IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 13 January 2024, accepted 18 March 2024, date of publication 25 March 2024, date of current version 3 April 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3381535

RESEARCH ARTICLE

Advanced Pigmented Facial Skin Analysis Using Conditional Generative Adversarial Networks

AN-CHAO TSAI[1], (Senior Member, IEEE), PATRICK PO-HAN HUANG[2], ZHONG-CHONG WU[3], AND JHING-FA WANG[3], (Life Fellow, IEEE)

[1]International Master Program of Information Technology and Application, National Pingtung University, Pingtung 91201, Taiwan
[2]Huang PH Dermatology and Aesthetics, Kaohsiung 80458, Taiwan
[3]Department of Electrical Engineering, National Cheng Kung University, Tainan 70101, Taiwan

Corresponding author: Patrick Po-Han Huang (skindr.huangph@gmail.com)

This work involved human subjects or animals in its research. The authors confirm that all human/animal subject research procedures and protocols are exempt from review board approval.

ABSTRACT In recent years, artificial intelligence (AI) approaches in computer vision and medical technology have been combined to create various convenient and accurate tools to assist medical treatments. In this work, we propose conditional generative adversarial networks (conditional GANs)-based pigmented facial skin analysis system for melasma diagnosis. In the past, melasma diagnosis was based on subjective diagnoses from doctors, and there were few automatic melasma analysis methods. The proposed system helps to determine the region according to the melasma's severity. Areas associated with melasma and hemoglobin are detected to determine whether they may require special treatments. Furthermore, the proposed work cooperates with HUANGDERM dermatology to collect a facial skin pigmented dataset. We divide the dataset into 3,000 groups for training datasets and 678 groups for testing. Each group contains four categories of images: standard white light, polarized light, melanin and hemoglobin distribution. As a result, the proposed system successfully generates melasma and hemoglobin images and performs well with respect to subjective and objective evaluations.

INDEX TERMS Conditional GANs, hemoglobin, melasma, pigmented facial skin analysis.

I. INTRODUCTION

The skin is an essential organ for humans. Apart from being a reference to a person's ethnicity, facial skin is often considered in a person's first impression. Studies have even shown a relevance between skin conditions and mental health [1]. As most people are eager to have clean and flawless facial skin, they may apply various skincare products to maintain the quality of the facial skin. The colors of the human skin are mainly determined by the distributions of melanin and hemoglobin [2]. The concentrations of melanin and hemoglobin are also related to various skin diseases. In normal skin, the melanin is evenly distributed, resulting in an even skin tone. However, prolonged sun exposure or skin diseases may cause excessive melanin precipitation in some regions, giving these areas darker appearances. For example, melasma causes the skin to turn brown, and hemoglobin makes the skin red. Skin diseases such as acne, rosacea and telangiectasia can also cause changes in the structure of blood vessels, increase the concentration of hemoglobin and cause uneven skin tones.

This study focuses on analyzing the melasma and hemoglobin concentrations to assist dermatologists during treatments. Melasma is a common hyperpigmented skin disease characterized by light yellow or dark brown patches on the surface of the face [3]. If left untreated, the patches will gradually expand and may usually be removed through laser treatments. However, the course of treatment for melasma varies from long to short, and some patients cannot distinguish the difference between each melasma treatment, leading to medical disputes.

Nowadays, the pigmented distribution can be examined through the Canfield device developed by Canfield Imaging Systems [4]. The Canfield device transforms skin images into hemoglobin and melanin distributions, which is well accepted

The associate editor coordinating the review of this manuscript and approving it for publication was Parikshit Sahatiya.

among dermatologists. The facial melanin distribution map from the Canfield device allows dermatologists to explain the severity of melasma so that the patient is more aware of the effectiveness of the treatment, thus avoiding medical disputes. However, the Canfield device is expensive, and most dermatologists are unwilling to invest in such an instrument.

This work aims to integrate AI and computer vision technologies to develop a cost-effective pigment distribution analysis system that generates an image with comparable quality to that from a Canfield device. Moreover, the severity and region of melasma are also provided based on the generated facial pigment distribution map and the proposed melasma analysis technology. As a result, the proposed system successfully generates melasma and hemoglobin images and performs well with respect to the MSE (Mean Square Error), MAE (Mean Absolute Error), PSNR (Peak Signal to Noise Ratio) and SSIM (Structural Similarity Index Measure).

The rest of this paper is organized as follows. Section II provides the related works. Section III presents the proposed pigmented facial skin analysis system in detail. Experimental results and evaluation are given in Section IV. Finally, the concluding remarks are shown in Section V.

II. RELATED WORKS

A. SKIN CONDITION ANALYSIS

1) PIGMENTED FACIAL SKIN ANALYSIS

Nowadays, the scope of skin treatment is not limited to diseases, and many people perform medical and cosmetic surgeries for aesthetic purposes. Deep neural network (DNN) based research is applied less to the fields of skin beauty technology than in the fields of pathological skin conditions.

Chang and Huang [5] replaced traditional contact probes with digital cameras to capture skin information and adopted computer vision techniques to analyze skin conditions. They proposed a series of standardized detection procedures to analyze skin features such as skin spots, wrinkles and acne and provided objective skin condition assessment reports. Wang and Li [6] proposed a facial pore detection algorithm that combines the characteristics of skin pigment distribution and optimal scale and effectively eliminates skin interference during detection. Both the speeded-up robust features (SURF) and scale-invariant feature transform (SIFT) algorithms were used to detect different cortical parts and calculate a threshold through the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) method [7]. The experimental results showed an improvement in the accuracy of facial pore detection.

Liu et al. [8] proposed an image-processing-based facial speckle analysis technique. They observed that the melanin in melasma is usually distributed on the skin surface, but the contrast between the dark spots and normal skin is negligible. Therefore, it is difficult to separate the pigmented region from normal skin and extract the contour from the facial spots. However, they found that the blue channel in RGB color space provided the clearest contour of the spots and used it for outline extraction.

Demirli et al. [4] proposed the RBX technology, which transfers an RGB image into the RBX color space. The RBX technology, referred to as a new color space, is a trading name developed by Canfield Imaging Systems. The skin image is represented according to the melanin and hemoglobin components taken under polarized illumination. The cross-polarization eliminates the specular reflection on the skin surface and improves the visibility of the epidermis and dermis, where melanin and hemoglobin are located. The RBX technology has been embedded in their products and is widely used in dermatology, cosmetic medicine and aesthetic skin care.

Sanjar et al. [9], proposed a fully convolutional network model for skin-lesion segmentation, the model used a binary cross-entropy loss function to compare the ground truth with the resultant image segmentation. The weights and biases are updated via backpropagation during training, and the final parameter values are saved as model checkpoints. The key advantage of this recommended architecture over the classical U-Net framework is its upsampling stage, which allows for better accuracy in segmenting skin lesions. Ding et al. [10] presented the automatic identification of benign pigmented skin lesions (PSLs) using deep convolutional neural network. The study aimed to develop a computer-aided detection system for accurate identification of PSLs from images captured using a digital camera or a smart phone. The authors proved that the YOLOv5-based system can potentially be used in the clinical identification of PSLs.

2) SKIN CONDITION DATASET

There are few skin datasets compared with face recognition datasets, and the skin datasets are mainly used for skin diseases and melanoma.

Mendonça et al. presented the PH2 dataset [11] with 200 dermoscopic images of melanocyte lesions obtained from Pedro Hispano Hospital. The dataset has 80 common moles, 80 atypical moles and 40 melanomas. Tschandl et al. released the HAM10000 (Human Against Machine with 10,000 training images) dataset [12]. The HAM10000 dataset contains dermoscopy images from different groups of people, consisting of 10,015 images in 7 categories and information on pigmented lesions. Combalia et al. released the BCN20000 dataset [13] collected from hospitals in Barcelona from 2010 to 2016 and consists of 19,424 dermatoscopic images of skin lesions. This dataset contains lesion locations that are difficult to diagnose, such as lesions found in nails and mucous membranes. However, the dataset is unsuitable for large-scale lesions and lesions with depigmentation.

B. SKIN DETECTION

1) SKIN SEGMENTATION

Skin detection is generally used to detect candidate regions such as faces, hands or other skin surfaces in the scene. It is

further applied to face detection, gesture analysis, and image filtering. In 2010, Enneha et al. [14] analyzed 11 color spaces and found that the phase pixel color converted to Hlab and YCbCr space has the best skin detection effect. Solanke and Gore [15] converted the image from RGB color space to YCbCr color space with a threshold for skin detection.

However, the results of this study indicated that if only a threshold is used as the criterion to separate the skin and non-skin pixels, the skin detection effect is easily affected by the background and brightness. Shaik et al. [16] used the RGB, HSV, and YCbCr color spaces to separate skin and non-skin regions. Experimental observations showed that if the picture has a simple background, both HSV and YCbCr color spaces have good results, but for a complex background, the YCbCr color space has a slightly better performance.

Traditional algorithms are used to detect the skin in [14], [15], and [16], but recently, researchers have applied deep learning for skin detection. Li et al. [17] proposed a hybrid network architecture for Traditional Chinese Medicine Inspection. Li's work is capable of detecting and segmenting facial components into 6-category (left eye, right eye, lips, tongue, face, and background) for diagnosis. With the rapid processing speed, its work can be further integrated into real applications. Hashemifard et al. [18] presented a robust and efficient method for weakly supervised human skin segmentation using guidance attention. They addressed challenges such as variability in skin color, pose, and illumination, and incorporates two attention modules and an efficient network architecture to improve segmentation results.

2) SKIN COLOR DATASET

Datasets for skin detection in training or testing are much more difficult to obtain than for face recognition since they require manual marking for the skin regions, and the dataset sizes are relatively small. The SFA dataset released by Casati et al. [19] in 2013 was a face dataset based on AR [20] and FERET [21]. The dataset contains 1,118 faces with more specifications for easier segmentation of the skin regions. Kawulok et al. [22] released the HGR (Hand Gesture Recognition) dataset that contains 899 photos taken in a controlled environment of the hands of 12 different people. The HAM10000 dataset comprises 10,000 training images used for detecting pigmented skin lesions. Tschandl et al. [23] gathered dermatoscopic images from diverse populations, obtained and stored through various modalities.

C. GENERATIVE ADVERSARIAL NETWORK IN COMPUTER VISION

Goodfellow et al. [24] proposed the CNN-based GAN (Generative Adversarial Networks), which is currently widely used in image super-resolution, image colorization, text-to-image translation [25], image augmentation [26], and video generator [27]. Mirza and Osindero [28] proposed the cGAN (Conditional Generative Adversarial Networks) framework with an additional condition for the generator, so the discriminator has to check based on that condition. However, the pictures generated by the traditional GAN and cGAN have low resolution. Naglah et al. [29] overcome the cGAN restriction and proposed a digital pathology system that accurately detects and quantifies the footprint of fibrous tissue in Hematoxylin and Eosin.

When there are two types of data, such as the paired dataset Cityscapes [30], the input data of the Pix2Pix framework [31] no longer need random codes but requires paired correlation for the generator and discriminator. The Pix2Pix framework [31] adopted the U-net-based framework and patch-wise discriminators (PatchGAN) for its network architecture. However, experimental results showed that the Pix2Pix framework only generates 256×256 images, and the generated image becomes blurred when the required resolution is higher. A considerable improvement is made in [32] to generate $2,048 \times 1.024$ images with visually appealing results. The key success in the improved Pix2Pix framework includes a coarse-to-fine generator, multi-scale discriminator and improved adversarial loss. Since paired datasets are hard to collect, the CycleGAN [33] is presented for unpaired data. The CycleGAN has two generators and discriminators for cycle consistency to guarantee the similarity between the generated image and the original input image. Ko et al. [34] presented the DCR (dense consistency regularization) with auxiliary self-supervision loss that enforces point-wise consistency for the discriminator. The experiment results demonstrated better performance than the CycleGAN. In [35], the authors introduce GP-UNIT. This model significantly enhances unsupervised image translation by utilizing generative priors from pre-trained class-conditional GANs. Furthermore, in study [36], Xie et al. have made significant strides by introducing DECENT, a method that leverages density estimators. Their approach focuses on mapping high-probability patches across different domains while preserving semantic integrity, a crucial aspect of image translation tasks. Lastly, in [37], a new framework named EnCo is introduced. EnCo effectively addresses the limitations of existing GAN-based methods in unpaired image-to-image translation. It maintains content fidelity through innovative latent space constraints.

III. PROPOSED SYSTEM

The proposed system architecture is shown in Fig. 1. This section mainly discusses the functionality of the facial skin pigment distribution image generator and the other modules used in the experiment.

The facial skin pigment distribution image generator aims to convert the image from the RGB to the RBX-like color space [4]. The proposed work adopts the conditional GANs framework from [32] to implement the face pigment distribution generator. While a low resolution has been a major issue in the GAN architecture, in pigmented skin analysis, more accurate diagnosis recommendations may be achieved with more retained details. Our collected dataset has a resolution

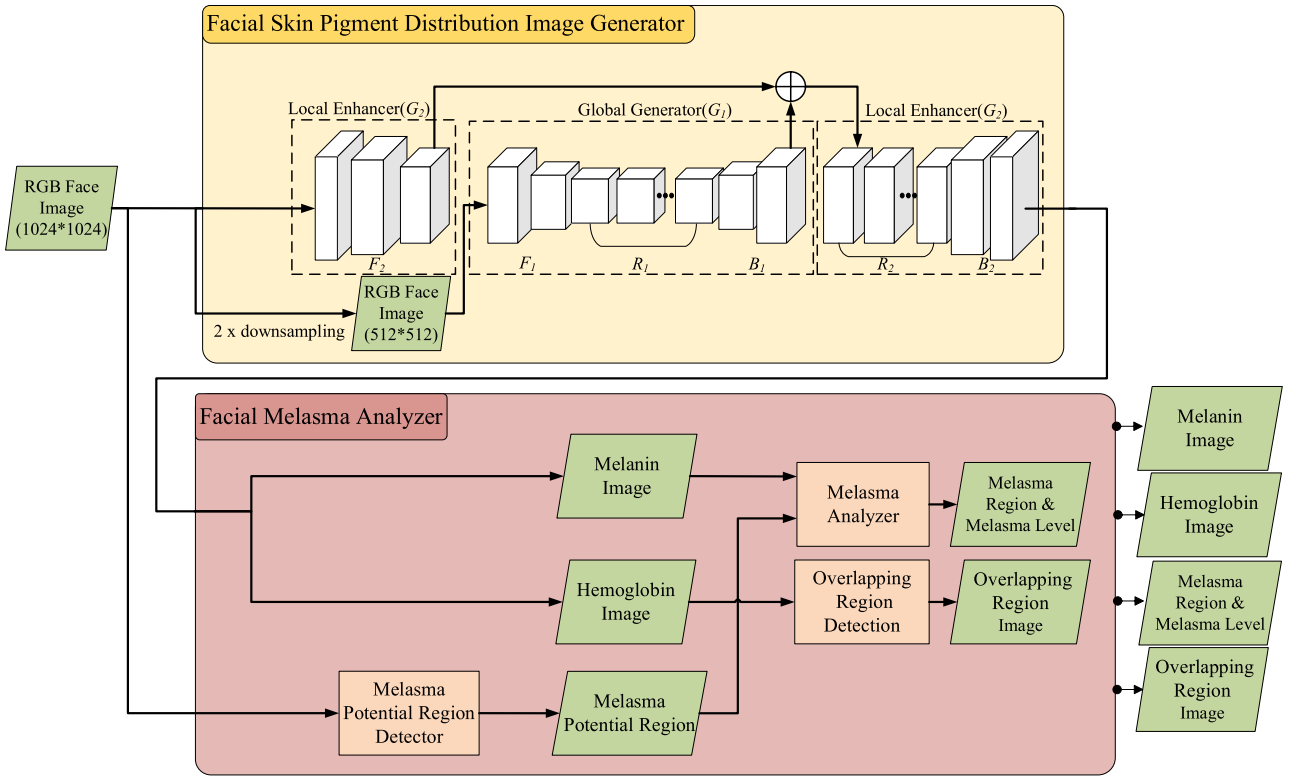

FIGURE 1. The proposed system architecture.

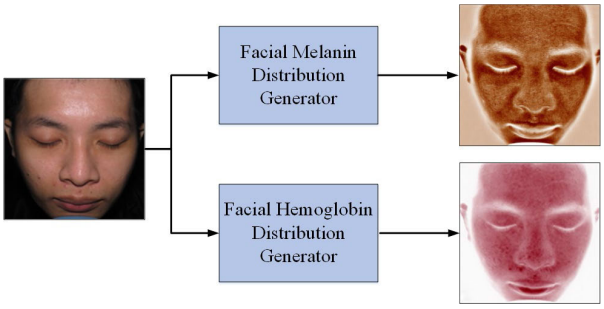

FIGURE 2. The overview of Facial skin pigment distribution image generator.

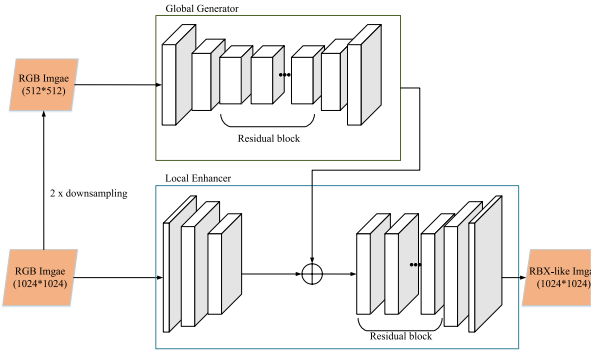

FIGURE 3. Data flow of the coarse-to-fine generator.

of 3, 312 × 3.808. However, due to the trade-off between hardware limitations and computation complexity, we down-sample the images to 1, 024 × 1.024 for input and output. The expected output of the facial melanin distribution map and hemoglobin are depicted in Fig. 2.

A. COARSE-TO-FINE GENERATOR

The conditional GANs framework improves image quality through a coarse-to-fine generator and multi-scale discriminator. The coarse-to-fine generator can be divided into two sub-generators, i.e., a global generator and a local enhancer, as shown in Fig. 1. The global generator generates a small-scale image that preserves the overall outline of the object in the image. The local enhancer deal with a large-scale image to preserve the details of the original images.

The global generator (G_1) is the image style conversion network architecture proposed by Johnson et al. [38], which has been proven successful in image style transformation. The global generator consists of three parts: 1) a down-sample convolutional network (F_1), 2) a string of residual block (R_1) networks, and 3) an up-sample deconvolution network (B_1). The global generator converts a 512 × 512 face image into a 512 × 512 facial pigment distribution map.

The local enhancer (G_2) network architecture is similar to the global generator, which is also composed of three parts: 1) a down-sample convolutional network (F_2), 2) a

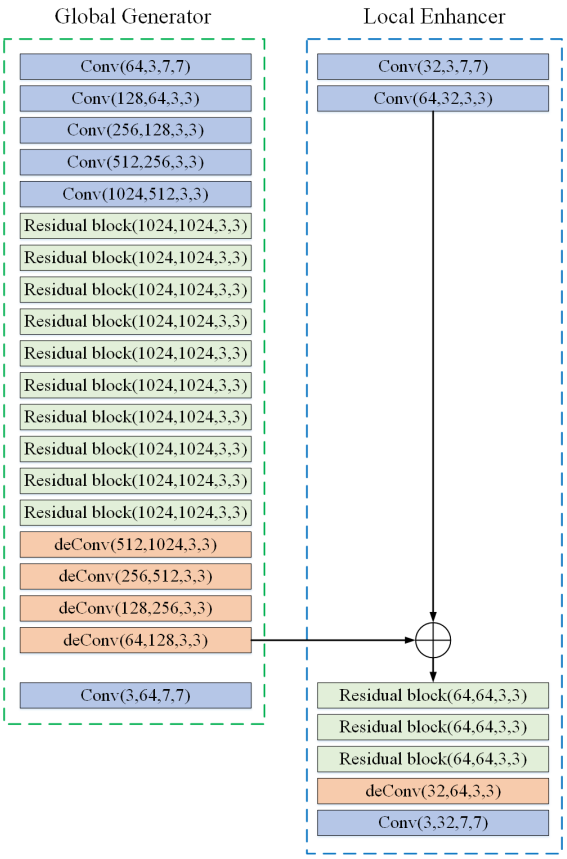

FIGURE 4. The network architecture of coarse-to-fine generator.

string of residual block networks (R_2), and 3) an up-sample deconvolution network (B_2). The local enhancer converts a 1024 × 1024 face image into a 1024 × 1024 facial pigment distribution map. The main difference between the local enhancer and global generator is the input of the residual block shown in Fig. 3. The input of the global generator comes from F_1, and the input of the local enhancer comes from the element-wise sum of the output of F_2 and the input of B_1. The network architecture diagram of the coarse-to-fine generator for the global generator and local enhancer is shown in Fig. 4. The Conv block in Fig. 4 can be decomposed into convolution, batch normalization and ReLU (Rectified Linear Units). Similar to the Conv block, the deConv block is composed of transpose convolution, batch normalization and ReLU, respectively.

The first convolutional layer of the global generator has a reflection padding and 64 filters (a 7 × 7 kernel and a stride of 1). The input and output feature map size remains the same, but the number of channels has changed. The proposed work adopted reflection padding since it symmetrically fills the edges to obtain better convolution results. In the down-sampling convolutional layer, the number of filters in each layer is doubled compared to the previous layer. The convolutional layers from the second to the fifth layers have a stride of 2, zero padding, and a kernel size of 3 × 3. Therefore, the length and width of the output feature map in each convolutional layer are half of the input feature map.

After five layers of down-sampling convolutional layers, the feature map enters the 10-layer residual block. When the neural network has too many layers, the gradient disappears, and training cannot converge. Therefore, He et al. [39] used the residual method to train a relatively deep neural network for image classification. Each convolution of the residual block has an initial padding; after the feature map passes through a series of residual blocks, the size of the feature map and the number of channels remain unchanged. In the up-sampling deconvolution network, the deconvolution network is composed of a transpose convolutional layer. The available output size of the convolutional layer is given as:

$$\boldsymbol{O} = \left\lfloor \frac{\boldsymbol{I} + 2P - K}{S} \right\rfloor + 1 \tag{1}$$

where $\boldsymbol{O}$ represents the output size, $\boldsymbol{I}$, P, K, and S stand for the input size, padding size, kernel size, and stride size, respectively.

When the output size of the convolutional layer is an integer, the output size formula of the transpose convolution layer would be pushed back and can be presented as:

$$\boldsymbol{O} = (\boldsymbol{I} - 1) \times S - 2P + K \tag{2}$$

where P and S are both from their corresponding convolutional layers. The decimal will be unconditionally rounded off if the output size $\boldsymbol{O}$ is divisible. In this case, the formula for the output size of the transpose convolutional layer is pushed back in the reverse direction and is corrected by:

$$\boldsymbol{O} = (\boldsymbol{I} - 1) \times S - 2P + K + \alpha \tag{3}$$

where α implies adding α columns and rows to the bottom and right edges after padding. The network details for local enhancers are shown in Fig. 4. The network architecture of the local enhancer is similar to the global generator, with differences in the number of layers and filters. The first-layer convolutional network of the local enhancer outputs a feature map with 32 channels. After another layer of the down-sampling convolutional network, a feature map with 64 channels is outputted. The fourth layer of the deConv output for the global generator also has a feature map with 64 channels. Therefore, an element-wise sum of these two feature maps can be used as the input of the residual block network in the local enhancer.

B. MULTI-SCALE DISCRIMINATOR

Fig. 5 illustrates the architecture of the multi-scale discriminator. The role of the discriminator determines whether the image is an actual sample or a generated one. However, the classification of high-resolution images is a challenge for GAN. A deep neural network and extensive hardware resources are required for calculation. In the proposed work, the generator that implements the 1024 × 1024 pictures has consumed most of the hardware resources. To overcome this problem, we adopted the conditional GANs network, which

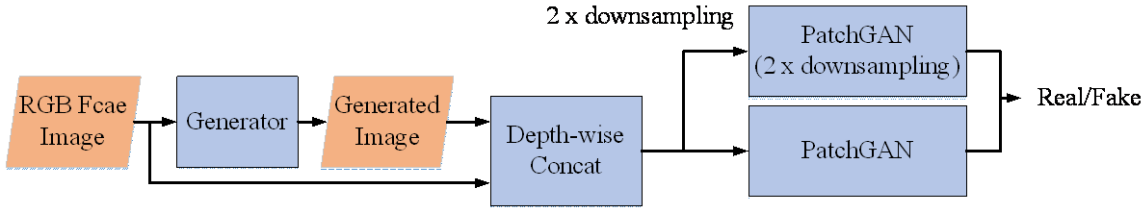

FIGURE 5. Data flow of Multi-Scale discriminator.

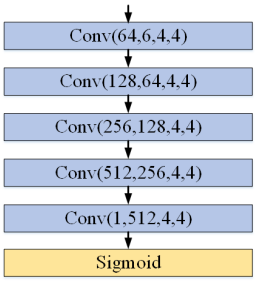

FIGURE 6. The network architecture of discriminator.

uses a multi-scale discriminator. The multi-scale discriminator can be extended to more than two-scale discriminators. In our facial pigmentation distribution image generator, there are two discriminators targeting different image scales (as shown in Fig. 5) with the same network architecture (as shown in Fig. 6). One of the discriminators identifies smaller-scale images to learn the overall integrity of the image and the other is used to distinguish large-scale images to retain the details.

The discriminator applies the PatchGAN from the Pix2Pix framework [31] and outputs an $N \times N$ matrix. Each element in the matrix represents a patch that depicts the probability value of the sample. The average value of the output feature map represents the result for the discriminator. The PatchGAN is responsible for small regions in the input picture, which could realize local image feature extraction and is more conducive for high-resolution image generation.

According to the practical suggestion in [31], we have adopted a patch size of 70 × 70 in our experiment, which also represents the receptive field size of the input image and can be depicted as follows:

$$RF_{N-1} = (RF_N - 1) \times S + K \tag{4}$$

where RF_N is the N^{th} layer receptive field size at the $(N\text{-}1)^{th}$ layer.

C. LOSS FUNCTION

In the conditional GANs, the global generator must be trained before the local enhancer. After training the local enhancer, we fine-tune the parameters for the local enhancer and global generator. The training phase is shown in Fig. 7. The objective function for training the GAN is represented as:

$$min_G\ max_D\ L_{GAN}\ (G, D) \tag{5}$$

$$L_{GAN}\ (G, D) = \left[\log D\ (\boldsymbol{s}, \boldsymbol{x})\right] + \left[\log\left(1 - D(\boldsymbol{s}, G(\boldsymbol{s}))\right)\right] \tag{6}$$

where $\boldsymbol{s}$ and $\boldsymbol{x}$ respectively represent the input image and ground truth, and $G(\boldsymbol{s})$ is the image generated by the generator. The training generator expects the result of the objective function to be as small as possible, which means that the generator tries to generate pictures to deceive the discriminator. When training the discriminator, we need to maximize $\log D(\boldsymbol{s}, \boldsymbol{x})$ and minimize $D(\boldsymbol{s}, G(\boldsymbol{s}))$ for the objective function to distinguish the real and fake pictures properly. The conditional GANs framework adopts the LSGANs (least squares generative adversarial networks) [40] as the loss function, which takes on the value of 1.0 for real images and 0.0 for fake images and is optimized by the MSE (mean squared error). The loss functions for the generator and discriminator are different. The GAN loss for the generator is

$$L_{GAN}(G) = \sum_{i=1}^{T} \frac{1}{N_i} \left[\left(1 - D^i(\boldsymbol{s}, G(\boldsymbol{s}))\right)^2\right] \tag{7}$$

and for discriminator, the GAN loss is

$$L_{GAN}(G) = \sum_{i=1}^{T} \frac{1}{N_i} \left[\left(1 - D^i(\boldsymbol{s}, \boldsymbol{x})\right)^2\right] + L_{GAN}(G) \tag{8}$$

where N_i is the number of the image element. When dealing with high-resolution images, the discriminator needs a deep neural network or a large convolutional kernel to achieve good results, which requires sizable memory for calculation. Therefore, two different scaled discriminators are used to avoid this problem. The first discriminator examines the details in a small-scale image, and the other verifies the original image's completeness. After involving the multi-scale discriminators, the objective function becomes

$$min_G\ max_{D_1, D_2} \sum_{k=1,2} L_{GAN}\ (G, D_k) \tag{9}$$

To ensure that the discriminator distinguishes the different features between real and synthesized pictures, the objective function integrates feature matching loss and VGG19 perceptual loss [20], which are proven suitable for high-resolution image conversion. Feature matching loss sends the generated sample and ground truth to the discriminator separately for feature extraction. VGG19 perceptual loss extracts feature values from generated samples and ground truth. We then apply element-wise loss on the feature map and L1 Loss on

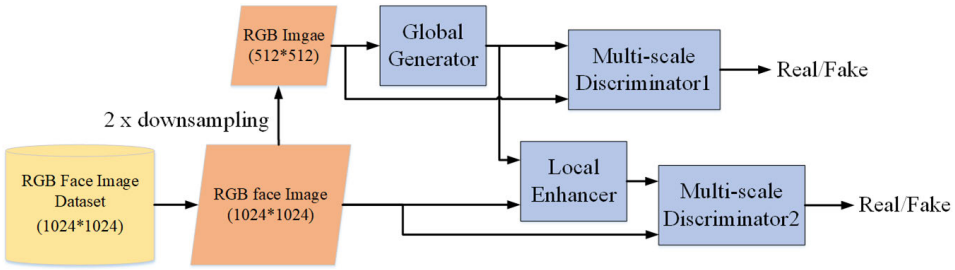

FIGURE 7. Training procedure of the Pix2PixHD.

each discriminator. As a result, the generator would be more stable during the training process, and the feature matching loss is presented as

$$L_{FM}(G, D_k) = \sum_{i=1}^{T} \frac{1}{N_i} \left[\left\| D_k^i(\boldsymbol{s}, \boldsymbol{x}) - D_k^i(\boldsymbol{s}, G(\boldsymbol{s})) \right\|_1 \right] \tag{10}$$

where T is the total number of layers, N_i is the number of elements for the i^{th} layer with layer index i. The VGG19 perceptual loss can be calculated as

$$L_{VGG}(G) = \sum_{i=1}^{N} \frac{1}{M_i} \left[\left\| F^{(i)}(\boldsymbol{x}) - F^{(i)}(G(\boldsymbol{x})) \right\|_1 \right] \tag{11}$$

where $F^{(i)}$ denotes the i^{th} layer with M_i elements of the VGG network. In total, the objective function integrates GAN loss, feature matching loss and VGG19 perceptual loss as follows:

$$min_G \left(\left(max_{D_1,D_2} \sum_{k=1,2} L_{GAN}(G, D_k) \right) + \lambda \sum_{k=1,2} L_{FM}(G, D_k) + \lambda L_{VGG}(G) \right) \tag{12}$$

The feature matching loss and VGG19 loss are responsible for ensuring the content consistency and the GAN loss takes care of the details. The experimental results show that the performance is greatly improved after the inclusion of the feature matching loss and VGG19 loss.

IV. EXPERIMENTAL RESULTS

A. PIGMENTED FACIAL SKIN DATASET

The pigmented facial skin dataset is collected in cooperation with HUANGDERM dermatology. The proposed work systematically organizes and labels each picture's category. The dataset is divided into visible light face images, RBX hemoglobin distribution images, RBX melanin distribution images and cross-polarized skin images. The portraits can be further divided into right profile face, frontal face and left profile face, as shown in Fig. 8. The original dataset has a resolution of 3, 312 × 3.808. This study has collected 14,712 pictures in 3,678 groups in total. The training data set includes 3,000 groups with 12,000 images, and the remaining images are treated as the test data set, which involves 678 groups with 4,712 images. The visible light face images depict the face in the real situation, as shown in Fig. 8 (a), Fig. 8 (e) and Fig. 8 (i). There is no reflection for the polarized light

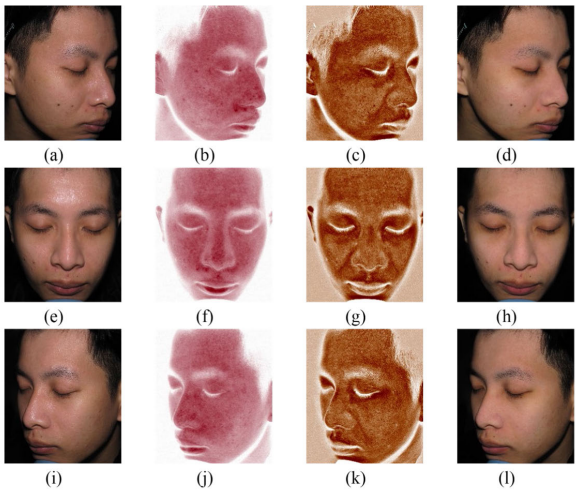

FIGURE 8. The sample images of Pigmented Facial Skin Dataset (a) Right profile original face image, (b) Right profile facial hemoglobin distribution image, (c) Right profile facial melanin distribution image, (d) Right profile facial cross-polarized image, (e) Frontal original face image, (f) Frontal facial hemoglobin distribution image, (g) Frontal facial melanin distribution image, (h) Frontal facial cross-polarized image, (i) Left profile original face image, (j) Left profile facial hemoglobin distribution image, (k) Left profile melanin distribution image, and (l) Left profile facial cross-polarized image.

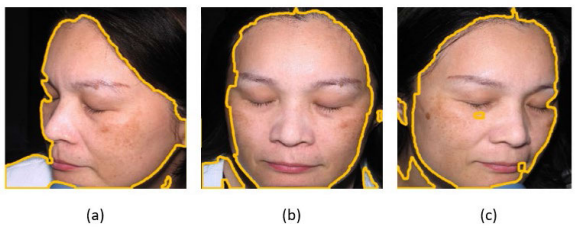

FIGURE 9. Skin detection result (a) Left profile face, (b) Frontal face, and (c) Right profile face.

images, and the details of the face are entirely preserved. Both the hemoglobin and the melanin distribution images were produced by RBX equipment. This work expects to generate facial pigment distribution maps, i.e., RBX-like images from photos taken by digital cameras or mobile phones. Therefore, we have chosen visible light face images as the input instead of cross-polarized light images.

FIGURE 10. Potential melasma region is found through facial features and skin detection, (a) Left profile face, (b) Frontal face, and (c) Right profile face.

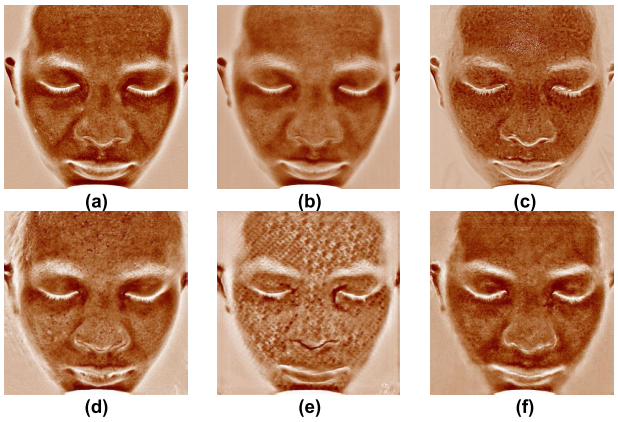

FIGURE 11. Comparison of generated melanin images for different models. (a) Dataset, (b) Proposed, (c) CycleGAN [33], (d) GP-UNIT [35], (e) DECENT [36], and (f) EnCo [37]. The proposed image is more close to the original image than the other models.

TABLE 1. Comparison of melanin for single image generated with different generator.

	Proposed	CycleGAN [33]	GP-UNIT [35]	Decent [36]	EnCo [37]
MSE	531.685	934.063	1044.838	1648.107	694.057
MAE	16.955	21.668	22.935	30.272	19.284
PSNR	20.874	18.427	17.940	15.961	19.717
SSIM	0.280	0.236	0.269	0.252	0.289

B. FACIAL MELASMA ANALYZER

We first transform the facial melanin distribution map into grayscale for melasma detection and apply the median blur filter for noise reduction. The output of the potential melasma region detection module is used as a mask. We then adopt the melasma area and severity index (MASI) [41] to divide the melanin into six levels. The melasma region is then determined with a severity index based on the average skin tone level. The melasma region information would be provided to dermatologists as a diagnosis reference.

1) SKIN DETECTION

Since training the generator is memory-intensive, skin detection is directly implemented through image processing. We apply the multi-color space thresholds [42], which involve three color spaces, i.e., RGB, HSV, and YCbCr, to detect the skin region. The thresholds are set individually, the lower and upper bounds for RGB, HSV, and YCbCr can be represented as:

$$\begin{aligned} &RGB_T\left\{R:108\sim255,\ G:52\sim255,\ B:45\ \sim255\right\}\\ &HSV_T\left\{H:0\sim120,\ S:50\sim150,\ V:0\ \sim255\right\}\\ &YCbCr_T\left\{Y:90\sim230,\ Cb:100\sim120,\ Cr:130\ \sim180\right\} \end{aligned} \tag{13}$$

When a pixel value is within these thresholds, the pixel is treated as a skin color that conforms to all these three color spaces. The skin detection result is presented in Fig. 9.

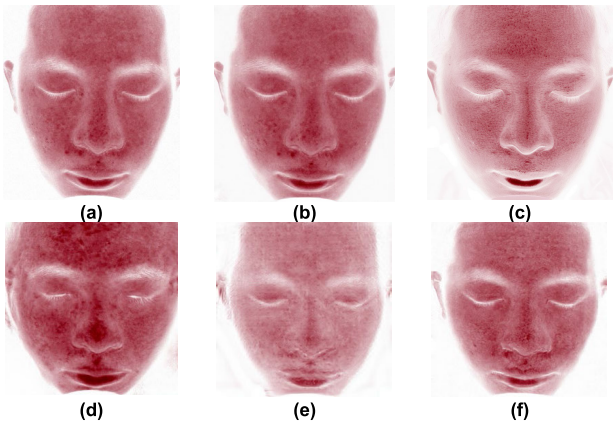

FIGURE 12. Comparison of generated hemoglobin images for different models. (a) Dataset, (b) Proposed, (c) CycleGAN [33], (d) GP-UNIT [35], (e) DECENT [36], and (f) EnCo [37].

TABLE 2. Comparison of hemoglobin for single image generated with different generator.

	Proposed	CycleGAN [33]	GP-UNIT [35]	Decent [36]	EnCo [37]
MSE	142.864	808.100	2077.941	674.463	151.107
MAE	8.780	20.795	33.283	19.242	8.672
PSNR	26.581	19.056	14.954	19.841	26.337
SSIM	0.529	0.491	0.523	0.582	0.599

2) POTENTIAL MELASMA REGION DETECTION

The melasma region is labeled and trained through the proposed system with help from the dermatologist. The melasma is generally distributed on both sides of the face in a C shape. The proposed module adopts the facial landmark detection model from the Dlib library [43] to locate the face before determining the potential melasma region. An additional 81-points facial landmark model [44] is added as the melasma may also appear on the forehead. With the integration of facial landmarks and skin detection, the possible melasma regions are circled and shown in Fig. 10.

3) PERFORMANCE EVALUATION OF GENERATED IMAGE

From the dermatologist's perspective, the original RBX image and the generated RBX-like image have 90% similarity, demonstrating the comparative quality with the Canfield device.

TABLE 3. Average generated result of melanin for test dataset (678 images).

	Proposed	CycleGAN [33]	GP-UNIT [35]	Decent [36]	EnCo [37]
MSE	747.932	1126.322	1604.8187	1477.936	1149.121
MAE	19.852	24.715	29.0750	28.404	24.703
PSNR	19.630	17.739	16.2956	16.531	17.696
SSIM	0.230	0.229	0.2750	0.285	0.312

TABLE 4. Average generated result of hemoglobin for test dataset (678 images).

	Proposed	CycleGAN [33]	GP-UNIT [35]	Decent [36]	EnCo [37]
MSE	151.644	464.306	1006.682	441.923	196.568
MAE	8.545	15.169	21.772	14.413	9.379
PSNR	26.788	21.789	19.147	22.303	25.684
SSIM	0.547	0.488	0.554	0.597	0.654

Furthermore, the proposed system also compared the generated results with the CycleGAN [33], GP-UNIT [35], DECENT [36], and EnCo [37] methods in MSE, MAE, PSNR and SSIM. The MSE and MAS are represented as:

$$MSE = \frac{1}{m}\sum_{i=1}^{m}\left(\boldsymbol{x}_i - \boldsymbol{y}_i\right)^2 \tag{14}$$

$$MAE = \frac{1}{m}\sum_{i=1}^{m}\left|\boldsymbol{x}_i - \boldsymbol{y}_i\right| \tag{15}$$

where $\boldsymbol{x}$ represents the produced image, $\boldsymbol{y}$ is the ground truth, and m is the number of elements in an image. The average PSNR is calculated through:

$$PSNR = 20log_{10}\frac{2^n - 1}{MSE} \tag{16}$$

where n is the bits per sample.

The SSIM is closer to the subjective perception of the human eye. It takes the brightness ($l(x, y)$), contrast ($c(x, y)$) and structure ($s(x, y)$) as indicators, which are given by:

$$SSIM = [l(x, y)]^{\alpha}[c\,(x, y)]^{\beta}[s\,(x, y)]^{\gamma} \tag{17}$$

$$l\,(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \tag{18}$$

$$c\,(x, y) = \frac{2\sigma_x\sigma_y + C_1}{\sigma_x^2 + \sigma_y^2 + C_1} \tag{19}$$

$$s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \tag{20}$$

where $\alpha > 0$, $\beta > 0$ and $\gamma > 0$ are the parameters for adjusting the relative importance of l, c and s. μ_x, μ_y and σ_x, σ_y are the average and variance of $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively. C_1, C_2 and C_3 are constants.

The experimental results of different melanin distribution generators are shown in Fig. 11, where we can see that the proposed image is quite similar to the dataset image and is better than the other models. Although Fig. 11 (f) appears to closely resemble the dataset, it is important to note that the enhanced deeper melanin near the mouth might affect decision-making during diagnosis. The evaluation results are provided in Table 1. The experimental results of different hemoglobin distribution generators are presented in Fig. 12, and the evaluation results are provided in Table 2. Table 3 and Table 4 present the average test results for 678 images. Overall, the proposed system has successfully generated RBX-like images. From the dermatologist's perspective, the generated image has 90% similarity in comparison to the RBX image. Objective comparisons with other network models also showed that the generated images perform better with respect to the MSE, MAE, and PSNR.

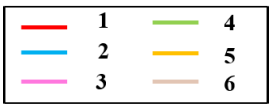

FIGURE 13. Six levels of melanin degree of severity.

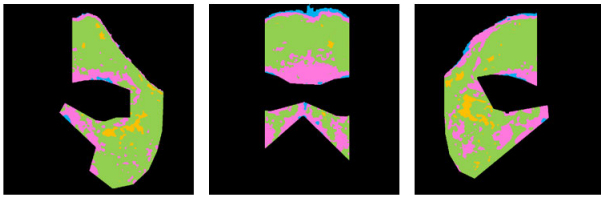

FIGURE 14. Melanin quantification (a) Left profile face (b) Frontal face, and (c) Right profile face.

FIGURE 15. Melasma region on melanin distribution image (a)Left profile face (b)Frontal face (c)Right profile face.

4) MELANIN QUANTIFICATION

This work applies melasma region and severity index (MASI) [41] for melanin quantification. We divide the grayscale values into six levels, i.e., (255, 230), (229, 170), (169, 110), (109, 66), (65, 25) and (24, 1). The levels are represented by different colors, as shown in Fig. 13. According to the quantified face, the proposed system finds the dominant level and treats it as the average skin color of the face. Therefore, the possible melasma region and severity can be obtained when the grayscale value is greater than the average skin color. The result is shown in Fig. 14. The potential melasma regions are further mapped onto the generated RBX-like and the original image to provide more reference, as shown in Fig. 15 and Fig. 16.

FIGURE 16. Melasma region on original image (a)Left profile face (b)Frontal face, and (c)Right profile face.

V. CONCLUSION

This work presented an accurate neural network that generates melanin and hemoglobin images for pigmented facial skin analysis. A pigmented facial skin dataset is collected, with 14,712 images in 3,678 groups. The experimental results show that the generated melanin image has a 90% similarity in comparison with the RBX image and can be adequately used as a reference during diagnosis. Furthermore, the proposed method performs better than other models with regard to the MSE, MAE, PNSR. Overall, the proposed system successfully demonstrates its usefulness in pigmented skin analysis.

REFERENCES

[1] Bhagwat, "The relationship of skin tone to physical and mental health outcomes in South Asian Americans," Ph.D. dissertation, Graduate Program Psychol., Rutgers Univ.-Graduate School-New Brunswick, New Brunswick, NJ, USA, 2012.

[2] N. G. Jablonski, "The evolution of human skin and skin color," *Annu. Rev. Anthropology*, vol. 33, no. 1, pp. 585–623, Oct. 2004.

[3] H. Y. Kang and J.-P. Ortonne, "What should be considered in treatment of melasma," *Ann. Dermatology*, vol. 22, no. 4, p. 373, 2010.

[4] R. Demirli, P. Otto, R. Viswanathan, S. Patwardhan, and J. Larkey. (2007). *RBX® Technology Overview*. [Online]. Available: https://www.canfieldsci.com/FileLibrary/RBX%20tech%20overview-LoRz1.pdf

[5] T. R. Chang and C. Y. Huang, "Skin condition detection based on image processing techniques," in *Proc. ITAOI*, Penghu, Taiwan, 2010.

[6] Z. Wang and R. Li, "Facial pore detection based on characteristics of skin pigment distribution," in *Proc. IEEE ICIP*, Taipei, Taiwan, 2019.

[7] T. N. Tran, K. Drab, and M. Daszykowski, "Revised DBSCAN algorithm to cluster data with dense adjacent clusters," *Chemometric Intell. Lab. Syst.*, vol. 120, pp. 92–96, Jan. 2013.

[8] X. Liu, J. Sun, and X. Wang, "Facial spot contour extraction based on color image processing," in *Proc. ICBIP*, Chengdu, China, 2019.

[9] K. Sanjar, O. Bekhzod, J. Kim, J. Kim, A. Paul, and J. Kim, "Improved U-net: Fully convolutional network model for skin-lesion segmentation," *Appl. Sci.*, vol. 10, p. 3658, May 2020. [Online]. Available: https://www.mdpi.com/2076-3417/10/10/3658

[10] H. Ding, E. Zhang, F. Fang, X. Liu, H. Zheng, H. Yang, Y. Ge, Y. Yang, and T. Lin, "Automatic identification of benign pigmented skin lesions from clinical images using deep convolutional neural network," *BMC Biotechnol.*, vol. 22, no. 1, p. 28, Oct. 2022. [Online]. Available: https://link.springer.com/article/10.1186/s12896-022-00755-5

[11] T. Mendonça, P. M. Ferreira, J. S. Marques, A. R. S. Marcal, and J. Rozeira, "PH2—A dermoscopic image database for research and benchmarking," in *Proc. EMBC*, Osaka, Japan, 2013.

[12] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Sci. Data*, vol. 5, no. 1, pp. 1–9, Aug. 2018.

[13] M. Combalia, N. C. F. Codella, V. Rotemberg, B. Helba, V. Vilaplana, O. Reiter, C. Carrera, A. Barreiro, A. C. Halpern, S. Puig, and J. Malvehy, "BCN20000: Dermoscopic lesions in the wild," 2019, *arXiv:1908.02288*.

[14] B. C. Ennehar, O. Brahim, and T. Hicham, "An appropriate color space to improve human skin detection," *INFOCOMP J. Comput. Sci.*, vol. 9, no. 4, pp. 1–10, Dec. 2010.

[15] G. B. Solanke and S. Gore, "GPU accelerated computing for human skin colour detection using YCbCr colour model," in *Proc. ICCUBEA*, Pune, India, 2017.

[16] K. B. Shaik, P. Ganesan, V. Kalist, B. S. Sathish, and J. M. M. Jenitha, "Comparative study of skin color detection and segmentation in HSV and YCbCr color space," *Proc. Comput. Sci.*, vol. 57, pp. 41–48, 2015.

[17] X. Li, D. Yang, Y. Wang, W. Zhang, F. Li, and W. Zhang, "TCMINet: Face parsing for traditional Chinese medicine inspection via a hybrid neural network with context aggregation," *IEEE Access*, vol. 8, pp. 93069–93082, 2020. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9094648

[18] K. Hashemifard, P. Climent-Perez, and F. Florez-Revuelta, "Weakly supervised human skin segmentation using guidance attention mechanisms," 2023, *arXiv:2302.04625*.

[19] J. P. B. Casati, D. R. Moraes, and E. L. L. Rodrigues, "SFA: A human skin image database based on FERET and AR facial images," in *Proc. IX Workshop de Visao Comput.*, Rio de Janeiro, 2013.

[20] A. Martinez and R. Benavente, "The AR face database," Robot Vision Lab., Purdue Univ., West Lafayette, IN, USA, CVC Tech. Rep., Jan. 1998.

[21] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1090–1104, Oct. 2000.

[22] M. Kawulok, J. Kawulok, and J. Nalepa, "Spatial-based skin detection using discriminative skin-presence features," *Pattern Recognit. Lett.*, vol. 41, pp. 3–13, May 2014.

[23] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," 2018, *arXiv:1803.10417*.

[24] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," 2014, *arXiv:1406.2661*.

[25] Y. Zhou, B. Hu, X. Yuan, K. Huang, Z. Yi, and G. G. Yen, "Multi-objective evolutionary generative adversarial network compression for image translation," *IEEE Trans. Evol. Comput.*, early access, Mar. 3, 2023, doi: 10.1109/TEVC.2023.3261135.

[26] Y. Lu, D. Chen, E. Olaniyi, and Y. Huang, "Generative adversarial networks (GANs) for image augmentation in agriculture: A systematic review," *Comput. Electron. Agricult.*, vol. 200, Sep. 2022, Art. no. 107208.

[27] M.-Y. Liu, X. Huang, J. Yu, T.-C. Wang, and A. Mallya, "Generative adversarial networks for image and video synthesis: Algorithms and applications," *Proc. IEEE*, vol. 109, no. 5, pp. 839–862, May 2021.

[28] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, *arXiv:1411.1784*.

[29] A. Naglah, F. Khalifa, A. El-Baz, and D. Gondim, "Conditional GANs based system for fibrosis detection and quantification in hematoxylin and Eosin whole slide images," *Med. Image Anal.*, vol. 81, Oct. 2022, Art. no. 102537.

[30] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *Proc. CVPR*, Las Vegas, NV, USA, 2016, pp. 3213–3223.

[31] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 5967–5976.

[32] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional GANs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 8798–8807.

[33] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 2242–2251.

[34] M. Ko, E. Cha, S. Suh, H. Lee, J.-J. Han, J. Shin, and B. Han, "Self-supervised dense consistency regularization for image-to-image translation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, New Orleans, LA, USA, Jun. 2022, pp. 18280–18289.

[35] S. Yang, L. Jiang, Z. Liu, and C. C. Loy, "Unsupervised image-to-image translation with generative prior," 2022, *arXiv:2204.03641*.

[36] S. Xie, Q. Ho, and K. Zhang, "Unpaired image-to-image translation with density changing regularization," in *Proc. NeurIPS*, New Orleans, LA, USA, 2022, pp. 1–14. [Online]. Available: https://openreview.net/forum?id=RNZ8JOmNaV4
[37] X. Cai, Y. Zhu, D. Miao, L. Fu, and Y. Yao, "Rethinking the paradigm of content constraints in GAN-based unpaired image-to-image translation," in *Proc. AAAI*, 2024. [Online]. Available: https://focus.zhuanzhi.ai/paper/1bc1ff45cf6c68d80d692d97cbaa7bcd
[38] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. ECCV*, Amsterdam, The Netherlands, 2016.
[39] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.
[40] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Italy, Oct. 2017, pp. 2813–2821.
[41] A. G. Pandya, L. S. Hynan, R. Bhore, F. C. Riley, I. L. Guevara, P. Grimes, J. J. Nordlund, M. Rendon, S. Taylor, R. W. Gottschalk, N. G. Agim, and J.-P. Ortonne, "Reliability assessment and validation of the melasma area and severity index (MASI) and a new modified Masi scoring method," *J. Amer. Acad. Dermatology*, vol. 64, no. 1, pp. 78–83, Jan. 2011.
[42] R. F. Rahmat, T. Chairunnisa, D. Gunawan, and O. S. Sitompul, "Skin color segmentation using multi-color space threshold," in *Proc. 3rd Int. Conf. Comput. Inf. Sci. (ICCOINS)*, Kuala Lumpur, Malaysia, Aug. 2016, pp. 391–396.
[43] D. King. (2002). *Dlib Library*. [Online]. Available: http://dlib.net/
[44] Codeniko. (2019). *81 Facial Landmarks Shape Predictor*. [Online]. Available: https://github.com/codeniko/shape_predictor_81_face_landmarks

AN-CHAO TSAI (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the National Cheng Kung University, Taiwan, in 2010. He is currently an Associate Professor with the International Master Program of Information Technology and Application, National Pingtung University, Pingtung, Taiwan. His research interests include artificial intelligence, virtual reality, and AIoT. He has been serving as the Track Chair and the Program Chair for the IEEE International Conference on Orange Technologies, since 2015.

PATRICK PO-HAN HUANG received the Medical degree in 1993. He completed dermatology residency training in 1997. In 2002, he was named as the Chair of the Department of Dermatology, Chang Gung Memorial Hospital, Kaohsiung, and began private practice in 2006. Since 2002, he has been devoting himself to skin imaging analysis in multidisciplinary collaboration. He is the author of 28 peer-reviewed publications, seven chapters, and seven invention patents in Taiwan, USA, and South Korea. He is an IFAAD and a FAADV.

ZHONG-CHONG WU received the M.S. degree from the Department of Electrical Engineering, National Cheng Kung University, Taiwan, in 2021. He is currently a Hardware Engineer with MediaTek Inc. His research interests include artificial intelligence, machine learning, image processing, image recognition, algorithm design, and pattern recognition.

JHING-FA WANG (Life Fellow, IEEE) is currently the Chair and a Distinguished Professor with the Department of Electrical Engineering, National Cheng Kung University, Taiwan. He has published about 173 journal articles on IEEE, SIAM, IEICE, and IEE; and about 396 conference papers. He has developed a Mandarin speech recognition system called Venus-Dictate, known as a pioneering system in Taiwan. His research interests include speech signal processing, image processing, biomedical signal processing, and VLSI system design. In 1999, he was elected as a fellow of IEEE, for his contribution on "Hardware and Software Co-Design on Speech Signal Processing." He served as the Editor-in-Chief for *International Journal of Chinese Engineering*, from 1995 to 2000.