Contents lists available at ScienceDirect

Digital Chinese Medicine

journal homepage: http://www.keaipublishing.com/dcmed

Facial color-preserving generative adversarial network-based privacy protection of facial diagnostic images in traditional Chinese medicine

Jilong SHEN[a], Aihua GUAN[a], Xinyu WANG[a], Jiadong XIE[a, b], Youwei DING[a, b], Kongfa HU[a, b, c, d*]

a. School of Artificial Intelligence and Information Technology, Nanjing University of Chinese Medicine, Nanjing, Jiangsu 210023, China

b. Jiangsu Province Engineering Research Center of Traditional Chinese Medicine Intelligence Health Service, Nanjing University of Chinese Medicine, Nanjing, Jiangsu 210023, China

c. Jiangsu Collaborative Innovation Center of Traditional Chinese Medicine in Prevention and Treatment of Tumor, Nanjing, Jiangsu 210023, China

d. Jiangsu Research Center for Major Health Risk Management and Traditional Chinese Medicine Control Policy, Nanjing University of Chinese Medicine, Nanjing, Jiangsu 210023, China

ARTICLE INFO

Article history
Received 18 June 2025
Accepted 30 September 2025
Available online 25 December 2025

Keywords
Traditional Chinese medicine (TCM) inspection
Facial complexion information
Image generation
Privacy preservation
Generative adversarial network
Color space

ABSTRACT

Objective To develop a facial image generation method based on a facial color-preserving generative adversarial network (FCP-GAN) that effectively decouples identity features from diagnostic facial complexion characteristics in traditional Chinese medicine (TCM) inspection, thereby addressing the critical challenge of privacy preservation in medical image analysis.

Methods A facial image dataset was constructed from participants at Nanjing University of Chinese Medicine between April 23 and June 10, 2023, using a TCM full-body inspection data acquisition equipment under controlled illumination. The proposed FCP-GAN model was designed to achieve the dual objectives of removing identity features and preserving colors through three key components: (i) a multi-space combination module that comprehensively extracts color attributes from red, green, blue (RGB), hue, saturation, value (HSV), and Lab spaces; (ii) a generator incorporating efficient channel attention (ECA) mechanism to enhance the representation of diagnostically critical color channels; and (iii) a dual-loss function that combines adversarial loss for de-identification with a dedicated color preservation loss. The model was trained and evaluated using a stratified 5-fold cross-validation strategy and evaluated against four baseline generative models: conditional GAN (CGAN), deep convolutional GAN (DCGAN), dual discriminator CGAN (DDCGAN), and medical GAN (MedGAN). Performance was assessed in terms of image quality [peak signal-to-noise ratio (PSNR) and structural similarity (SSIM)], distribution similarity [Fréchet inception distance (FID)], privacy protection (face recognition accuracy), and diagnostic consistency [mean squared error (MSE) and Pearson correlation coefficient (PCC)].

Results The final analysis included facial images from 216 participants. Compared with baseline models, FCP-GAN achieved superior performance, with PSNR = 31.02 dB and SSIM = 0.908, representing an improvement of 1.21 dB and 0.034 in SSIM over the strongest baseline (MedGAN). The FID value (23.45) was also the lowest among all models, indicating superior distributional similarity to real images. The multi-space feature fusion and the ECA mechanism contributed significantly to these performance gains, as evidenced by ablation studies. The stratified 5-fold cross-validation confirmed the model's robustness, with results reported

*Corresponding author: Kongfa HU, E-mail: kfhu@njucm.edu.cn.

Peer review under the responsibility of Hunan University of Chinese Medicine.

DOI: 10.1016/j.dcmed.2025.12.002

Citation: SHEN JL, GUAN AH, WANG XY, et al. Facial color-preserving generative adversarial network-based privacy protection of facial diagnostic images in traditional Chinese medicine. Digital Chinese Medicine, 2025, 8(4): 455-466.

as mean ± standard deviation (SD) across all folds. The model effectively protected privacy by reducing face recognition accuracy from 95.2% (original images) to 60.1% (generated images). Critically, it maintained high diagnostic fidelity, as evidenced by a low MSE (< 0.051) and a high PCC (> 0.98) for key TCM facial features between original and generated images.

Conclusion The FCP-GAN model provides an effective technical solution for ensuring privacy in TCM diagnostic imaging, successfully having removed identity features while preserving clinically vital facial color features. This study offers significant value for developing intelligent and secure TCM telemedicine systems.

1 Introduction

Facial diagnosis, a cornerstone of traditional Chinese medicine (TCM) inspection, has been developed over millennia since it was systematically documented in *Huangdi Neijing* (《黄帝内经》, *Inner Cannon of Huangdi*) [1]. During the era of the *Inner Canon*, the Five-Color Inspection Method (五色诊法) was established, which diagnoses diseases by observing subtle chromatic variations among five specific hues, as articulated in *Suwen · Wuzang Shengcheng* (《素问 · 五藏生成》, *Plain Questions · Generation of the Five Zang Organs*): "Minute chromatic signs can be discerned visually" (五色微诊, 可以目察). In TCM, facial complexion serves as a critical diagnostic indicator, enabling practitioners to "deduce internal pathologies from external manifestations" (司外揣内), which is one of the fundamental principles of TCM diagnostics [2]. This principle forms a fundamental approach for assessing visceral disorders through external observation [1, 3]. It also aligns with *Lingshu · Benzang* (《灵枢 · 本藏》, *Spiritual Pivot · The Root of the Viscera*): "By observing the external manifestations, one may comprehend the internal organs, thereby discerning the nature of the disease" (视其外应, 以知其内藏, 则知所病也) [4]. Consequently, facial attributes, including color, luster, and lip tone, directly reflect the functional state of Qi and blood circulation, underscoring the clinical significance of facial complexion in TCM diagnosis and treatment [5]. In TCM facial diagnosis, chromatic information plays a vital role across clinical applications, particularly in dermatology, where complexion variations directly indicate the prosperity or decline of Qi and blood and the health status of visceral systems.

The advancement of medical imaging technologies has facilitated the application of digitized facial diagnostic images in TCM telemedicine. Besides, contemporary medical practitioners increasingly integrate advanced scientific methods with TCM principles to precisely analyze facial complexion, thereby strengthening the objectivity and accuracy of diagnostic results [6, 7]. Furthermore, the relevant methods and principles for information collection have been further clarified [8], and comparative studies across systems on concepts such as health-related quality of life also provide a theoretical foundation for this integration [9]. DONG et al. [10] analyzed five-color indicators (cyan, red, yellow, white, and black) in patients with coronary heart disease, chronic renal failure, and chronic hepatitis B, and found that patients with coronary heart disease predominantly showed yellow and red complexions, patients with chronic renal failure showed yellow, cyan, and white complexions, and those with hepatitis B yellow or black complexions, validating the discriminative power of the five-color indicators. CHEN et al. [11] achieved 89.5% classification accuracy for identifying white complexions using backpropagation neural networks. Concurrently, deep learning models like convolutional neural networks (CNNs) have been applied to automate facial color recognition with high accuracy. Related research has confirmed the effectiveness of this technology in extracting TCM facial color features [12].

Face recognition technology based on deep learning has become increasingly mature and widely applied [13]. However, these facial biometrics pose significant privacy risks. Personal Information Protection Law of the People's Republic of China (2021) categorizes biometric data as sensitive personal information, mandating strict de-identification measures [14]. Furthermore, for scenarios involving cross-border data transfer, relevant security assessment requirements must also be met [15]. This creates a critical challenge: balancing diagnostic utilization with identity privacy preservation. Current research demonstrates that non-de-identified facial data poses multifaceted risks. DAVID et al. [16] experimentally validated that standard facial recognition models achieve 76.8% accuracy in recognizing identity by medical images, suggesting potential disclosure of sensitive clinical records through data leakage. International Organization for Standardization explicitly mandates "irreversible de-identification" for medical biometric data, requiring complete identity removal while preserving diagnostic features [17]. GAFNI et al. [18] revealed that conventional methods cause 41.3% attrition in facial chromatic features, compromising TCM complexion analysis. Conventional approaches such as blurring or mosaicking, while partially obscuring identities, incur substantial loss of complexion characteristics, rendering them clinically inadequate. Conversely, standard generative adversarial networks (GANs) [19], while promising for image synthesis, often fail

to address the specific need for color fidelity in TCM. Conventional GANs can induce visually perceptible chromatic shifts in color spaces crucial for TCM analysis, thereby compromising diagnostic reliability. Notably, medical GAN (MedGAN) [20] introduced an adaptive model leveraging Wasserstein loss to address model collapse in the generation of medical images, significantly enhancing its training stability. WANG et al. [21] leveraged conditional GANs (CGANs) to generate high-quality facial expression images while preserving identity information. LI et al. [22] developed a dual-layer GAN architecture for synthesizing structurally coherent and high-definition sketch facial images. WU et al. [23] introduced privacy-protective-GAN (PP-GAN), which employed a novel verifier design for facial de-identification. YANG et al. [24] proposed a mask generation model to overlay high-resolution facial masks for privacy protection. However, these methods do not explicitly prioritize the retention of pathological color features essential for TCM diagnosis.

To resolve this conflict between privacy protection and diagnostic color preservation, we propose a facial color-preserving GAN (FCP-GAN). Our model incorporates three key innovations: (i) building a facial feature extraction model that integrates multiple color spaces [red, green, blue (RGB), hue, saturation, vision (HSV), and Lab] to comprehensively capture facial chromatic features for TCM diagnosis; (ii) introducing an Efficient Channel Attention (ECA) module in the generator to enhance sensitivity to diagnostic-relevant color channels; (iii) designing a dual loss function (adversarial loss and color retention loss) to achieve a balance between the removal of identity features and the retention of color features during adversarial training. This approach aims to generate de-identified facial images that retain diagnostic value, providing a technical foundation for secure and intelligent TCM diagnostic systems.

2 Data and methods

2.1 Data collection

The experimental data were derived from voluntary participants in TCM inspection research at Nanjing University of Chinese Medicine. The study protocol was approved by the Medical Ethics Committee of Affiliated Hospital of Nanjing University of Chinese Medicine (2023NL-255-01). Participants were included if they had unobstructed facial exposure without acute skin diseases; individuals with facial surgery history or systemic conditions causing significant facial discoloration were excluded. All participants provided written informed consent before enrollment. The study data were collected between April 23 and June 10, 2023.

The research data were collected using a proprietary acquisition system developed by our institution. Participants were required to maintain unobstructed facial exposure during image capture. Under standardized natural-spectrum illumination, the device captured whole-body images for TCM diagnosis, from which facial regions were extracted, recorded, and analyzed. Figure 1 illustrates the device's acquisition methodology and the general setup of the capture process.

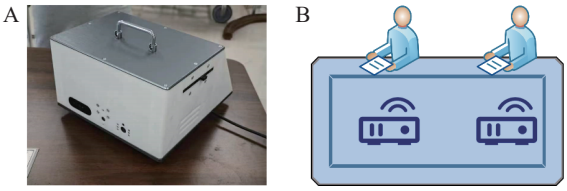

Figure 1 The data acquisition protocol and representative sample data

A, independently developed data acquisition equipment. B, schematic diagram of the data collection process.

Color calibration was performed using the X-Rite Color Checker Classic included in each capture session. A third-order polynomial correction matrix was derived via least-squares fitting:

$$\begin{bmatrix} R_{\text{corr}} \\ G_{\text{corr}} \\ B_{\text{corr}} \end{bmatrix} = M_{3\times 10} \times \begin{bmatrix} R \\ G \\ B \\ R^2 \\ G^2 \\ B^2 \\ RG \\ RB \\ GB \\ 1 \end{bmatrix} \tag{1}$$

Calibration performance was quantified using the CIEDE2000 color difference metric:

$$\Delta E = \sqrt{\left(\frac{\Delta L'}{k_L S_L}\right)^2 + \left(\frac{\Delta C'}{k_C S_C}\right)^2 + \left(\frac{\Delta H'}{k_H S_H}\right)^2 + R_T \frac{\Delta C' \Delta H'}{S_C S_H}} \tag{2}$$

with parametric factors $k_L = k_C = k_H = 1$ and rotation term R_T accounting for chroma-hue interaction.

The mean ΔE across all 216 participants was 1.32 ± 0.18, satisfying the "imperceptible difference" threshold ($\Delta E < 2.3$).

2.2 TCM classification of facial complexions

Within TCM theoretical framework, facial complexions are categorized into physiological and pathological complexions. The pathological complexion comprises five colors—red, yellow, white, black, and cyan—collectively termed as the Five-Color Inspection (五色诊断). Detailed definition of the five colors are as follows: (i) red complexion: "crimson hue resembling cinnabar wrapped in white silk" (如白裹朱); (ii) yellow complexion: "amber

tone akin to orpiment enclosed in yellow gauze" (如罗裹雄黄); (iii) white complexion: "lustrous pallor comparable to polished jade" (如白璧之泽); (iv) black complexion: "deep ebony coloration similar to multilayered lacquer" (如重漆色); (v) cyan complexion: "bluish cast mirroring patinated jade texture" (如苍璧之泽).

The constructed TCM facial complexion dataset consists of 216 participants, stratified as follows: red complexions (n = 54), yellow (n = 117), white (n = 15), black (n = 15), and cyan (n = 15).

2.3 Extraction of facial complexion features

Facial complexion features serve as the primary diagnostic basis in TCM inspection, which are represented through different channels of color spaces in digital image processing. Different color spaces exhibit distinct capabilities in characterizing facial complexion features. To comprehensively describe facial color information, this study employs three color spaces—RGB, HSV, and Lab—for extracting facial features. RGB space is the most common color representation in computer vision, which preserves original skin tone information. HSV space enhances the discriminative power of colors and reduces illumination interference, making it particularly suitable for TCM Five-Color Inspection. Lab space provides illumination-independent color representation, improving the stability and discriminative power of facial colors as well. The complementary combination of these three-color spaces facilitate the construction of a comprehensive facial complexion descriptor, thereby enhancing classification accuracy and model generalizability.

2.3.1 RGB The RGB space represents chromatic information through the intensity of red, green, and blue channels, serving as a fundamental model for color analysis in computer vision. Given TCM facial diagnosis's emphasis on subtle chromatic variations, the RGB space provides an intuitive basis for extracting complexion features. As depicted in Figure 2, the workflow of RGB feature extraction involves: (i) image standardization: resizing raw images to 224 × 224 pixels for dimensional uniformity; (ii) pixel normalization: linearly transforming pixel values from (0, 255) to (– 1, 1) via Formula (3), accelerating gradient descent convergence while mitigating vanishing gradient risks.

$$X_{\text{normalized}} = \frac{X_{\text{original}}}{127.5} - 1 \tag{3}$$

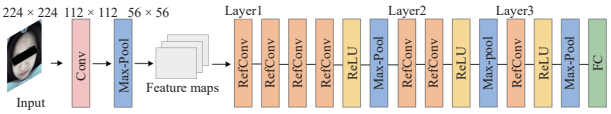

Figure 2 The workflow of extracting RGB facial color features using an enhanced CNN

To extract facial complexion features in the RGB space, this study implemented an enhanced CNN incorporating Refocusing Convolution (RefConv). As a re-parameterized convolutional variant, RefConv establishes supplementary inter-kernel connections among convolutional parameters to amplify sensitivity to local pixel correlations. The core operational mechanism is formalized in Formula (4):

$$y = f(W_{\text{conv}} \times x) + \alpha \times f[\sigma(W_{\text{ref}} \times x)] \tag{4}$$

where W_{conv} denotes the base convolution kernel, W_{ref} represents the refocused kernel, and α (α= 0.2) is trainable coefficients governing cross-kernel feature combined intensity.

Through multi-layer convolutional operations, the original RGB images were transformed into hierarchical feature maps, where each layer discriminatively encoded region-specific color distributions and textural patterns across facial regions. The resultant RGB feature vector was subsequently integrated into the multi-space combined model through concatenation operations.

2.3.2 HSV The facial images were initially converted from the RGB to the HSV space.

Value (V) quantifies color brightness, calculated as Formula (5):

$$V = \max(R, G, B) \tag{5}$$

Saturation (S) measures color purity, defined by Formula (6):

$$S = \begin{cases} \dfrac{V - \min(R,G,B)}{\mathrm{V}}, & \text{if } V \neq 0 \\ 0, & \text{else} \end{cases} \tag{6}$$

Hue (H) represents color category, calculated as Formula (7):

$$H = \begin{cases} 60 \times (G-B)/[V - \min(R,G,B)], & \text{if } V = R \\ 120 + 60 \times (B-R)/[V - \min(R,G,B)], & \text{if } V = G \\ 240 + 60 \times (R-B)/[V - \min(R,G,B)], & \text{if } V = B \end{cases} \tag{7}$$

For Asian skin tones (predominant in our Nanjing University of Chinese Medicine cohort), the HSV range is empirically defined as Formula (8):

$$H_{\min} \leqslant H \leqslant H_{\max},\ S_{\min} \leqslant S \leqslant S_{\max},\ V_{\min} \leqslant V \leqslant V_{\max} \tag{8}$$

where the minimum value of H is 0 and the maximum value is 180, and the minimum value of both S and V is 0 and the maximum is 255.

The extracted skin regions were subsequently converted into a "mask", where skin areas were represented by white pixels while non-skin regions were marked as black. This mask was then applied to the original image through pixel-wise multiplication, selectively preserving visual information exclusively in the mask's white

regions (corresponding to detected skin areas) while suppressing all other content. Formula (9) is the mathematical calculation governing this masking operation:

$$\text{mask}(H,S,V) = \begin{cases} 1, & \text{if } H_{\min} \leqslant H \leqslant H_{\max},\ S_{\min} \leqslant S \leqslant S_{\max}, \text{ and} \\ & V_{\min} \leqslant V \leqslant V_{\max} \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

Finally, the histograms for the H, S, and V channels were computed, and color histograms were plotted to extract the facial image's color features (Figure 3).

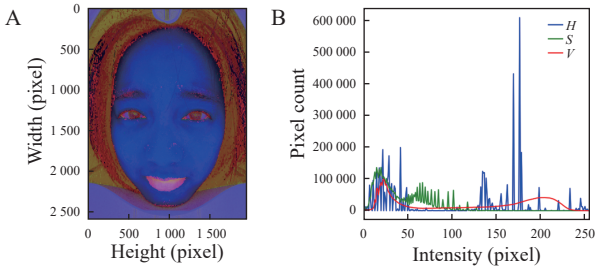

Figure 3 HSV facial color image and color histogram of extracted complexion features

A, HSV facial image. B, HSV color histogram.

2.3.3 Lab The CIELab space, designed to align with human visual perception, comprises three channels: luminance (L), red-green chromaticity (a), and yellow-blue chromaticity (b). This perceptually uniform color space offers superior advantages for characterizing facial colors in TCM diagnosis. The Lab-based facial feature extraction model primarily involves three steps.

(i) Image preprocessing. Facial images in RGB space were first converted to the Lab space using standardized color transformation functions. The luminance channel (L) quantifies light intensity, while the a and b channels encode chromatic opposition along the red-green and yellow-blue axes respectively. This conversion is mathematically defined by Formula (10):

$$\begin{aligned} L &= 116 \times f\left(\frac{R}{255}\right) - 16 \\ a &= 500 \times \left[f\left(\frac{R}{255}\right) - f\left(\frac{G}{255}\right)\right] \\ b &= 500 \times \left[f\left(\frac{G}{255}\right) - f\left(\frac{B}{255}\right)\right] \end{aligned} \tag{10}$$

where f represents the standardized function of color space transformation formalized in Formula (11):

$$f(X) = \begin{cases} X^{1/3}, & \text{if } X > 0.008\,856 \\ \dfrac{903.3 \times X + 16}{116}, & \text{if } X < 0.008\,856 \end{cases} \tag{11}$$

(ii) Clustering analysis. A K-means clustering algorithm (k = 5) partitions Lab color pixels into five diagnostically significant chromatic clusters, corresponding to the Five-Color Inspection in TCM. Cluster optimization minimizes the Euclidean distance between pixels and their respective centroids, as formalized in Formula (12):

$$\min = \sum_{i=1}^{k} \sum_{x \in C_i} \|x - \mu_i\|^2 \tag{12}$$

where C_i denotes the i-th cluster and μ_i represents its centroid.

(iii) Feature extraction. The average luminance, red-green component, and yellow-blue component were extracted from each color region as main features. Guided by TCM theories, which link specific facial regions with internal organ systems, pathological states may manifest as abnormal chromatic distributions. For instance, insufficiency of Qi and blood often corresponds to an elevated proportion of white clusters. Damp-heat syndrome typically exhibits increased yellow cluster ratios. To enhance diagnostic robustness, cluster area ratios, defined as the percentage of each cluster's area relative to the total facial area, are incorporated as auxiliary features. These metrics are formally expressed in Formula (13):

$$F = \left[L_{\text{avg}}, a_{\text{avg}}, b_{\text{avg}}, A_{\text{ratio}}\right] \tag{13}$$

where L_{avg}, a_{avg}, and b_{avg} denote the mean values of luminance, red-green, and yellow-blue components, and A_{ratio} represents the normalized area ratio of the i-th cluster.

Integrating these Lab-derived facial features with RGB and HSV-derived facial features through multi-space combination, a multi-dimensional facial color descriptor was constructed for subsequent FCP-GAN model training and image generation.

2.4 FCP-GAN model

As illustrated in Figure 4, the proposed model comprises three core components: facial data acquisition, complexion feature extraction, and model optimization. During the acquisition phase, facial images were captured using self-developed data acquisition equipment.

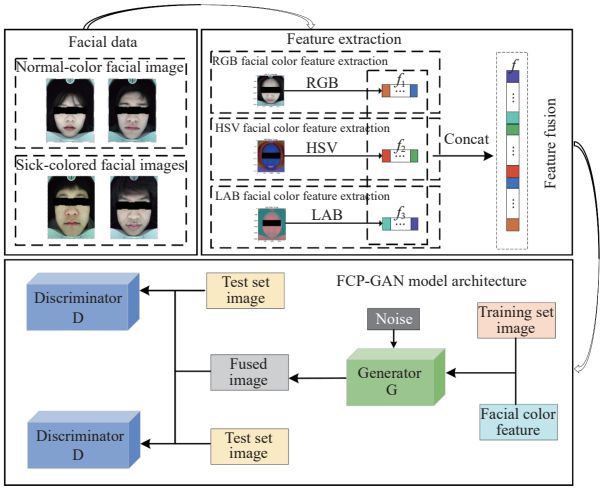

Figure 4 Framework of the proposed FCP-GAN model

The extraction of facial complexion employed a multi-space analytical approach, systematically deriving chromatic characteristics from RGB, HSV, and Lab color spaces. First, the CNN was used to automatically learn and extract color features (f_1) from facial RGB images. Then, the facial RGB images were converted into HSV space to obtain the HSV color features (f_2) through the color histogram. Next, the facial RGB images were converted into the Lab space, and the pixels were clustered by the K-means ($k = 5$) to diagnostically identify significant chromatic clusters for extracting the Lab color features (f_3). Finally, the three features of f_1, f_2, and f_3 were pooled to obtain the facial color features $f = (f_1, f_2, f_3)$

During the model training phase, the FCP-GAN model generated facial images with identity features removed while facial complexion information preserved for diagnostic purpose. The training process utilized the Adam optimizer with a learning rate of 0.000 2 and momentum parameters. For each generator update, the discriminator was trained twice. An ECA module was incorporated after each convolutional layer in the generator to dynamically optimize channel-wise feature weight distributions, thereby enhancing the representation of facial complexion characteristics and suppressing identity-related interference. Throughout the experiments, quantitative evaluations [peak signal-to-noise ratio (PSNR), structural similarity (SSIM), and Fréchet inception distance (FID)] were conducted every 1 000 iterations. (i) PSNR: measuring pixel-level similarity between generated and real images, with higher values indicating lower distortion. (ii) SSIM: evaluating perceptual structural consistency, where values closer to 1 denote higher fidelity. (iii) FID: quantifying distributional divergence in feature space, with lower values reflecting greater visual authenticity.

2.5 Model training strategy and loss function

To effectively remove identity features while preserving diagnostically critical facial color features, this study proposed the FCP-GAN model based on the GAN model with the integration of an ECA mechanism. The ECA module enhances cross-channel interaction without reducing dimensionality, enabling the network to prioritize subtle color variations associated with pathological states while maintaining computational efficiency (Figure 5). The kernel size k in ECA determines the local receptive field for channel attention, controlling how many neighboring channels are considered during attention weight computation.

The generator was designed to transform a latent vector z and conditional information c into compliant synthetic images $\tilde{x}$. In FCP-GAN, the generator employed a sequence of transposed convolutional layers for image synthesis, augmented with an ECA module after each transposed convolution to enhance feature channel

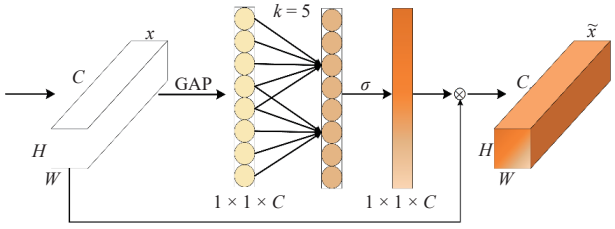

Figure 5 Structure of the ECA module

saliency. The workflow proceeds as follows. The latent vector was projected into an initial feature map of predefined dimensions through a fully connected layer. Then, the hierarchical feature on this feature map was extracted and the spatial resolution restored via a series of transposed convolutional layers. Following each transposed convolution, an ECA module dynamically recalibrated inter-channel attention distributions, selectively amplifying features critical to diagnostic image details while suppressing image noise. The optimized generator synthesized images that strictly adhered to input conditional labels, maximally preserving diagnostically relevant facial color features while obfuscating identity-sensitive attributes. The enhanced generator network is illustrated in Figure 6.

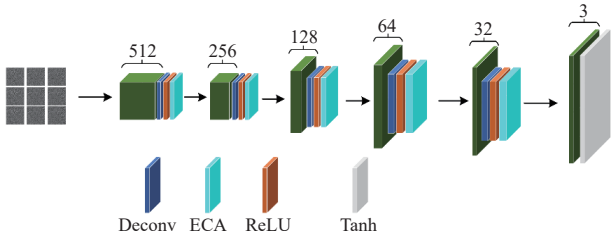

Figure 6 Architecture of the generator network in FCP-GAN

The generator aimed to synthesize facial images with diagnostic value while suppressing identity-related attributes, whereas the discriminator distinguished between real and generated images. To achieve this dual objective, the proposed model integrated two loss functions. The composite loss function combined these components through a weighted summation, as formalized in Formula (14) – (16).

Adversarial loss:

$$L_{\mathrm{GAN}} = E\left[\log D(x)\right] + E\left[\log(1 - D(G(z)))\right] \tag{14}$$

Color preservation loss:

$$L_{\mathrm{color}} = \|H_{\mathrm{real}} - H_{\mathrm{fake}}\|^2 \tag{15}$$

where H_{real} and H_{fake} represent real and generated images.

Composite loss:

$$L_{\mathrm{total}} = L_{\mathrm{GAN}} + \lambda L_{\mathrm{color}} \tag{16}$$

where λ was empirically set to 0.7 through sensitivity

analysis, balancing the removal of identity features and the preservation of facial colors.

2.6 Evaluation methods and metrics

To comprehensively evaluate the model's performance, the following assessment scheme and metrics were employed.

PSNR, SSIM, and FID were used to quantify the pixel-level quality, perceptual structural consistency, and distributional similarity between generated and real images, respectively. Their mathematical formulations are defined as follows:

$$\mathrm{PSNR} = 10 \times \log_{10}\left(\frac{\max_I^2}{\mathrm{MSE}}\right) \tag{17}$$

where max = 255 (for 8-bit images) and MSE is the mean squared error.

$$\mathrm{MSE} = \frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W}[I_r(i,j) - I_g(i,j)]^2 \tag{18}$$

where H and W are the height and width of the image, and the pixel values of the real image and the generated image, respectively.

$$\mathrm{SSIM}(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{19}$$

where μ_x and μ_y are pixel means; σ_x^2 and σ_y^2 represent variances; σ_{xy} denotes covariance; C_1, C_2 stabilize the denominator.

$$\mathrm{FID} = \left\|\mu_r - \mu_g\right\|^2 + Tr\left[\sum_r + \sum_g - 2\left(\sum_r\sum_g\right)^{\frac{1}{2}}\right] \tag{20}$$

where μ_r and μ_g as feature means and $\sum_r$, $\sum_g$ as covariance matrices for real and generated images, respectively.

To validate the removal of identity information, three pre-trained face recognition models—FaceNet, ArcFace, and VGGFace2—were employed as assessment tools. These models map original and generated images into a feature space. Identity similarity scores were computed and converted into metrics including accuracy, precision, recall, and F1 score to quantify the risk of identity leakage.

To validate the diagnostic utility of the generated images, a ResNet-34 classifier adapted for Five-Color Inspection was implemented. The model, pre-trained on ImageNet, had its final fully connected layer replaced with a new structure comprising a 512-dimensional feature layer, a dropout layer (p = 0.5), and a 5-node softmax output layer. This classifier infers TCM-specific facial attributes by analyzing multi-space color distributions (RGB, HSV, and Lab). The fidelity of diagnostic features was assessed by calculating the MSE and Pearson correlation coefficient (PCC) for these TCM attributes between original and generated images.

To mitigate overfitting, a stratified 5-fold cross-validation strategy was adopted. The dataset was stratified by skin tone and partitioned into five folds. In each round, four folds were used for training and the remaining one for validation. This procedure was repeated five times with three independent random splits to reduce variance. Final performance metrics were reported as the mean ± standard deviation (SD) across all runs.

3 Experiment and analysis

3.1 Experimental environment and evaluation metrics

The experimental model was deployed on a high-performance computing platform equipped with an NVIDIA L20 GPU (48GB VRAM). Python 3.7 served as the programming language, with PyTorch 2.3.1 and CUDA 12.1 utilized for model construction, training, and inference, ensuring computational efficiency and stability. Under this configuration, the following quantitative results were obtained to evaluate the performance of the model.

Facial images from 216 participants were collected and divided into the training and testing datasets in a 7 : 3 ratio. The training dataset was employed for the model to learn, while the testing dataset evaluated the performance of the model. A critical focus of the experiment was to assess the model's ability to retain diagnostically relevant facial colors while removing identity-related information for the sake of privacy protection. To address potential overfitting due to limited sample size, cross-validation was implemented during training to enhance generalization robustness. The dataset was stratified by skin tone and partitioned into five folds.

3.2 FID variation trends and result analysis

The FID variation trends during training iterations, serving as a key indicator of image generation quality, are presented in Figure 7. As illustrated, the FID value of FCP-GAN started at 40.56 during initial training and progressively decreased to 23.45, significantly outperforming all baselines (CGAN, DCGAN, DDCGAN, and MedGAN). Notably, FCP-GAN achieved lower FID than MedGAN and maintained this advantage throughout the training process. These FID results demonstrate that FCP-GAN could achieve better performance in generating high-quality images while preserving diagnostically critical facial color features and enhancing privacy protection through effectively removing identity features.

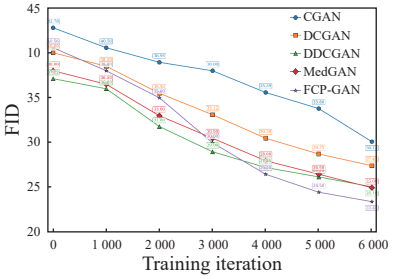

Figure 7 FID variation trends during training iterations for different models

DCGAN, deep convolutional GAN. DDCGAN, dual discriminator conditional GAN.

3.3 Comparison of quality of the generated images

The proposed FCP-GAN was evaluated against four conditional GAN variants (CGAN, DCGAN, DDCGAN, and MedGAN) to assess its efficacy in retaining facial colors and eliminating identity features. Quantitative comparisons are summarized in Table 1. FCP-GAN has obtained ideal results in both PSNR and SSIM metrics. (i) PSNR: FCP-GAN improved the images of PSNR by 4.55 dB compared with CGAN, 4.01 dB compared with DCGAN, 2.78 dB compared with DDCGAN, and 1.21 dB compared with MedGAN. (ii) SSIM: in terms of SSIM, FCP-GAN outperformed CGAN by 0.105, DCGAN by 0.094, DDCGAN by 0.074, and MedGAN by 0.034. These results confirm that FCP-GAN excels in synthesizing high-fidelity facial images that preserve diagnostically relevant facial colors while effectively obfuscating identity-sensitive attributes, and establishing its superiority over existing approaches.

Table 1 Quality assessment of generated images by different models

Model	PSNR (dB)	SSIM
CGAN	26.47	0.803
DCGAN	27.01	0.814
DDCGAN	28.24	0.834
MedGAN	29.81	0.874
FCP-GAN	31.02	0.908

3.4 Impacts of color space combinations on model performance

To validate the enhancement of multi-color space combination on extracting facial colors and the quality of the generated images, comparative experiments were conducted under identical networks and training strategies, evaluating scenarios with single or dual color space features. As shown in Table 2, results demonstrated that single color spaces exhibited limitations in capturing facial chromatic variations and structural information, while multi-space combination significantly improved the model's representational capacity for capturing diagnostic facial colors. The integration of RGB, HSV, and Lab spaces achieved optimal performance across all metrics (PSNR, SSIM, and FID), confirming the necessity of the fusion of multi-spaces in their applications in TCM facial diagnosis.

Table 2 Effects of different color space combinations on model performance

Color space	PSNR (dB)	SSIM	FID
RGB	28.45	0.801	35.12
HSV	27.98	0.794	36.47
Lab	29.13	0.812	34.65
RGB + HSV	29.87	0.829	30.94
RGB + Lab	30.23	0.834	28.76
HSV + Lab	29.99	0.831	29.03
RGB + HSV + Lab	31.02	0.908	23.45

3.5 Validation of the preservation of facial color features

The generated images maintained high fidelity in preserving diagnostically critical color features. Quantitative evaluation revealed minimal MSE values (< 0.051) and PCC values approaching 1 (> 0.98) for key TCM attributes between original and generated images (Table 3). These results indicate that FCP-GAN successfully retains the facial color information essential for TCM diagnosis while removing identity features.

Table 3 Consistency of diagnostic features from facial complexion on images

Facial feature	Original image	Generated image	MSE	PCC
Cold/heat	Heat	Heat	0.051	0.98
Deficiency/excess	Excess	Excess	0.042	0.99
Complexion	Rosy	Rosy	0.033	0.99

3.6 Contribution analysis of multi-space features

To quantify the diagnostic value of different color spaces, we employed SHapley Additive exPlanations (SHAP) to interpret the combined feature vector $f = (f_1, f_2, f_3)$ (Figure 8A). The results revealed that: the a channel in the Lab space alone contributed 42.7%, corroborating its pivotal role in discerning the five TCM facial colors. Collectively, the Lab space accounted for 68.3% of the total contribution, significantly exceeding RGB (19.1%) and HSV (12.6%). This dominant contribution was further corroborated by the high classification performance achieved using the combined features on a test set of 30 generated images. As detailed in the confusion matrix (Figure 8B), the model achieved an overall classification accuracy of 93.3% (28/30 correct) in identifying the five TCM facial colors. The model demonstrated exceptional recognition capability for red (100% accuracy), yellow (100% accuracy), and white (100% accuracy) complexions. Notably, the two misclassifications occurred within the cyan category, with one instance being misidentified

as black and another as white. This minor confusion can be attributed to the inherent chromatic proximity and lower sample representation of these darker, less saturated colors in the dataset, which aligns with known challenges in TCM color differentiation. Nevertheless, these results quantitatively validate that the color features preserved by FCP-GAN retain sufficient discriminative power for accurate Five-Color Inspection.

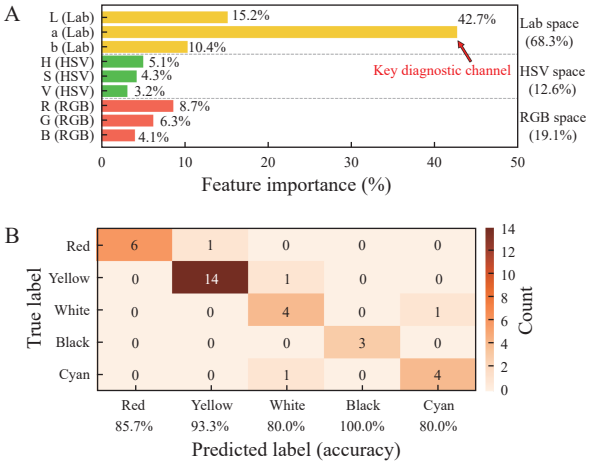

Figure 8 Contribution analysis of multi-space color features and classification of TCM five colors

A, SHAP feature importance visualization. B, five-color classification confusion matrix (n = 30).

3.7 Validation of identity feature removal

To validate the diagnostic utility of the generated images, a ResNet-34 classifier adapted for Five-Color Inspection was implemented. The model, pre-trained on ImageNet, had its final fully connected layer replaced with a new structure comprising a 512-dimensional feature layer, a dropout layer (p = 0.5), and a 5-node softmax output layer. This classifier infers TCM-specific facial attributes by analyzing multi-space color distributions (RGB, HSV, and Lab). The fidelity of diagnostic features was assessed by calculating the MSE and PCC for these TCM attributes between original and generated images. Furthermore, the similarity scores between original and generated images across different models were quantified and compared (Figure 9), providing a direct measure of identity feature removal.

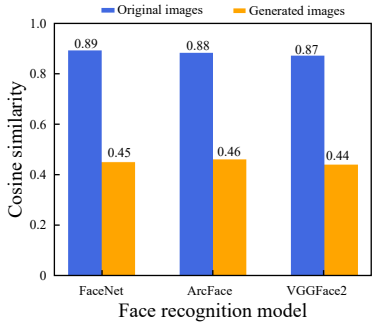

Figure 9 Similarity scores of different models

To further validate the effectiveness of identity obfuscation, we evaluated the generated images using three state-of-the-art face recognition models (FaceNet, ArcFace, and VGGFace2). As summarized in Table 4, the recognition accuracy for generated images dropped drastically to approximately 60%, compared to over 94% for original images across all models. This significant performance degradation in metrics including accuracy, precision, recall, and F1 score conclusively demonstrates that the proposed FCP-GAN successfully removes personal identity features while preserving the diagnostically critical facial colors.

Table 4 Image recognition efficiency by the FCP-GAN model

Recognition model	Image type	Accuracy	Precision	Recall	F1 score
FaceNet	Original	95.2	94.6	96.1	95.3
	Generated	60.1	58.3	62.5	60.4
ArcFace	Original	94.8	93.2	95.6	94.4
	Generated	58.7	57.1	61.0	59.0
VGGFace2	Original	95.6	94.8	96.3	95.5
	Generated	61.2	59.3	63.2	61.2

3.8 Optimization of the color preservation loss weight (λ)

Table 5 shows the impact of different λ values on model performance. λ realized the optimal balance between privacy protection (face recognition accuracy 60.1%) and diagnostic utility (PSNR = 31.02 dB, SSIM = 0.908).

Table 5 Impact of the color preservation loss weight (λ) on FCP-GAN performance

λ value	Accuracy (%)	PSNR (dB)	SSIM
0.3	55.8	29.41	0.846
0.5	57.3	30.15	0.878
0.7	60.1	31.02	0.908
1.0	65.4	31.45	0.922
1.2	69.7	31.62	0.926

3.9 Ablation study on key components of FCP-GAN

To validate the contribution of each key component in FCP-GAN, comprehensive ablation experiments were conducted by comparing four model variants: (i) FCP-GAN w/o ECA: removing the ECA module; (ii) FCP-GAN w/o Lab: using only RGB and HSV features, excluding the Lab space; (iii) FCP-GAN w/o dual-loss: using only adversarial loss, removing the color preservation loss; (iv) FCP-GAN w/o RefConv: replacing the refocusing convolution with standard convolution.

As illustrated in Figure 10, the complete FCP-GAN outperformed all variants across three metrics: PSNR, SSIM, and FID. Removing the ECA module resulted in a

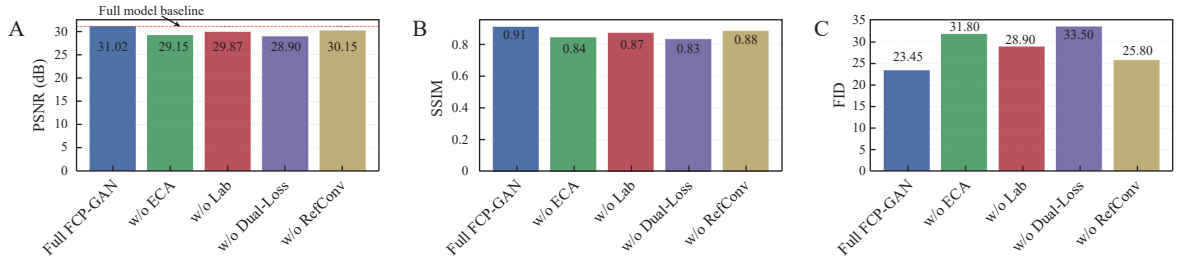

Figure 10 Results of the ablation study on FCP-GAN components

A, PSNR comparison. B, SSIM comparison. C, FID comparison.

PSNR decrease of 1.21 dB and an SSIM reduction of 0.042, indicating that ECA effectively enhanced the channel-wise perception of color features. Excluding the Lab space increased the FID value to 30.94, demonstrating the indispensable role of Lab space in characterizing the five colors. Using only the adversarial loss led to a significant degradation in color preservation capability (MSE increased to 0.098), whereas introducing the color preservation loss (λ = 0.7) markedly improved color consistency in the generated images. Replacing RefConv also caused a decrease in structural similarity (SSIM dropped by 0.028), reflecting its effectiveness in modeling local color textures. In summary, each module contributed positively to the model's performance, with the Lab feature extraction and the dual-loss mechanism being particularly critical for preserving diagnostic color information.

4 Discussion

The proposed FCP-GAN model has realized the dual objectives of effectively removing identity features and well preserving facial colors on TCM diagnostic images. Its superior performance can be attributed to several key mechanisms.

First, the multi-space combination strategy constructed a more comprehensive facial color representation system through the complementarity of RGB, HSV, and Lab spaces. Specifically, the chrominance separation capability of the Lab space in the a/b channels highly aligns with the TCM theory, the Five-Color Inspection, and SHAP analysis confirmed its dominant contribution of 68.3%, which corroborates findings from LI et al. [25] on the discriminative power of the a channel. This underscores that a physiologically-grounded color space is crucial for TCM-oriented models, an aspect not prioritized in general-purpose medical GANs like MedGAN [20].

Second, the integration of the ECA module represents a significant enhancement over standard generator architectures. By enabling dynamic, cross-channel attention weight adjustment, the ECA mechanism allows the generator to focus on pathology-relevant color channels while suppressing identity-related interference. This contributed to an SSIM improvement of approximately 0.042 over the variant without ECA, demonstrating its efficacy in enhancing structural and textural fidelity. In contrast, conventional GANs, which lack such targeted attention mechanisms, have been shown to induce perceptible chromatic shifts (ΔE > 2.3) that compromise diagnostic reliability [26]. Our approach mitigates this risk by explicitly prioritizing color fidelity at the feature channel level.

Third, the dual-loss function provides explicit optimization guidance for the inherently challenging feature decoupling process. Unlike privacy-protection GANs like PP-GAN [23], which primarily focus on de-identification, our model introduces a dedicated color preservation loss. The weighted balance (with λ = 0.7) achieved a 35.1% reduction in face recognition accuracy while maintaining high image fidelity (PSNR > 31 dB). This structured loss design is pivotal for navigating the trade-off between privacy and utility, ensuring that the de-identification process does not come at the cost of diagnostic value.

Nevertheless, this study has its limitations. First, the sample size remains relatively limited even after augmentation, especially for rare complexion types like cyan, which may affect the model's generalizability. Second, although color calibration was applied, illumination variations during image acquisition could still pose challenges to color feature stability. In TCM practices, facial diagnosis must be integrated with other diagnostic methods such as tongue inspection and inquiry; relying solely on facial color images cannot fully substitute for clinical diagnosis.

Our future work will proceed in three directions: first, expanding the dataset's size and diversity; second, exploring multi-modal data fusion to enhance diagnostic robustness; and third, validating the clinical usability and acceptance of the generated images in actual TCM telemedicine scenarios.

5 Conclusion

This study successfully addresses the critical challenge of balancing privacy protection with diagnostic fidelity in TCM telemedicine by proposing a novel FCP-GAN model. By integrating multi-space color features, an ECA mechanism, and dual-loss optimization, our method effectively decouples and removes identity information while robustly preserving the diagnostically vital facial color features. Experimental validations confirm that

FCP-GAN outperforms existing methods, achieving superior de-identification and setting a new benchmark for color fidelity in generated TCM facial images. This work provides a viable and reliable technical solution for ensuring data privacy and security, thereby laying a solid foundation for the development of clinically applicable and ethically compliant intelligent TCM diagnostic systems.

Fundings

National Key Research and Development Program of China (2022YFC3502302), and Graduate Research Innovation Program of Jiangsu Province (KYCX25_2 269).

Competing interests

The authors declare no conflict of interest.

References

[1] GUAN Q, XU Y, YANG S, et al. Research on the relationship between the characteristics of traditional Chinese medicine facial diagnosis and disease. China Journal of Traditional Chinese Medicine and Pharmacy, 2022, 37(2): 902–905.

[2] LI CD. Traditional Chinese Medicine Misdiagnostics. Beijing: China Press of Traditional Chinese Medicine, 2021.

[3] HAO N. A brief discussion on the five colors diagnosis of the Inner Canon of Huangdi. Clinical Journal of Chinese Medicine, 2018, 10(33): 7–9.

[4] SHI LS, HE J. Discussion on life philosophical view of “the order of explicit and implicit states in different regions” in *Huangdi Neijing*. China Journal of Traditional Chinese Medicine and Pharmacy, 2022, 37(4): 1868–1874.

[5] ZHANG HK, LI FF. Research progress on traditional Chinese medicine facial diagnosis information acquisition and recognition. Modernization of Traditional Chinese Medicine and Materia Medica-World Science and Technology, 2015, 17(2): 400–404.

[6] ZHANG YY, ZHOU H, ZAN SH, et al. Application and research progress on new technology of four diagnosis in traditional Chinese medicine. Chinese Computed Medical Imaging, 2021, 27(1): 83–86.

[7] WANG SY, YANG J, TIAN Y. Research progress of key technologies of TCM clinical information collection based on TCM data science. Chinese Archives of Traditional Chinese Medicine, 2024, 42(3): 10–18.

[8] LI SH, XIA SJ, ZHAO W, et al. Methods and principles of information collection in four diagnosis of traditional Chinese medicine. Tianjin Journal of Traditional Chinese Medicine, 2020, 37(3): 266–269.

[9] YU CH, SUN YN, HE LY, et al. Comparative study on the concepts of health related quality of life from TCM and modernized medicine measures. China Journal of Traditional Chinese Medicine and Pharmacy, 2016, 31(11): 4773–4778.

[10] DONG MQ, LI FF, ZHOU R, et al. Facial color feature analysis in the different organs of disease based on image processing. China Journal of Traditional Chinese Medicine and Pharmacy, 2013, 28(4): 959–963.

[11] CHEN MZ, CEN YG, XU JT, et al. Study on observation diagnosis automatic complexion recognition based on image processing. Chinese Journal of Information on Traditional Chinese Medicine, 2018, 25(12): 97–101.

[12] SUN KN, SUN Q, LI XX, et al. Research on TCM color extraction and recognition based on the convolutional neural network. China Journal of Traditional Chinese Medicine and Pharmacy, 2021, 36(7): 4286–4290.

[13] LIANG QQ. Face recognition study of fusion features on CNN and SVM. Modern Information Technology, 2020, 4(19): 62–65,70.

[14] Standing Committee of the National People's Congress. Personal Information Protection Law of the People’s Republic of China. Gazette of the Standing Committee of the National People’s Congress of the People’s Republic of China, 2021, (6): 1117–1125.

[15] National Internet Information Office. Measures of data cross-border transfer security assessment. Gazette of the State Council of the People’s Republic of China, 2022, (24): 39–41.

[16] DAVID M, DAVID H. Anonymising pathology data using generative adversarial networks. Medical Imaging 2022: Digital and Computational Pathology. Bellingham: SPIE, 2022: 1203917.

[17] ISO/IEC. Information technology-Security techniques-Biometric information protection: ISO/IEC 24745: 2022. Geneva: ISO Copyright Office, 2022.

[18] GAFNI O, WOLF L, TAIGMAN Y. Live face de-identification in video. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 9378–9387.

[19] GOODFELLOW IAN, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets. Advances in Neural Information Processing Systems. 2014, 27: 2672–2680.

[20] GUO KH, CHEN J, QIU T, et al. MedGAN: an adaptive GAN approach for medical image generation. Computers in Biology and Medicine, 2023, 163: 107119.

[21] WANG XX, LI FF, CHEN Q. Facial expression generation method based on enhanced conditional generative adversarial networks. Journal of Chinese Computer Systems, 2020, 41(9): 1987–1992.

[22] LI KX, CAO L, DU KN. Face sketch synthesis method based on double layer generative adversarial networks. Computer Applications and Software, 2019, 36(12): 176–183.

[23] WU YF, YANG F, XU Y, et al. Privacy-protective-GAN for privacy preserving face de-identification. Journal of Computer Science and Technology, 2019, 34(1): 47–60.

[24] YANG Y, HUANG YY, SHI M, et al. Invertible mask network for face privacy preservation. Information Sciences, 2023, 629: 566–579.

[25] LI FF, LI GZ, ZHOU R, et al. PLS and LDA for gloss recognition in facial inspection of traditional Chinese medicine. Modernization of Traditional Chinese Medicine and Materia Medica-World Science and Technology, 2011, 13(6): 977–981.

[26] LIU J, ZHANG Y, WANG H, et al. A novel facial color recognition method for traditional Chinese medicine based on hyperspectral imaging. Journal of Innovative Optical Health Sciences, 2021, 14(3): 2150006.

基于保色生成对抗网络的中医面诊图像隐私保护方法

沈纪龙[a], 管爱华[a], 王欣宇[a], 谢佳东[a, b], 丁有伟[a, b], 胡孔法[a, b, c, d*]

a. 南京中医药大学人工智能与信息技术学院, 江苏 南京 210023, 中国
b. 南京中医药大学江苏省智慧中医药健康服务工程研究中心, 江苏 南京 210023, 中国
c. 江苏省中医药防治肿瘤协同创新中心, 江苏 南京 210023, 中国
d. 南京中医药大学江苏重大健康风险管理与中医药防控政策研究中心, 江苏 南京 210023, 中国

【摘要】目的 提出一种基于保色生成对抗网络（FCP-GAN）的面部图像生成方法，旨在有效分离中医面诊中的身份特征与诊断面色特征，以解决医学图像分析中隐私保护的关键挑战。**方法** 研究数据来源于2023年4月23日至6月10日在南京中医药大学采集的参与者面部图像，采集过程使用中医全身望诊数据采集设备并在受控光照下进行。所提出的FCP-GAN模型通过三个关键组件实现去除身份特征与保留面色特征的双重目标：（1）多空间融合模块，全面提取红、绿、蓝（RGB），色调、饱和度、明度（HSV）和Lab颜色空间的颜色属性；（2）生成器中引入高效通道注意力（ECA）机制，以增强对诊断关键颜色通道的表征能力；（3）结合去标识化对抗损失与专用颜色保持损失的双重损失函数。模型采用分层5折交叉验证策略进行训练与评估，并与4种基线生成模型进行比较：条件生成对抗网络（CGAN）、深度卷积生成对抗网络（DCGAN）、双判别器条件生成对抗网络（DDCGAN）、医学生成对抗网络（MedGAN）。主要从以下4个维度评估性能：图像质量[峰值信噪比（PSNR）、结构相似性（SSIM）]、分布相似性[弗雷歇起始距离（FID）]、隐私保护效果（人脸识别准确率）和诊断一致性[均方误差（MSE）、皮尔逊相关系数（PCC）]。**结果** 分析最终纳入216名受试者的面部图像。与基线模型相比，FCP-GAN取得了最优性能，其PSNR为31.02 dB，SSIM为0.908，相较最强基线模型（MedGAN）分别提升了1.21 dB和0.034；其FID值（23.45）也为所有模型中最低，表明生成图像与真实图像的分布最为接近。消融实验证实，多空间特征融合与ECA机制对此性能提升贡献显著。分层5折交叉验证确认了模型的稳健性，所有结果均以多次运行的平均值±标准差报告。该模型能有效保护隐私，将人脸识别准确率从原始图像的95.2%降至生成图像的60.1%。至关重要的是，它保持了较高的诊断保真度，原始图像与生成图像在关键中医面部特征上表现出低MSE（< 0.051）和高PCC（> 0.98）。**结论** FCP-GAN模型为保障中医诊断图像隐私提供了一种有效的技术方案，能够成功去除身份特征，同时保留临床至关重要的面部颜色特征。本研究对开发智能、安全的中医远程医疗系统具有重要价值。

【关键词】中医面诊；面色信息；图像生成；隐私保护；生成对抗网络；颜色空间