared using the paired Wilcoxon signed-rank test). As an overall abstraction, the results support the effectiveness of the proposed 1D-WOA as a balanced optimization algorithm that can enhance multiple aspects of image quality simultaneously, albeit the results also indicate that simpler non-optimization strategies are still significantly faster and hence can be considered as computational-quality trade-offs, rather than competitors for perceptual quality.

Table 4 Standalone quantitative ablation comparison of *post-processing* alternatives averaged over 1,081 raw GAN image outputs.

Method	PSNR (dB) ↑	LPIPS ↓	Identity ↑	Fitness ↑	Edge Contrast ↑	Skin Smooth-ness ↑	Color-fulness ↑	Over-head (ms) ↓	Wilcoxon p-value
AnimeGAN2	18.72	0.3322	0.7178	0.5694	0.1731	0.8090	0.4131	0.0	$p < 0.001$
CLAHE Enhancement	15.96	0.3514	0.6843	0.5888	0.1897	0.7845	0.4139	9.2	$p < 0.001$
GWO Optimization	18.85	0.3316	0.7188	0.5736	0.1726	0.8098	0.4136	1167.9	$p < 0.001$
Histogram Equalization (HE)	15.19	0.3831	0.6623	0.5065	0.1837	0.7658	0.4000	7.1	$p < 0.001$
Multi-Scale Retinex (MSR)	7.94	0.5648	0.4171	0.3923	0.1241	0.9186	0.1472	929.6	$p < 0.001$
PSO Optimization	18.87	0.3315	0.7189	0.5741	0.1723	0.8099	0.4138	1164.1	$p < 0.001$
Proposed 1D-WOA post-processing (Ours)	18.88	0.3314	0.7189	0.5742	0.1722	0.8101	0.4140	969.1	Ref.

4.3 Comparative Validation

The proposed (1D-WOA) framework has been quantitatively compared with AnimeGAN2 and (EPI) [43] for 1,081 facial images as shown in Table 5. The main goal of the proposed framework is to refine the appearance of the cartoonized face while removing distracting elements, such as skin texture, that are not part of the face structure. The proposed approach had the highest spatial intensity variation or Spatial Frequency (101.3095 ± 41.5052) and tonal contrast (RMS Contrast = 0.3515 ± 0.0711) when compared to the other approaches, resulting in greater spatial intensity variations and tonal contrast. Additionally, it has a lower image entropy value (H = 5.3930 ± 0.5532), which is lower than AnimeGAN2 (H = 5.7490 ± 0.4276) and EPI (H = 7.0913 ± 0.3785), showing that the intensity of the images is less complex in general, but not automatically better for smooth skin. The similarity value (S_{id}) for identity was also relatively low, 0.6747 ± 0.0141. The results show that there is a trade-off between structural improvement and perceptual similarity, but not a general superiority for any metric. The quantitative results should be considered alongside qualitative comparisons in Fig. 6 and the skin-smoothness and edge-contrast measurements in Table 2, since the principal contribution is the beauty of the face, not the preservation of identity.

Table 5 Quantitative Comparison of Cartoonization Methods on the Full Dataset (N = 1081).

Method	Identity Similarity (S_{id}) ↑	Entropy (H) ↑	Spatial Frequency (SF)	RMS Contrast (C_RMS) ↑
Source (Original)	—	7.3283 ± 0.3914	12.6612 ± 6.1841	0.2299 ± 0.0452
AnimeGAN2	0.6868 ± 0.0106	5.7490 ± 0.4276	77.4008 ± 19.6602	0.2868 ± 0.0561
Enhanced Pixel Integration (EPI)	0.6964 ± 0.0114	7.0913 ± 0.3785	65.1680 ± 16.8728	0.2662 ± 0.0511
Proposed 1D-WOA	0.6747 ± 0.0141	5.3930 ± 0.5532	101.3095 ± 41.5052	0.3515 ± 0.0711

Note: Values are reported as mean ± standard deviation. The source-image row does not report identity similarity.

 ISSN:2302-8149
e-ISSN:2540-9719

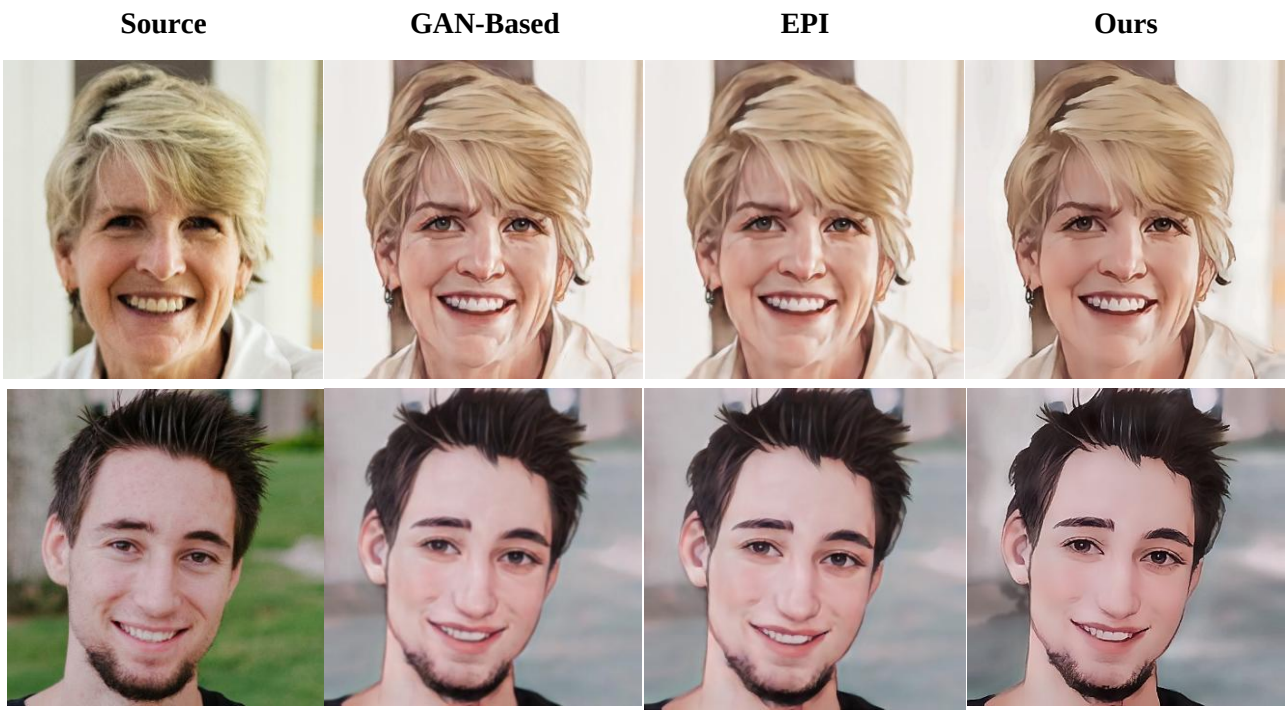


Fig. 6. Qualitative comparison of baseline cartoonization methods against the proposed 1D-WOA.

4.4 Hardware Resource Profiling and Edge Deployment Feasibility

The performance of the operational throughput and resource utilization was assessed on two execution platforms, a cloud-based server GPU (NVIDIA T4 16 GB) and a power-constrained embedded hardware platform (NVIDIA Jetson Orin Nano 8 GB in 15 W mode). Tables 6 and 7 list each of these processing stages with their execution latency, peak memory usage, power usage, and frame rate. The NVIDIA T4 server GPU shows a 1D-WOA integration of 401.2 MB VRAM with a total pipeline processing time of 271.0 ms (3.7 FPS) compared to 178.9 ms (5.6 FPS) when it is not integrated. The entire pipeline runs at 335.0 ms per frame (3.0 FPS) with a measured power consumption of 11.45 W and system memory usage of 428.5 MB (around 5.3% of its total onboard memory). When comparing the post-processing techniques, it can be seen that while baseline Fixed Gamma has lower latency (194.4ms at 5.1FPS on the edge platform Jetson Orin Nano), 1D-WOA adds extra processing time (140.6ms compared to Fixed Gamma on the edge platform) for computing content-adaptive gamma values. These empirical measurements define the precise conditions and limitations of the deployment of the framework in embedded IoT edge devices compared to infrastructure in the cloud.

Table 6 End-to-end hardware resource profiling results (NVIDIA T4 server GPU)

Hardware Platform	Processing Pipeline	Latency (ms)	Peak Memory (MB)	Power Draw (W)	Throughput (FPS)	Edge Feasible?
Server GPU (NVIDIA T4 16GB)	AnimeGAN2 Baseline	178.9 ± 2.7	401.2	~48.5 W (TDP Constrained)	5.6	Yes (Server)
Server GPU (NVIDIA T4 16GB)	AnimeGAN2 + Fixed Gamma	188.4 ± 5.3	401.2	~48.5 W (TDP Constrained)	5.3	Yes (Server)
Server GPU (NVIDIA T4 16GB)	AnimeGAN2 + 1D-WOA (Optimized)	271.0 ± 10.2	401.2	~48.5 W (TDP Constrained)	3.7	Yes (Server)

Note: ms = milliseconds; MB = megabytes; W = watts; FPS = frames per second; TDP = Thermal Design Power.

 ISSN:2302-8149
e-ISSN:2540-9719

Table 7 Hardware resource profiling results (NVIDIA jetson orin nano 8GB – 15 W Mode)

Hardware Platform	Processing Pipeline	Latency (ms)	Peak Memory (MB)	Power Draw (W)	Throughput (FPS)	Edge Feasible?
Jetson Orin Nano (8GB)	AnimeGAN2 Baseline	199.0 ± 11.4	428.5	11.45 W	5.0	Yes (Edge-Deployable)
Jetson Orin Nano (8GB)	AnimeGAN2 + Fixed Gamma	194.4 ± 2.8	428.5	11.45 W	5.1	Yes (Edge-Deployable)
Jetson Orin Nano (8GB)	AnimeGAN2 + 1D-WOA (Optimized)	335.0 ± 12.4	428.5	11.45 W (15 W Mode)	3.0	Yes (Edge-Deployable)

5. Conclusions

In this work, a facial cartoon beautification system based on an adaptive framework is studied, where the skin refinement and preservation of important structural details are examined. The framework combines four techniques: bilateral filtering (to maintain facial boundaries), AnimeGAN2-based stylization, skin-aware blending, and 1D-WOA-based adaptive luminance optimization, to achieve cosmetic smoothing while preserving facial boundaries. The proposed framework incorporates four techniques: 1) Bilateral filtering to preserve facial boundaries; 2) AnimeGAN2-based stylization; 3) Skin-aware blending; and 4) 1D-WOA based optimization of the luminance, with the aim of obtaining cosmetic smoothing while maintaining the facial boundaries. The results reported show that the first beautification step applies a direction to the skin-smoothness measure with a value of 0.7031 and substantially decreases the value of edge contrast. Optimization and edge refinement further improved the normalized edge contrast from 0.1435 to 0.2667. For the NVIDIA Jetson Orin Nano, it was 3.0 frames per second, 11.45 W power consumption, and 428.5 MB peak memory usage. The results indicate that the framework provides increased structural definition and tonal contrast, and preserves the computational feasibility of the study conditions. The results indicate that recovering facial structural contrast following facial smoothing is the main contribution, not simultaneous improvement in all the evaluation metrics. However, this final result did not fully restore the skin's baseline edge contrast, and the enhancement of the skin-smoothness measure is not sufficient to conclude that wrinkles or facial defects were removed. Future studies are needed to include region specific skin quality assessment, standard facial identity assessment, and additional validation across the various generative architectures. Such studies would be able to give more solid evidence for the aesthetics effectiveness and generalizability of the proposed framework.

Acknowledgment

The researcher thanks the Computer Science Department in the Computer Science and Mathematics College, University of Mosul /Iraq, for their support.

References

[1] P. L. Rosin and Y. K. Lai, "*Non-Photorealistic Rendering of Portraits*," in Proc. Workshop Comput. Aesthetics, Eurographics Assoc., 2015. Available: https://orca.cardiff.ac.uk/id/eprint/76344.

[2] K. Wang, X. Liu, D. Zhao, X. Geng, T. Xian, and Y. Chang, *"OIDSty: One-Shot Identity-Preserving Face Stylization*," Image Vis. Comput., Vol. 167, Art. No. 105899, 2026. DOI: https://doi.org/10.1016/j.imavis.2026.105899.

[3] X. Gao, Y. Z. Zhang, and Y. T. Xu, "*Learning to Incorporate Texture Saliency Adaptive Attention to Image Cartoonization*," in Proc. Int. Conf. Mach. Learn. (ICML), 2022. DOI: https://doi.org/10.48550/arXiv.2208.01587.

[4] V. S. Venu and S. Samreen, "*Expressive Cartoon Card (ECC) Generation using Deep Learning for Privacy Preserving Social Media Posts*," Traitement du Signal, Vol. 42, No. 3, pp. 1367-1381, 2025. DOI: https://doi.org/10.18280/ts.420313.

[5] H. Winnemöller, S. C. Olsen, and B. Gooch, "*Real-Time Video Abstraction*," ACM Trans. Graph., Vol. 25, No. 3, pp. 1221-1226, 2006. DOI: https://doi.org/10.1145/1141911.1142018.

[6] Z. Lu and Y. Zhou, "*Advancing Facial Stylization through Semantic Preservation Constraint and Pseudo-Paired Supervision*," arXiv preprint arXiv:2506.22022v1, 2025. DOI: https://doi.org/10.48550/arXiv.2506.22022.

 ISSN:2302-8149
e-ISSN:2540-9719

[7] T. Wang, "*A Comprehensive Review of Image-to-Image Translation, Face Manipulation, and Neural Style Transfer Methods*," TechRxiv, 2024. DOI: https://doi.org/10.36227/techrxiv.174431621.17773362/v1.
[8] F. Zhang, H. Zhao, Y. Li, Y. Wu, and X. Sun, "*CBA-GAN: Cartoonization Style Transformation based on the Convolutional Attention Module*," Comput. Electr. Eng., Vol. 106, Art. No. 108575, 2023. DOI: https://doi.org/10.1016/j.compeleceng.2022.108575.
[9] M. Liu, *Learn Generative AI with PyTorch*. New York, NY, USA: Simon and Schuster, 2024. Available: https://books.google.iq/book
[10] Y. Zhang and L. X. L. Wang, "*Cartoon-Style Image Rendering Transfer based on Neural Networks and Attention Mechanisms*," Comput. Intell. Neurosci., Vol. 2022, Article ID 2958338, 2022. DOI: https://doi.org/10.1155/2022/2958338.
[11] A. Purwanto, Kusrini, E. Utami, and D. Agustriawan, "*A Comprehensive Literature Review on Generative Adversarial Networks (GANs) for AI Anime Image Generation*," in Proc. IEEE Int. Conf. Artif. Intell. Mechatronics Syst. (AIMS), Bandung, Indonesia, 2024. DOI: 10.1109/AIMS61812.2024.10513308.
[12] A. Odena, V. Dumoulin, and C. Olah, "*Deconvolution and Checkerboard Artifacts*," Distill, Vol. 1, No. 10, p. e3, 2016. https://doi.org/10.23915/distill.00003
[13] M. Braik, "*Hybrid Enhanced Whale Optimization Algorithm for Contrast and Detail Enhancement of Color Images*," Cluster Comput.,Vol. 27, No. 1, pp. 231-267, 2024.DOI: https://doi.org/10.1007/s10586-022-03920-9.
[14] C. Tomasi and R. Manduchi, "*Bilateral Filtering for Gray and Color Images*," in Proc. 6th Int. Conf. Comput. Vis. (ICCV), 1998, pp. 839-846. DOI: 10.1109/ICCV.1998.710815.
[15] N. Rana et al., "*Whale Optimization Algorithm: A Systematic Review of Contemporary Applications, Modifications and Developments*," Neural Comput. Appl., Vol. 32, pp. 16245–16277, 2020. DOI: https://doi.org/10.1007/s00521-020-04849-z.
[16] J. E. Kyprianidis and J. Döllner, "*Image Abstraction by Structure Adaptive Filtering*," Theory Pract. Comput. Graphics, pp. 51-58, 2008.
[17] Y. Chen, Y. K. Lai, and Y. J. Liu, "*CartoonGAN: Generative Adversarial Networks for Photo Cartoonization*," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2018, pp. 9465-9474. DOI: https://doi.org/10.1109/CVPR.2018.00986.
[18] J. Kim, M. Kim, H. Kang, and Lee, "*U-GAT-IT: Unsupervised Generative Attentional Networks with Adaptive Layer-Instance Normalization for Image-to-Image Translation*," in Proc. 8th Int. Conf. Learn. Representations (ICLR), 2020. DOI: https://doi.org/10.48550/arXiv.1907.10830.
[19] Z. Zhang, R. Ren, C. Chen, J. Wang, W. Wang, and X. Li, "*Cartoon Image Processing: A Survey*," Int. J. Comput. Vis., Vol. 130, No. 11, pp. 2733–2769, 2022. DOI: 10.1007/s11263-022-01645-1.
[20] J. Long, K. Zhang, Y. Liu, S. Chen, and Q. Luo, "*Edge-Preserving Image Smoothing via Sparse Gradient Enhancement*," Pattern Anal. Appl., Vol. 28, No. 4, p. 160, 2025. DOI: 10.1007/s10044-025-01539-1.
[21] N. E. S. Alhamuy et al., "*Analog Image Processing using Gaussian based Filtering*," in Proc. Int. Conf. Electr. Electron. Eng. (ELECO), Bursa, Turkey, 2023. DOI: 10.1109/ELECO60389.2023.10415976
[22] S. Mirjalili and A. Lewis, "*The Whale Optimization Algorithm*," Adv. Eng. Softw., Vol. 95, pp. 51-67, 2016. DOI: 10.1016/j.advengsoft.2016.01.008.
[23] Y. Han, P. Hu, Z. Su, L. Liu, and J. Panneerselvam, "*An Advanced Whale Optimization Algorithm for Grayscale Image Enhancement*," *Biomimetics*, Vol. 9, No. 12, Art. No. 760, 2024, DOI: 10.3390/biomimetics9120760.
[24] C. Wei, W. Wang, W. Yang, and J. Liu, "*Deep Retinex Decomposition for Low-Light Enhancement*," in Proc. British Mach. Vis. Conf. (BMVC), 2018. Available: https://arxiv.org/abs/1808.04560
[25] T. Karras, S. Laine, and T. Aila, "*A Style-based Generator Architecture for Generative Adversarial Networks*," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Long Beach, CA, USA, 2019, pp. 4401–4410.
[26] N. Tomar et al., "*An Effective Cartoonifying of an Image using Machine Learning*," in Proc. IEEE 4th Annu. Flagship India Council Int. Subsections Conf. (INDISCON), Mysore, India, 2023. DOI: 10.1109/INDISCON58499.2023.10270137.
[27] R. C. Gonzalez and R. E. Woods, Digital Image Processing, 3rd ed. Upper Saddle River, NJ, USA: Pearson Prentice Hall, 2008.
[28] Y. H. Chang, P. H. Chung, Y. H. Chai, and H. W. Lin, "*Color Face Image Generation with Improved Generative Adversarial Networks*," Electronics, Vol. 13, No. 7, p. 1205, 2024. DOI: 10.3390/electronics13071205.
[29] Y. S. Younis and I. M. Khudher, "*Classification of Lateral Incisor Teeth for Aesthetic Dentistry using Intelligent Techniques*," J. Educ. SCI., Vol. 32, No. 1, pp. 91–108, 2023. DOI: 10.33899/edusj.2023.137076.1296.
[30] B. A. Khalaf, M. J. Ridha, I. M. Ahmed, A. H. Fadel, S. Mostafa, S. Kasim, and R. Hidayat, "*Improved Iris Biometric Identification through Hybrid Stochastic Search and Shark Smell Optimization*," Int. J. Adv. SCI. Eng. Inf. Technol., Vol. 15, No. 4, 2025. Available: https://openurl.ebsco.com/EPDB

 ISSN:2302-8149
e-ISSN:2540-9719

[31] D. Chai and K. N. Ngan, "*Face Segmentation using Skin-Color Map in Videophone Applications*," IEEE text Trans. Circuits Syst. Video Technol., Vol. 9, No. 4, pp. 551-564, 1999. DOI: *10.1109/76.767122*.
[32] A. Q. Aloraibi, "*Image Morphing Techniques: A Review*," Technium, Vol. 9, pp. 41–53, 2023. Available: *https://www.researchgate.net/publication/389030873_ID8699_image_morphing*
[33] S. L. Phung, A. Bouzerdoum, and D. Chai, "*Skin Segmentation using Color Pixel Classification: Analysis and Comparison*," IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI), Vol. 27, No. 1, pp. 148-154, 2005. DOI: *10.1109/TPAMI.2005.17*.
[34] M. A. Ali Masaud, S. A. Avcı, and J. Rahebi, "*Detecting Cyberattacks in Smart Grids using VGG16 and Whale Fisher Mantis Optimization Algorithm (WOA-FMO)*," J. Supercomput., Vol. 81, p. 857, 2025. DOI: *10.1007/s11227-025-07345-0*.
[35] M. Pérez Delgado and M. E. Celebi, "*A Comparative Study of Color Quantization Methods using Various Image Quality Assessment Indices*," Multimedia Syst., Vol. 30, No. 1, p. 40, 2024. DOI: *10.1007/s00530-023-01206-7*.
[36] S. M. Abdul Razzaq and B. I. Khaleel, "*Detection of Plants Leaf Diseases using Swarm Optimization Algorithms*," AL-Rafidain J. Comput. SCI. Math., Vol. 15, No. 2, pp. 193–212, 2021.
[37] N. Shahraki, M. Zamani, H. A. Varzaneh, and S. Mirjalili, "*A Systematic Review of the Whale Optimization Algorithm: Theoretical Foundation, Improvements, and Hybridizations*," Arch. Comput. Methods Eng., Vol. 30, pp. 4113–4159, 2023. DOI: *https://doi.org/10.1007/s11831-023-09928-7*.
[38] L. R. Khaleel and B. A. Mitras, "*Hybrid Whale Optimization Algorithm with Modified Conjugate Gradient Method to Solve Global Optimization Problems*," Open Access Library Journal, Vol. 7, No. 6, 2020, DOI: *10.4236/oalib.1106459*.
[39] A. Horé and D. Ziou, "*Image Quality Metrics: PSNR vs. SSIM*," in Proc. 20th Int. Conf. Pattern Recognit. (ICPR), 2010. DOI: *10.1109/ICPR.2010.579*.
[40] I. M. Ahmed, A. K. Ali, and M. S. Mahmood, "*Employing Hybrid Watermarking to Improve Email Security Against Cyber Attacks*," J. Soft Comput. Data Mining, Vol. 6, No. 1, pp. 435–447, 2025. Available: *https://publisher.uthm.edu.my/ojs/index.php/jscdm/article/view/20220*.
[41] J. Chen, G. Liu, and H. Y. Chen, "*AnimeGAN: A Novel Lightweight Generative Adversarial Network for Photo Anime Colorization*," in *International Symposium on Intelligence Computation and Applications*, Springer, Singapore, 2020, pp. 242-256.
[42] B. Lee, "*AnimeGAN2-PyTorch: Novel Face2Paint Portrait Stylization*," GitHub Repository, 2021. [Online]. Available: *https://github.com/bryandlee/animegan2-pytorch*.
[43] S. Choi, H. Ko, K. Bae, H. Lee, H. Joo, and W. Choi, "*Optimizing 2D Image Quality in CartoonGAN: A Novel Approach using Enhanced Pixel Integration*," Computers, Materials & Continua, Vol. 83, No. 1, pp. 335–355, 2025, *https://doi.org/10.32604/cmc.2025.061243*.
[44] Q. Huynh-Thu and M. Ghanbari, "*Scope of Validity of PSNR in Image/Video Quality Assessment*," Electron. Lett., Vol. 44, No. 13, pp. 800–801, 2008. DOI: *https://doi.org/10.1049/el:20080522*.
[45] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "*The Unreasonable Effectiveness of Deep Features as a Perceptual Metric*," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2018, pp. 586–595. DOI: *https://doi.org/10.48550/arXiv.1801.03924*.
[46] T. Leyvand, D. Cohen-Or, G. Dror, and D. Lischinski, "*Data-Driven Enhancement of Facial Attractiveness*," ACM Trans. Graph., Vol. 27, No. 3, pp. 1–9, 2008. DOI: https://doi.org/10.1145/1399504.1360637.
[47] K. He, J. Sun, and X. Tang, "*Guided Image Filtering*," IEEE Trans. Pattern Anal. Mach. Intell., Vol. 35, No. 6, pp. 1397–1409, 2013.
[48] D. Hasler and S. Süsstrunk, "*Measuring Colorfulness in Natural Images*," in *Proceedings of SPIE - Human Vision and Electronic Imaging VIII*, Vol. 5007, pp. 87-95, 2003. DOI: 10.1117/12.477378. *https://doi.org/10.1117/12.477378*.
[49] Y. Blau and T. Michaeli, "*The Perception-Distortion Tradeoff*," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, 2018, pp. 6228-6237. *https://doi.org/10.1109/cvpr.2018.00652*.