

# Pathology-Preserving Transformer Based on Multi-Color Space for Low-Quality Medical Image Enhancement

Qingshan Hou, Yaqi Wang, Peng Cao, Jianguo Ju, Huijuan Tu, Xiaoli Liu, Jinzhu Yang, Huazhu Fu, *Senior member, IEEE*, Yih Chung Tham, and Osmar R. Zaiane, *Senior member, IEEE*

**Abstract**—Medical images acquired under suboptimal conditions often suffer from quality degradation, such as low-light, blurring, and artifacts. Such degradations obscure the lesions and anatomical structures in medical images, making it difficult to distinguish key pathological regions. This significantly increases the risk of misdiagnosis by automated medical diagnostic systems or clinicians. To address this challenge, we propose a multi-color space-based quality enhancement network (MSQNet) that effectively eliminates global low-quality factors while preserving pathology-related characteristics for improved clinical observation and analysis. We first revisit the properties of image quality enhancement in different color spaces, where the V-channel in the HSV space can better represent the contrast and brightness enhancement process, whereas the A/B-channel in the LAB space is more focused on the color change of low-quality images. The proposed framework harnesses the unique properties of different color spaces to optimize the image enhancement process. Specifically, we propose a pathology-preserving transformer, designed to selectively aggregate features across different color spaces and enable comprehensive multiscale feature fusion. Leveraging these capabilities, MSQNet effectively enhances low-quality RGB medical images while preserving key pathological features, thereby establishing a new paradigm in medical image enhancement. Extensive experiments on three public medical image datasets demonstrate that MSQNet outperforms traditional enhancement techniques and state-of-the-art methods, in terms of both quantitative metrics and qualitative visual assessment. MSQNet successfully improves image quality while preserving pathological features and anatomical structures, facilitating accurate diagnosis and analysis by medical professionals and automated systems.

**Index Terms**—Low-Quality Medical Image, Quality Enhancement, Multi-Color Space, Pathology-Preserving, Transformer

## I. INTRODUCTION

Medical images are utilized extensively by clinicians and computer-aided diagnostic systems for early disease detection and diagnosis [1], [2], owing to their safety and cost-effectiveness [3]. However, medical images captured under

improper conditions often have significant variations in quality. Low-quality images always suffer from low-contrast, low-brightness, artifacts, etc. [4], [5]. These low-quality factors result in visually poor images and hamper subsequent high-level clinical applications, ranging from disease diagnosis and lesion segmentation to tissue structure detection [6]–[8]. Therefore, it is necessary to enhance low-quality medical images, which not only improves their visual qualities but also benefits clinical observation and analysis.

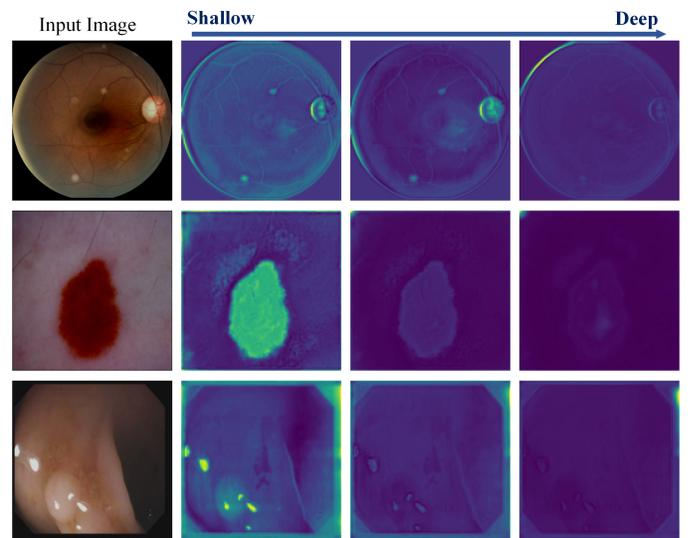


Fig. 1. Illustration of the degradation of lesion characteristics and anatomical structures caused by current image enhancement methods. The input images (left) contain retinal lesions, skin lesions, and polyps. As the enhancement procedure progresses from shallow to deep (from left to right), the lesion features and anatomical details gradually degrade and disappear. This can significantly hinder accurate diagnosis and clinical decision-making.

Recent advancements in deep learning, especially the refined designs of deep neural networks, have shown significant superiority over traditional methods in improving the visual quality of natural and medical images [9]–[13]. Despite these achievements, most deep learning-based approaches [14], [15] rely heavily on aligned low/high-quality image pairs for training. This reliance poses a critical challenge in medical imaging, where obtaining a substantial number of such image pairs in real-world scenarios is impractical due to the limitations in operational procedures and ethical constraints. To alleviate the scarcity of image pairs, several unsupervised learning methods [16], [17] have been proposed. Unfortunately, these

Qingshan Hou\*, Yaqi Wang\*, Peng Cao, Xiaoli Liu, and Jinzhu Yang, are with the School of Computer Science and Engineering, Northeastern University, Shenyang 110000, China. (Corresponding author: Peng Cao, e-mail: caopeng@mail.neu.edu.cn, \* equal contribution.)

Jianguo Ju is with the School of Information Science and Technology, Northwest University, Shaanxi, China.

Huijuan Tu is an associate chief radiologist of Kunshan Hospital of Chinese Medicine, Jiangsu, China.

Huazhu Fu is with the Institute of High-Performance Computing, Agency for Science, Technology, and Research, Singapore.

Qingshan Hou and Yih Chung Tham are with NUS Yong Loo Lin School of Medicine, Singapore.

Osmar R. Zaiane is with Amii, University of Alberta, Edmonton, Canada.

methods often exhibit instability and fail to effectively highlight local details, which are crucial for decision-making in medical imaging. Ideally, high-quality medical images should exhibit uniform illumination, clear lesion visibility, and well-preserved anatomical structures. As shown in Figure 1, existing enhancement methods often degrade pathological features, significantly impairing the visibility of lesions and anatomical structures in the enhanced images. This degradation not only compromises diagnostic reliability but also hinders downstream clinical analysis tasks. Taking all of the above into consideration, the main challenges in enhancing the quality of medical images are as follows:

**1) Data Efficiency:** How can robust quality enhancement be achieved with limited medical image pairs, thus alleviating the data scarcity problem?

**2) Feature Retention:** How can it be ensured that the enhancement process does not degrade lesion characteristics and anatomical structures that are essential for accurate diagnosis?

To address these challenges, this study focuses on bridging the gap between effective image enhancement and reliable clinical applicability. We propose a semi-supervised framework for low-quality medical image enhancement, which not only enhances image quality but also preserves critical lesion features, even in scenarios with limited image pairs. Specifically, inspired by image fusion methods, such as MATR [18], which has shown exceptional performance in medical image fusion by integrating metabolic information from SPECT images with anatomical details from MRI, highlighting its potential for supporting diagnosis and treatment planning. Furthermore, in non-medical domains, the YDTR method [19] effectively integrates complementary features from infrared and visible images using a Y-shape dynamic Transformer, demonstrating its generalization capability and broad applicability across diverse imaging tasks. Considering the color spaces that form part of the human visual system, we first analyze the effects of different color spaces, including LAB and HSV, on image quality enhancement [20], [21]. For the A and B channels of the LAB color space, the distances between colors and perceptual differences are uniform. This property enables the model to control and predict color variations more easily when color enhancements are performed. For the HSV color space, the detailed analysis in Section IV reveals that the H and S channels remain unchanged between the low-quality and enhanced images, indicating that image quality enhancement affects only the V channel. We then integrate the representations and advantages of different color spaces to improve image quality. More specifically, we construct the VAB color space based on the HSV and LAB color spaces of the original RGB image, which more effectively addresses issues such as low-contrast and low-brightness. Finally, to effectively utilize the complementary information from different color spaces (RGB and VAB) while preserving essential pathological features, we propose the pathology-preserving transformer ( $P^2Trans$ ) component.  $P^2Trans$  is designed to selectively aggregate multi-level global features from the VAB space, ensuring that the enhancement process of RGB medical images remains sensitive to clinically significant structures, such as exudates, skin lesion boundaries, polyps,

etc. By integrating  $P^2Trans$  into the enhancement pipeline, the proposed framework achieves a dual objective: improving the visual quality of RGB medical images and preserving essential pathological features for downstream clinical analysis tasks. In other words,  $P^2Trans$  helps establish a novel paradigm in medical image enhancement by aligning visual improvements with diagnostic relevance. Our contributions can be summarized as follows:

1) We propose MSQNet, which leverages multi-scale features from different color spaces to enhance image quality, maintaining pathological features while eliminating low-quality factors.

2) We construct the VAB color space by combining the advantages of LAB and HSV for medical image quality enhancement.

3) The interaction between color spaces integrates global and local representations, providing complementary information. Based on VAB and RGB spaces, a pathology-preserving transformer is designed to perform feature-level interaction and aggregation.

4) Extensive experiments show that MSQNet generates superior quality enhancements and achieves state-of-the-art performance across medical image datasets. Its effectiveness is further validated in downstream tasks, highlighting its broad clinical applicability. The source code for MSQNet is publicly available at <https://github.com/HouQingshan/MSQNet>.

## II. RELATED WORKS

### A. Non-Learning-Based Image Quality Enhancement

The traditional non-learning-based medical image quality enhancement methods rely mainly on hand-crafted priors. For example, contrast-limited adaptive histogram equalization (CLAHE) [22] is widely applied to medical images as an effective image enhancement method. It adjusts the local contrast of an image by redistributing the lightness values of the pixels, resulting in improved visibility of structures in the image. Mitra et al. [23] further enhanced the capabilities of CLAHE by combining it with the Fourier transform to improve the contrast of cataractous color fundus images, which often suffer from poor contrast and hazy appearance. In the field of retinal image enhancement, Cao et al. [24] presented a method for overcoming blurring caused by factors such as refractive medium turbidity and imperfect imaging conditions. Their approach involves a combination of low-pass filtering to remove noise and an  $\alpha$ -rooting operation to enhance the local contrast and sharpness of the retinal images while preserving the global luminance. Similarly, Cheng et al. [25] proposed a structure-preserving guided retinal image filtering (SGRIF) method for restoring fundus images and improving contrast. SGRIF leverages the structural information of retinal images to guide the filtering process, effectively enhancing image quality while maintaining the integrity of retinal structures. In addition, Wang et al. [26] proposed a low-light image enhancement method based on a virtual exposure strategy and image fusion, which generates multiple virtual exposure images and employs multi-scale fusion techniques to effectively enhance image brightness and details.

Despite the effectiveness of these traditional enhancement methods for specific medical images, their reliance on global

image statistics and hand-crafted priors severely limits their applicability and generalizability. These methods often struggle to adapt to the diverse characteristics of medical images acquired under different conditions and from various imaging modalities. Moreover, the hand-crafted nature of these methods requires extensive domain knowledge and manual tuning of parameters, making them less flexible and harder to optimize for specific tasks. As a result, interest in learning-based approaches that can automatically learn the optimal enhancement strategies from large datasets of medical images, potentially overcoming the limitations of traditional methods, is increasing.

### B. Learning-Based Image Quality Enhancement

1) *Unpaired data-based medical image enhancement*: Traditional image enhancement methods rely on paired high- and low-quality image data [27], [28]. For example, Wang et al. [28] proposed a low-light image enhancement framework based on Retinex theory, which decomposes the input image into an illumination map and a reflection map, and adjusts brightness and details separately, significantly improving image quality under low-light conditions. However, numerous challenges are often faced in acquiring such data in clinical settings. In recent years, enhancement techniques based on unpaired data, combined with the advantages of deep learning, have provided a series of novel approaches for addressing the scarcity of paired data. For example, You et al. [29] introduced Cycle-CBAM, which is a CycleGAN-based method for enhancing fundus images. This approach learns to transform low-quality fundus images into high-quality images via unpaired data. Similarly, Ma et al. [16] proposed StillGAN for medical image quality enhancement, which takes into account both structure and illumination indicators and imposes related constraints. However, the cycle consistency constraint of the GAN model has limitations in preserving detailed retinal structures when processing unpaired fundus images, which may compromise the accuracy of medical diagnosis. Park et al. [30] proposed contrastive unpaired translation (CUT), which is an unpaired image translation network that employs contrastive learning to enhance the quality of unpaired image-to-image translation. This approach inspired the development of I-SECRET, a medical image enhancement method introduced by Cheng et al. [31]. I-SECRET combines contrastive learning with a semi-supervised learning framework to guide the enhancement of fundus images based on region importance. In another study, Cheng et al. [17] introduced a medical image enhancement method called LED. This method first learns a degradation mapping from unpaired high-quality images to low-quality images via a data-driven degradation framework. Subsequently, it learns a reverse enhancement process in a paired manner via a conditional diffusion model. Nevertheless, during the learning procedure of degradation mapping, some essential lesion information may be inadvertently degraded, potentially impacting clinical diagnosis.

2) *Synthetic data-based medical image enhancement*: Medical image enhancement has long been constrained by the scarcity of high-quality paired data. To address this predicament,

researchers have begun exploring enhancement techniques based on synthetic data. By integrating advanced algorithms and domain knowledge, these techniques can generate realistic synthetic paired data, opening new avenues for training deep learning models. The application of synthetic data is expected to reduce the reliance on clinical data and accelerate the progress of medical image enhancement techniques. Raj et al. [32] and Shen et al. [33] utilized degradation models that consider factors such as blurring, noise, and uneven illumination to generate synthetic low-quality medical images. These approaches address the scarcity of the high- and low-quality medical image pairs required for training deep learning frameworks. However, medical image enhancement methods based on synthetic data still have limitations, such as the differences between synthetic and real data, the difficulty in preserving pathological features, and the simplification of the degradation process. Li et al. [15] presented the structure-consistent restoration network (SCR-Net) to address quality degradation in fundus images of cataract patients. By leveraging synthetic data for training, SCR-Net overcomes the scarcity of paired high- and low-quality fundus images. This approach enables the network to effectively improve the visual quality of fundus images obtained from cataract patients, resulting in clearer and more detailed representations of the retinal structures. Liu et al. [34] introduced the pyramid constraint enhancement network (PCE-Net) to address the issues associated with quality deterioration in retinal fundus images. The incorporation of a pyramid constraint enables PCE-Net to provide reliable image enhancement in clinical settings, mitigating the reliance on large volumes of clinical data. While the aforementioned methods have made significant strides in improving overall image quality, they fall short in effectively addressing the critical issue of pathological feature degradation.

### III. VAB COLOUR SPACE INVESTIGATION

Using a fundus image as an example, this section demonstrates the necessity of the VAB color space in preserving lesion structures and details by analyzing its intensity histogram in both the RGB and VAB color spaces. We focus on comparing the histogram distribution characteristics of the two color spaces in this fundus image, thereby revealing the unique advantages of the VAB space.

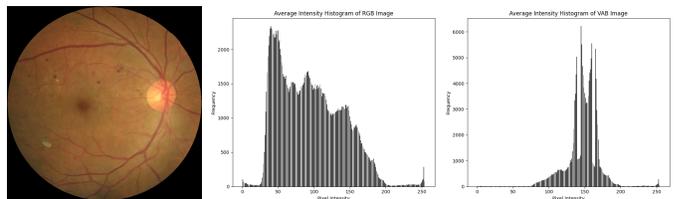


Fig. 2. Comparative analysis of a color fundus image (left) using intensity histograms in the RGB space (middle) and VAB space (right) highlights the superiority of VAB in capturing and preserving pathological details.

As shown in Figure 2(middle), the histogram in the RGB space has a distinct peak in the low-intensity range (0-50), indicating the presence of many dark pixels in the image. This

may be due to the imaging conditions and uneven illumination of the fundus image. Although the RGB space can reflect the overall brightness distribution of the image, the darker regions may obscure important lesion details, which is not conducive to further analysis and diagnosis. In contrast, the histogram in the VAB space (Figure 2 (right)) exhibits a more balanced distribution characteristic. Although the number of pixels in the low-intensity range (0-50) is lower in the VAB space than in the RGB space, the VAB space clearly has advantages in the medium- to high-intensity range (50-200), with a richer pixel distribution and more prominent detail information. In Particular, in the medium-intensity range (120-180), there are significant peak regions in the VAB histogram, which are speculated to correspond to important structures or features related to lesions. This finding indicates that the VAB space may effectively separate the key information in the fundus image and reduce the interference of factors such as illumination.

The above preliminary analysis demonstrates that, compared with the RGB space, the VAB color space has unique advantages in representing pathological fundus images. It can better preserve and highlight the detailed features of lesion areas at different intensity levels, providing more valuable information for computer-aided diagnosis of fundus diseases.

#### IV. METHODOLOGY

Before introducing our framework, we first define the notations used in this paper. The training dataset  $\mathcal{D} = \{X, \mathcal{X}\}$  contains two subsets: subset  $X = \{(X_{RGB}^i, R_{RGB}^i)_{i=1}^N\}$  with reference images and subset  $\mathcal{X} = \{(\mathcal{X}_{RGB}^i)_{i=1}^M\}$  without reference images. Here,  $X_{RGB}^i$  and  $\mathcal{X}_{RGB}^i$  represent the  $i_{th}$  images in subsets  $X$  and  $\mathcal{X}$ , respectively, whereas  $R_{RGB}^i$  is the high-quality reference image corresponding to  $X_{RGB}^i$ . The variables  $N$  and  $M$  represent the number of samples in subsets  $X$  and  $\mathcal{X}$ , respectively. Given the medical image training set  $\mathcal{D}$ , MSQNet is trained to learn a model  $f(\cdot; \theta)$  that performs quality enhancement on medical image test sets. Table I provides a summary of the key notations used in this study.

##### A. Overview of the MSQNet Framework

Low-quality medical images often contain various types of quality defects that interfere with the accurate identification of lesions and anatomical structures. However, most existing image quality enhancement methods suffer from severe feature degradation during processing, making it difficult to achieve an effective balance between improving image quality and preserving important details. This imbalance can lead to the loss of key detail information in the processed images or the over-enhancement of artifacts. To better balance the relative relationship between quality enhancement and detail restoration during training, we consider the advantages of different color spaces and integrate the complementary information from these spaces. With this motivation, we propose MSQNet for low-quality medical image enhancement, which consists of two main stages: 1) construction of the VAB color space, and 2) dual-branch image quality enhancement. Figure 3 shows the pipeline of the MSQNet framework.

TABLE I  
DESCRIPTION OF IMPORTANT NOTATIONS

Notation	Description
<b>Image notations</b>	
$X_{RGB}^i, \hat{X}_{RGB}^i$	Low-quality RGB image with reference and corresponding enhanced one;
$R_{RGB}^i$	The reference of $X_{RGB}^i$ ;
$\mathcal{X}_{RGB}^i, \hat{\mathcal{X}}_{RGB}^i$	Low-quality RGB image without reference and corresponding enhanced one
$X_{VAB}^i, \hat{X}_{VAB}^i$	Low-quality VAB image with reference and corresponding enhanced one;
$R_{VAB}^i$	The reference of $X_{VAB}^i$ ;
$\mathcal{X}_{VAB}^i, \hat{\mathcal{X}}_{VAB}^i$	Low-quality VAB image without reference and corresponding enhanced one
<b>Discriminators</b>	
$D_{RGB}$	Discriminator for distinguishing between $\hat{X}_{RGB}^i$ and $R_{RGB}^i$ ;
$D_{VAB}$	Discriminator for distinguishing between $\hat{\mathcal{X}}_{VAB}^i$ and $R_{VAB}^i$
<b>Feature map notions</b>	
$\mathcal{F}_{RGB}^i$	Low-quality factors in low-quality RGB image;
$F_{RGB}^i$	Feature maps of low-quality RGB image;
$F_{VAB}^i$	Feature maps of low-quality VAB image;
$M_S^i$	Similarity matrix between Q and K;
$V_A^i, V_C^i$	Attention vector, and Critical similarity vector;
$F_C^i/F_{C'}^i, F_A^i$	Critical VAB feature map, and Aggregated feature map of $F_{RGB}^i$ and $F_{C'}^i$ ;
$F_T^i$	Fused feature map of $F_{RGB}^i$ and $F_{VAB}^i$
<b>Key operations</b>	
$Unfold(\cdot), Fold(\cdot)$	Unfold and Fold operations;
$Gt(\cdot), Rp(\cdot), Ep(\cdot)$	Gather, Reshape, and Expansion operations

##### B. Construction of the VAB Color Space

To integrate the advantages and representations of different color spaces and further construct the VAB color space, we analyze the effects of the HSV and LAB color spaces on the quality enhancement task.

**Effect of the HSV Color Space on Quality Enhancement.** According to the Retinex model, the relationship between the enhanced image  $\hat{X}_{RGB}^i$  and the low-quality image  $X_{RGB}^i$  can be represented as follows:

$$\hat{X}_{RGB}^i = X_{RGB}^i / (\mathcal{F}_{RGB}^i + \delta) \quad (1)$$

where  $\delta$  indicates a small constant to prevent the denominator from being zero and  $\mathcal{F}_{RGB}^i$  represents the low-quality factor in  $X_{RGB}^i$ .

The low-quality RGB image can be decomposed into three separate channels based on the pixel values:

$$\begin{cases} L = \max_{c \in \{R, G, B\}} X_{RGB}^i(p^c) \\ M = \text{median}_{c \in \{R, G, B\}} X_{RGB}^i(p^c) \\ S = \min_{c \in \{R, G, B\}} X_{RGB}^i(p^c) \end{cases} \quad (2)$$

where  $p$  denotes an individual pixel. Furthermore, the low-quality image  $X_{RGB}^i$  can be represented in the HSV color space as follows:

$$\begin{cases} V(X_{RGB}^i) = L \\ S(X_{RGB}^i) = (L - S)/L \\ H(X_{RGB}^i) = C_1 + C_2(M - S)/(L - S) \end{cases} \quad (3)$$

where  $V(X_{RGB}^i)$ ,  $S(X_{RGB}^i)$ , and  $H(X_{RGB}^i)$  represent the V, S, and H channels of the low-quality image  $X_{RGB}^i$ , respectively.  $C_1$  and  $C_2$  indicate constant values. Based on equation 1, the enhanced image  $\hat{X}_{RGB}^i$  in HSV space can be expressed as equation 4. For clarity, we omit the constant  $\delta$  from equation 1 and simply use  $\mathcal{F}_{RGB}^i$  to represent  $\mathcal{F}_{RGB}^i + \delta$ .

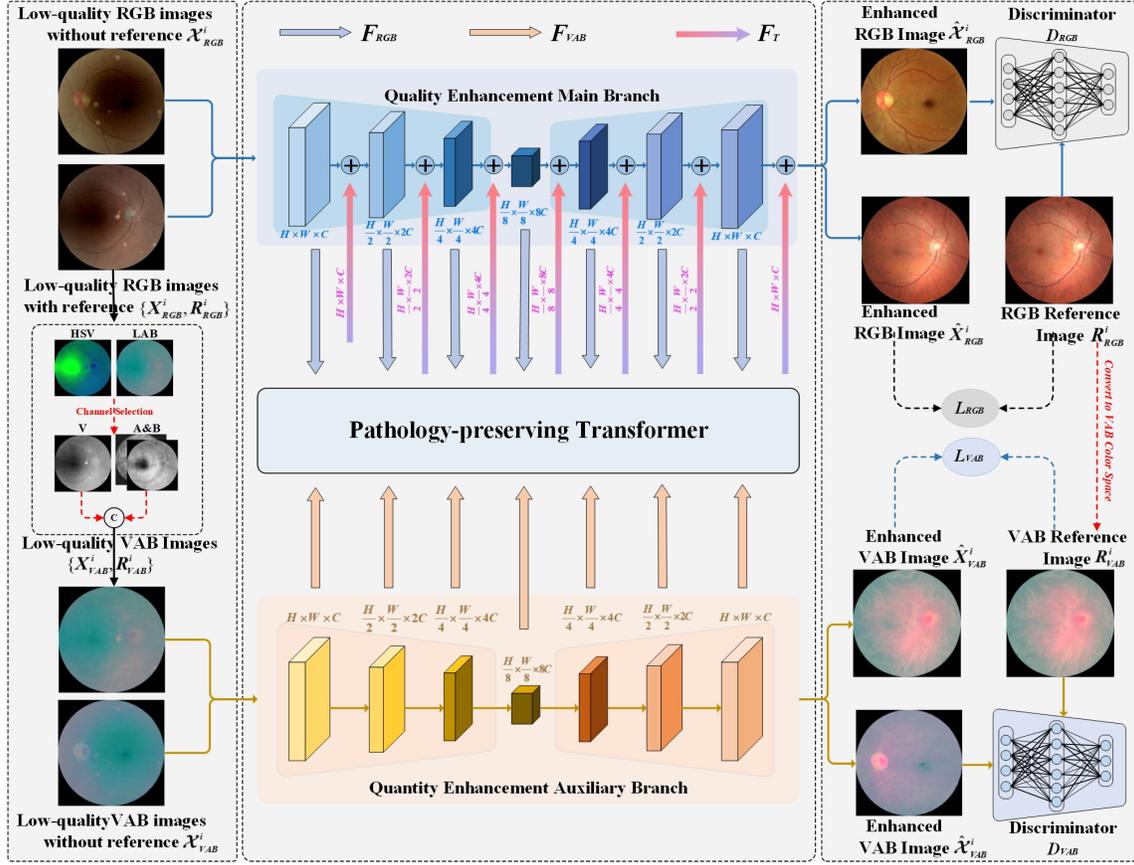


Fig. 3. Overview of the MSQNet framework. Generally, the training of MSQNet consists of two main stages. In the first stage, a new color space, VAB, is constructed by considering the advantages of multi-color spaces. Then, the low-quality RGB images  $\mathcal{X}_{RGB}^i$  and  $X_{RGB}^i$  and RGB reference image  $R_{RGB}^i$  are transformed into the low-quality VAB images  $\mathcal{X}_{VAB}^i$  and  $X_{VAB}^i$  and VAB reference image  $R_{VAB}^i$ . In the second stage,  $\mathcal{X}_{RGB}^i/X_{RGB}^i$  are fed into the main branch for original RGB image quality enhancement, whereas  $\mathcal{X}_{VAB}^i/X_{VAB}^i$  are fed into an auxiliary branch for guidance information extraction from the VAB color space. A pathology-preserving transformer component is employed to perform multi-scale feature interaction and aggregation between the VAB and RGB color spaces.

$$\begin{cases} V(\hat{X}_{RGB}^i) = L/\mathcal{F}_{RGB}^i = V(X_{RGB}^i)/\mathcal{F}_{RGB}^i \\ S(\hat{X}_{RGB}^i) = \frac{(L/\mathcal{F}_{RGB}^i - S/\mathcal{F}_{RGB}^i)}{(L/\mathcal{F}_{RGB}^i)} = S(X_{RGB}^i) \\ H(\hat{X}_{RGB}^i) = C_1 + \frac{C_2(M/\mathcal{F}_{RGB}^i - S/\mathcal{F}_{RGB}^i)}{(L/\mathcal{F}_{RGB}^i - S/\mathcal{F}_{RGB}^i)} = H(X_{RGB}^i) \end{cases} \quad (4)$$

where  $V(\hat{X}_{RGB}^i)$ ,  $S(\hat{X}_{RGB}^i)$ , and  $H(\hat{X}_{RGB}^i)$  represent the V, S, and H channels of the enhanced image  $\hat{X}_{RGB}^i$ , respectively. We can observe that the H and S channels remain consistent between the low-quality image  $X_{RGB}^i$  and the enhanced image  $\hat{X}_{RGB}^i$ . The quality enhancement focuses exclusively on the V channel, resulting in a positive effect.

#### Effect of the LAB Color Space on Quality Enhancement.

In contrast to the RGB color space, the LAB color space presents the following advantages [20]. The A and B channels in the LAB color space are designed to align more closely with the human visual system's perception. This ensures that color adjustments made in this space do not introduce additional color loss, which is crucial for medical image quality enhancement tasks. Moreover, the LAB color space separates luminance and color information, with the L channel representing luminance and the A and B channels representing

color. This separation makes it significantly easier to perform precise color adjustments and achieve optimal color balance via the A and B channels.

As illustrated in Figure 3, by combining the advantages of both the HSV and LAB color spaces for quality enhancement, we ultimately construct the VAB color space to guide the quality enhancement of RGB medical images.

#### C. Dual-branch image quality enhancement

To obtain a higher-quality medical image  $\hat{X}_{RGB}^i$  from a low-quality medical image  $X_{RGB}^i$ , we design an end-to-end quality enhancement framework, which includes a main quality enhancement branch, an auxiliary branch for extracting guidance information from the VAB color space, and a pathology-preserving transformer component for encouraging multi-scale feature interaction and aggregation.

As shown in Figure 3, low-quality RGB images  $\mathcal{X}_{RGB}^i$  and  $X_{RGB}^i$  are accepted as inputs to train the main quality enhancement branch in a semi-supervised manner. The loss  $\mathcal{L}_m$  of the main branch is formulated as follows:

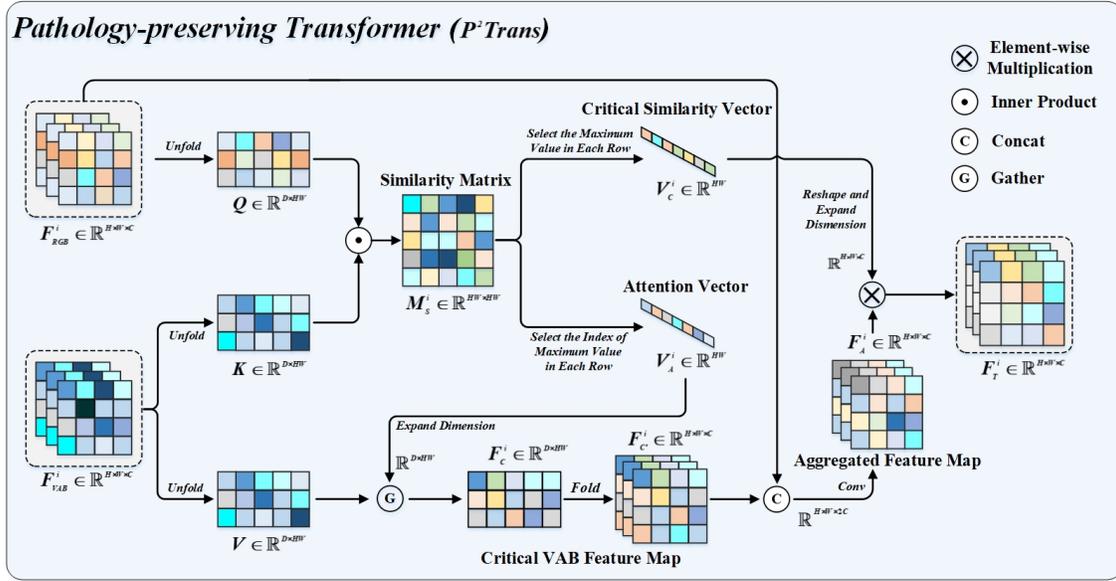


Fig. 4. Illustration of the pathology-preserving transformer component.

$$\begin{aligned}
 \mathcal{L}_m = & \lambda_1 \mathbb{E}_{X_{RGB}} [\|R_{RGB} - f_m(X_{RGB} | \theta_{RGB})\|_1] \\
 & + \mathbb{E}_{R_{RGB}} [\log D_{RGB}(R_{RGB})] \\
 & + \mathbb{E}_{X_{RGB}} [\log(1 - D_{RGB}(f_m(X_{RGB} | \theta_{RGB})))]
 \end{aligned} \quad (5)$$

where  $f_m(\cdot; \theta_{RGB})$  represents the mapping function of the main branch quality enhancement with parameter  $\theta_{RGB}$ ,  $D_{RGB}(\cdot)$  represents a discriminator for distinguishing between the enhanced image  $\hat{X}_{RGB}^i$  and the reference image  $R_{RGB}^i$ , and  $\lambda_1$  and  $\lambda_2$  are the regularization weights that balance the losses of the main branch.

Given the transformed low-quality VAB images  $\hat{X}_{VAB}^i$  and  $X_{VAB}^i$ , as well as the reference VAB image  $R_{VAB}^i$ , as inputs, the auxiliary branch is jointly trained to produce the enhanced VAB images  $\hat{X}_{VAB}^i$  and  $\hat{X}_{VAB}^i$  and to extract guidance information. The loss  $\mathcal{L}_a$  of the auxiliary branch can be formulated as follows:

$$\begin{aligned}
 \mathcal{L}_a = & \lambda_2 \mathbb{E}_{X_{VAB}} [\|R_{VAB} - f_a(X_{VAB} | \theta_{VAB})\|_1] \\
 & + \mathbb{E}_{R_{VAB}} [\log D_{VAB}(R_{VAB})] \\
 & + \mathbb{E}_{X_{VAB}} [\log(1 - D_{VAB}(f_a(X_{VAB} | \theta_{VAB})))]
 \end{aligned} \quad (6)$$

where  $f_a(\cdot; \theta_{VAB})$  is the mapping function of the auxiliary branch with parameter  $\theta_{VAB}$ . Combining all the losses, the final loss for MSQNet can be defined as follows:

$$\mathcal{L}_{MSQNet} = \mathcal{L}_m + \lambda_3 \mathcal{L}_a \quad (7)$$

Owing to the auxiliary branch possessing a stronger ability to remove specific low-quality factors (e.g., low-light and artifacts) and maintain more complete biomarkers than the main branch dose, we design a pathology-preserving transformer ( $P^2Trans$ ) component to assist the main branch in learning specific low-quality representations more effectively from the VAB color space. As shown in Figure 4, we take the feature maps  $F_{RGB}^i \in \mathbb{R}^{H \times W \times C}$  and  $F_{VAB}^i \in \mathbb{R}^{H \times W \times C}$ ,

inherited from the main and auxiliary branches, as inputs of the  $P^2Trans$  component. The query ( $Q$ ), key ( $K$ ), and value ( $V$ ) matrices are obtained by applying the *unfold* operation to the feature maps  $F_{RGB}^i$  and  $F_{VAB}^i$  from the main and auxiliary branches, respectively. Specifically,  $Q$  is derived from  $F_{RGB}^i$ , whereas both  $K$  and  $V$  are derived from  $F_{VAB}^i$ , i.e.,  $Q = Unfold(F_{RGB}^i) \in \mathbb{R}^{D \times HW}$ , and  $K = V = Unfold(F_{VAB}^i) \in \mathbb{R}^{D \times HW}$ . Specifically, we first obtain the similarity matrix  $M_S^i \in \mathbb{R}^{HW \times HW} = Q \odot K$  by calculating the relevance between  $Q$  and  $K$ . To obtain the anatomical structure features without low-quality factors from the auxiliary branch, we calculate the critical VAB feature map  $F_{C'}^i \in \mathbb{R}^{H \times W \times C}$  via equation 8.

$$F_{C'}^i = Fold(F_C^i) = Fold(Gt(V, Ep(V_A^i))) \quad (8)$$

where  $V_A^i$  is the attention vector obtained via index selection on  $M_S^i$ , which contains the most relevant positions of the auxiliary branch with respect to the main branch.  $Ep(\cdot)$  represents the dimension expansion operation performed to make the dimensions of  $V_A^i$  and  $V$  consistent.  $Gt(\cdot)$  denotes the gather operation used to obtain the unfolded critical VAB feature map  $F_C^i$ . Furthermore, we concatenate  $F_{C'}^i$  and  $F_{RGB}^i$  and feed them into a convolutional layer to obtain the aggregated feature map  $F_A^i$ . In the end, we fuse the complementary information from the two branches in our framework by combining  $F_A^i$  with the critical similarity vector  $V_C^i$  obtained through maximum selection on  $M_S^i$ . The final output  $F_T^i$  of the pathology-preserving transformer block can be represented as follows:

$$F_T^i = F_A^i \otimes Ep(Rp(V_C^i)) \quad (9)$$

where  $\otimes$  indicates the element-wise multiplication, and  $Rp(\cdot)$  and  $Ep(\cdot)$  denote the reshaping and dimension expansion operations, respectively.

## V. EXPERIMENTS

### A. Datasets and Performance Metrics

In our experiments, we evaluate the effectiveness of MSQNet by comparing it against existing quality enhancement methods on four different types of medical image benchmark datasets:

(1) **Retinal fundus dataset:** The retinal fundus dataset from the EyeQ, DDR, IDRiD, and DRIVE datasets, which includes 8,347 synthesized paired images and 1,876 unpaired low-quality fundus images. We randomly split 10% from the training dataset for validation. The testing dataset includes 4,559 low-quality images. Importantly, synthetic paired images are generated from the EyeQ dataset via the algorithm introduced by Shen et al. [33].

(2) **Messidor dataset:** The Messidor dataset [35] contains a total of 1200 fundus images, sourced from three different ophthalmology departments. Each image is provided with a label indicating the severity of diabetic retinopathy (DR). Similarly, the synthetic images are generated using the algorithm proposed by Shen et al. [33].

(3) **Skin lesion dataset:** The skin lesion dataset is constructed based on the ISIC Challenge Dataset 2017 [36], which includes 2,750 images. All the skin lesion images are randomly divided into three partitions: 1,679 for training, 186 for validation, and 885 for testing.

(4) **Endoscopy dataset:** Specular highlights and uneven illumination can seriously affect the visual quality of endoscopic images. Three clinicians selected 912 images based on the work of Bernal et al. [37] from the public CVC-EndoSceneStill dataset [38] for endoluminal scene enhancement. We randomly split these images into 450, 100, and 362 images for training, validation, and testing, respectively.

**Performance Metrics.** Two types of evaluation metrics are selected to quantify the quality enhancement performance of the relevant datasets. Specifically, the peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) are used as full-reference metrics to evaluate the differences between enhanced images and high-quality reference images. Additionally, the image quality assessment score [31] is adopted as a non-reference metric to focus on the overall quality enhancement effect across various medical datasets.

### B. Implementation Details

The proposed MSQNet framework is implemented via the PyTorch library, and all the experiments are conducted on a hardware platform equipped with four NVIDIA Quadro RTX GPUs, each with 24 GB of memory. All training images are resized to  $512 \times 512$ , followed by combinations of horizontal flipping, random rotation, and vertical flipping. All the considered methods are trained from scratch. We train MSQNet via the Adam optimizer with a *weight decay* of 0.1 and an *initial learning rate* of  $1 \times 10^{-4}$ . The MSQNet is trained for 150 epochs with a *batch size*=32. Learning rate warmup is adopted in the first 60 epochs, and then the cosine annealing schedule is applied in the following epochs to automatically adjust the learning rate. The weight parameters in  $\mathcal{L}_m$ ,  $\mathcal{L}_a$  and  $\mathcal{L}_{MSQNet}$  are experimentally set to  $\lambda_1 = \lambda_2 = 5$ , and

$\lambda_3 = 1$ . For the loss  $\mathcal{L}_m/\mathcal{L}_a$  of the main/auxiliary branch, the weight coefficient  $\lambda_1/\lambda_2$  effectively controls the content consistency between the enhanced RGB/VAB images and their corresponding RGB/VAB reference images. As a result,  $\lambda_1/\lambda_2$  should maintain a certain value to ensure alignment between the enhanced and reference images.

### C. Comparison with State-of-the-Art Methods

To demonstrate the effectiveness of MSQNet, we compare it with representative state-of-the-art medical image quality enhancement methods on various datasets. The comparable methods include the following: 1) Three traditional methods, e.g. LIME [39], latent structure-drive [41], and distribution fitting [40]. 2) Supervised deep learning-based methods, e.g. SCR-Net [15], PCE-Net [34], ArcNet [42], cGAN [43], and RFormer [14]. 3) A semi-supervised deep learning-based methods, e.g. I-SECRET [31]. 4) Unsupervised deep learning-based methods, e.g. StillGAN [16], CutGAN [30], LED [17], and Cycle-CBAM [29].

As reported in Table II, poor quality enhancement performance is observed when traditional methods are used. Compared with the traditional methods, considerable improvements are obtained by all deep learning-based methods. According to the full-reference metrics, the proposed MSQNet framework demonstrates superior quality enhancement performance and generalizability, consistently outperforming all other traditional and deep learning-based quality enhancement methods. MSQNet consistently outperforms the previous best method, I-SECRET, across all three datasets. On the retinal fundus dataset, MSQNet achieves a PSNR of 30.47 and an SSIM of 0.937, which exceed those of I-SECRET by 3.11 and 0.029, respectively. Similarly, on the skin lesion dataset, MSQNet obtains a PSNR of 30.59 and an SSIM of 0.906, outperforming I-SECRET by 2.06 and 0.014, respectively. Finally, on the endoscopy Dataset, MSQNet achieves a PSNR of 35.37, which exceeds that of I-SECRET by 2.69. The overall results imply that the proposed MSQNet framework can better enhance the quality of different medical image datasets than can I-SECRET in terms of full-reference metrics. In terms of the non-reference metric IQAS, it is also observed that our method outperforms all other comparable methods, implying that the proposed MSQNet framework can better enhance the overall quality of different medical image datasets. On the retinal fundus dataset, MSQNet achieves an IQAS of 0.698, which exceeds that of I-SECRET by 0.034. On the skin lesion dataset, MSQNet achieves an IQAS of 0.724, outperforming I-SECRET by 0.021. On the endoscopy dataset, MSQNet achieves an IQAS of 0.776, outperforming I-SECRET by 0.042, even with a training set of only 450 images. These results demonstrate the superiority of MSQNet for enhancing low-quality images with limited data, and its ability to leverage unlabeled data effectively across various medical image domains.

Figure 5 presents the qualitative comparisons on different types of medical image datasets. It can be observed that the low-quality images suffer from various issues such as uneven illumination, blurry details, and poor contrast. Deep learning-based methods such as PCE-Net, ArcNet and Cycle-CBAM

TABLE II  
COMPARISON WITH STATE-OF-THE-ART QUALITY ENHANCEMENT METHODS ON THREE BENCHMARK DATASETS.

Methods	Types	Retinal Fundus Dataset			Skin Lesion Dataset			Endoscopy Dataset		
		PSNR↑	SSIM↑	IQAS↑	PSNR↑	SSIM↑	IQAS↑	PSNR↑	SSIM↑	IQAS↑
LIME [39]	Traditional	13.54	0.868	0.346	12.93	0.824	0.387	14.86	0.854	0.425
Fu et al. [40]	Traditional	9.76	0.564	0.235	10.24	0.633	0.312	11.75	0.692	0.397
He et al. [41]	Traditional	15.56	0.759	0.368	16.32	0.732	0.425	17.32	0.776	0.473
SCR-Net [15]	Supervised	26.37	0.876	0.653	26.72	0.882	0.675	31.64	0.923	0.745
PCE-Net [34]	Supervised	22.44	0.872	0.573	24.70	0.895	0.582	23.63	0.803	0.618
ArcNet [42]	Supervised	21.24	0.796	0.562	20.23	0.856	0.549	22.05	0.856	0.663
cGAN [43]	Supervised	26.35	0.894	0.634	26.38	0.851	0.645	27.33	0.908	0.704
RFormer [14]	Supervised	24.37	0.862	0.607	25.36	0.865	0.614	26.42	0.874	0.636
StillGAN [16]	Unsupervised	25.38	0.896	0.619	25.27	0.864	0.627	29.25	0.916	0.726
CutGAN [30]	Unsupervised	22.76	0.872	0.576	25.32	0.836	0.597	26.45	0.854	0.659
LED [17]	Unsupervised	26.23	0.860	0.658	27.58	0.822	0.684	32.69	0.924	0.742
Cycle-CBAM [29]	Unsupervised	21.56	0.843	0.534	24.26	0.874	0.536	25.27	0.832	0.627
I-SECRET [31]	Semi-supervised	27.36	0.908	0.664	28.53	0.892	0.703	32.68	<b>0.931</b>	0.734
MSQNet(Ours)	Semi-supervised	<b>30.47</b>	<b>0.937</b>	<b>0.698</b>	<b>30.59</b>	<b>0.906</b>	<b>0.724</b>	<b>35.37</b>	0.926	<b>0.776</b>

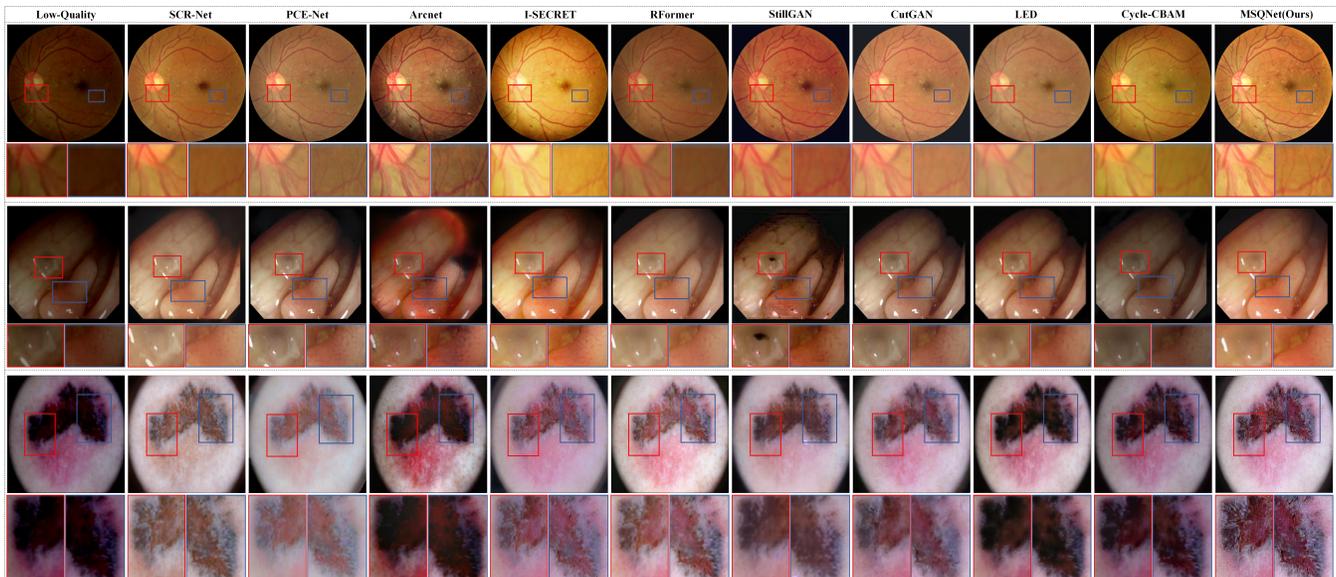


Fig. 5. Visual comparisons of medical image enhancement between MSQNet and other comparable SOTA methods.

exhibit limited ability to improve the uneven illumination of low-quality endoscopy images, resulting in unsatisfactory enhancement results. I-SECRET, SCR-Net, and LED yield over-smoothed results and fail to recover clear skin lesion boundaries, leading to a loss of important diagnostic details. StillGAN, on the other hand, yields over-enhanced results and fails to restore more fine-grained content and structural details of endoscopy images, which may hinder accurate interpretation.

In contrast, the proposed MSQNet demonstrates superior performance in enhancing the quality of various medical images. For the endoscopy images, MSQNet effectively improves the illumination uniformity, contrast, and sharpness while preserving the essential structural details. In the case of skin lesion images, MSQNet successfully restores clear lesion boundaries and enhances fine-grained textures, which are crucial for accurate diagnosis. The retinal fundus images enhanced by MSQNet also exhibit improved clarity, contrast, and visibility

of important anatomical structures. These improvements can be attributed to MSQNet’s ability to encourage the encoder to learn more fine-grained and structural details by leveraging unidirectional feature guidance derived from the VAB space. The integration of multi-scale contextual information further enables MSQNet to capture both local and global features effectively while alleviating the feature degradation of lesions and anatomical structures. Consequently, low-quality medical images can be significantly enhanced, facilitating better visual interpretation and potentially aiding in accurate diagnosis. To further evaluate the generalization of MSQNet, we also perform 5-fold cross-validation on the Messidor dataset. The experimental results on the Messidor dataset are summarized in Table III. MSQNet outperforms other comparable methods, achieving the best performance across different metrics, such as PSNR ( $29.82 \pm 0.47$ ), SSIM ( $0.908 \pm 0.019$ ), and IQAS (0.706). These results further demonstrate the generalization of MSQNet compared to other methods on different fundus

datasets.

TABLE III  
PERFORMANCE COMPARISON OF STATE-OF-THE-ART METHODS ON THE MESSIDOR DATASET USING 5-FOLD CROSS-VALIDATION.

Method	PSNR $\uparrow$	SSIM $\uparrow$	IQAS $\uparrow$
LIME [39]	12.36	0.852	0.338
Fu et al. [40]	11.32	0.573	0.263
He et al. [41]	16.23	0.764	0.374
SCR-Net [15]	26.45 $\pm$ 0.16	0.879 $\pm$ 0.008	0.641
PCE-Net [34]	24.53 $\pm$ 0.36	0.859 $\pm$ 0.023	0.582
ArcNet [42]	22.94 $\pm$ 0.26	0.848 $\pm$ 0.012	0.556
cGAN [43]	26.21 $\pm$ 0.23	0.871 $\pm$ 0.015	0.625
RFormer [14]	23.56 $\pm$ 0.08	0.862 $\pm$ 0.004	0.613
StillGAN [16]	26.18 $\pm$ 0.45	0.865 $\pm$ 0.035	0.649
CutGAN [30]	24.82 $\pm$ 0.21	0.887 $\pm$ 0.012	0.617
LED [17]	26.08 $\pm$ 0.21	0.868 $\pm$ 0.029	0.652
CycleCBAM [29]	23.29 $\pm$ 0.37	0.853 $\pm$ 0.027	0.583
I-SECRET [31]	26.89 $\pm$ 0.34	0.883 $\pm$ 0.021	0.658
MSQNet (Ours)	<b>29.82 <math>\pm</math> 0.47</b>	<b>0.908 <math>\pm</math> 0.019</b>	<b>0.706</b>

As shown in Table IV, we also analyze the computational complexity of MSQNet by comparing metrics such as parameters, floating-point operations (FLOPs), inference time, and GPU memory usage during the training stage. Although its computational complexity is not the most optimal compared to other methods, MSQNet achieves a good balance between efficiency and performance. For example, it has only 24.07M parameters and 236.39G FLOPs, demonstrating certain advantages over methods like PCE-Net, StillGAN, and LED. While its inference time (1.572 seconds) and GPU memory usage (16,031M) are slightly higher than those of comparable methods, the significant improvements in quality metrics offset these drawbacks. Moreover, the analysis of computational complexity provides valuable insights for future optimization, such as refining the model structure and reducing parameters to further enhance the overall efficiency of MSQNet.

TABLE IV  
THE COMPARISON OF DIFFERENT DEEP LEARNING-BASED METHODS IN TERMS OF COMPUTATIONAL PERFORMANCE METRICS.

Methods	Params (M)	FLOPs (G)	Inference time(s)	GPU memory usage (M)
SCR-Net [15]	89.29	137.29	1.462	2,906
PCE-Net [34]	26.65	343.45	0.103	9,554
Arc-Net [42]	54.42	72.81	0.084	4,730
cGAN [43]	54.41	72.61	0.605	2,774
RFormer [14]	21.11	183.25	0.228	18,193
StillGAN [16]	78.64	268.35	1.083	16,490
CutGAN [30]	11.38	256.53	0.081	8,948
LED [17]	113.68	996.37	244.424	10,850
CycleCBAM [29]	11.38	227.46	0.076	9,556
I-SECRET [31]	11.76	228.76	0.089	9,168
MSQNet (Ours)	24.07	236.39	1.572	16,031

Finally, to show the effectiveness of MAQNet from a subjective perspective, we invite two ophthalmologists to participate in a user study. Specifically, we randomly select 100 low-quality fundus images from the test set of the EyeQ dataset and enhance these images using MSQNet and other comparable methods. The experts re-evaluate the enhanced images based on two key factors: the mitigation of low-quality factors and the preservation of pathological features. As shown in

the table V, the number of enhanced images re-classified as high-quality by the ophthalmologists is significantly higher for MSQNet compared to other methods. This indicates that MSQNet excels not only in removing low-quality factors but also in preserving pathological features, thereby improving the diagnostic utility of medical images.

TABLE V  
THE QUALITY RE-CLASSIFICATION OF ENHANCED FUNDUS IMAGES IN THE USER STUDY.

Methods	Ophthalmologist 1		Ophthalmologists 2	
	low-quality	high-quality	low-quality	high-quality
Low-quality Image	88	12	90	10
LIME [39]	72	28	74	26
Fu et al. [40]	78	22	81	19
He et al. [41]	67	33	69	31
SCR-Net [15]	36	64	37	63
PCE-Net [34]	47	53	46	54
ArcNet [42]	51	49	53	47
cGAN [43]	39	61	35	65
RFormer [14]	42	58	38	62
StillGAN [16]	36	64	39	61
CutGAN [30]	45	55	43	57
LED [17]	40	60	39	61
CycleCBAM [29]	48	52	49	51
I-SECRET [31]	31	69	29	71
MSQNet (Ours)	27	73	25	75

#### D. Ablation Studies

In this section, we perform a series of ablation studies to investigate the effectiveness of the  $P^2Trans$  component and the VAB color space within MSQNet.

1) *Ablation Studies on  $P^2Trans$* : To analyze the effectiveness of  $P^2Trans$  in guiding multiscale feature interaction and aggregation, and to assess its ability to preserve lesions and anatomical structures, we compare MSQNet with its variant MSQNet *w/o  $P^2Trans$*  (MSQNet is trained independently without the  $P^2Trans$  component.).

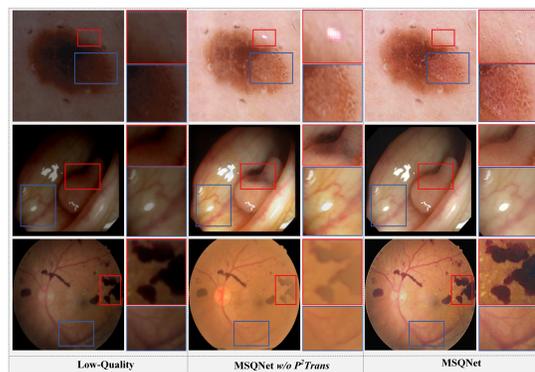


Fig. 6. Effect of  $P^2Trans$  on the enhancement results.

The low-quality medical images shown in Figure 6 exhibit uneven illumination, blur, and artifacts that obscure important anatomical and pathological details. MSQNet *w/o  $P^2Trans$*  generates relatively smooth enhanced images, which improves the overall visibility, but the images still lack sharpness in the boundaries and details. For example, the boundaries of polyps in the colonoscopy images and hemorrhages in the fundus images are degraded, and fine structures such as tiny vessels

in the fundus images are also not well-preserved. In addition, MSQNet *w/o*  $P^2Trans$  introduces undesired artifacts into the enhanced skin lesion image. In contrast, under the guidance of  $P^2Trans$ , the images enhanced by MSQNet exhibit clear details of lesions and anatomical structures across the fundus, colonoscopy, and skin lesion images. These comparison results suggest that  $P^2Trans$  can effectively guide multiscale feature interaction and aggregation, preserving more information about lesions and anatomical structures while avoiding the introduction of undesired artifacts during the quality enhancement process.

2) *Ablation Studies on Key Parameter for Feature Fusion Between RGB and VAB Color Spaces:*  $P^2Trans$  plays a critical role in the multi-scale global feature fusion between the RGB main branch and the VAB auxiliary branch. Therefore, we further conducted an in-depth investigation into the impact of the hyperparameter  $\lambda_3$  on the feature fusion between the RGB and VAB spaces.

TABLE VI  
THE EFFECT OF THE KEY PARAMETER  $\lambda_3$  FOR IMAGE QUALITY ENHANCEMENT ON RETINAL FUNDUS DATASET.

Method	Value of $\lambda_3$	PSNR $\uparrow$	SSIM $\uparrow$	IQAS $\uparrow$
MSQNet	0.3	25.42	0.867	0.629
MSQNet	0.5	26.75	0.873	0.647
MSQNet	0.7	31.26	0.942	0.706
MSQNet	1.0	30.47	0.937	0.698
MSQNet	1.3	29.46	0.924	0.684
MSQNet	1.5	28.21	0.916	0.672

As shown in the table VI,  $\lambda_3$  directly influences the balance between the contributions of the main branch and the auxiliary branch during the training process. When  $\lambda_3$  takes small values (e.g.,  $\lambda_3 = 0.3$  or  $0.5$ ), the influence of the auxiliary branch is limited, potentially leading to degraded pathological features or incomplete removal of low-quality factors. This weakens the guidance provided by the VAB color space for removing specific low-quality factors and preserving pathological details. Consequently, quality enhancement performance metrics such as PSNR, SSIM, and IQAS are relatively low. As  $\lambda_3$  increases (e.g.,  $\lambda_3 = 0.7$  and  $1.0$ ), the contribution of the auxiliary branch becomes more pronounced, enhancing feature fusion across different color spaces and providing stronger guidance to the main branch. This improvement is reflected in the enhancement of all performance metrics. Under this setting, the auxiliary branch effectively guides the main branch in removing specific low-quality artifacts while simultaneously preserving critical pathological details, resulting in optimal image quality. However, when  $\lambda_3$  further increases (e.g.,  $\lambda_3 = 1.3$  or  $1.5$ ), the performance exhibits a declining trend. This may result from an overemphasis on the auxiliary branch, which degraded the main branch's ability to capture the unique characteristics of RGB images (e.g., specific brightness and color information in the RGB space). Consequently, the overall feature fusion process becomes imbalanced, leading to reduced enhancement effects and slight degradation in performance metrics, such as PSNR and SSIM.

3) *Ablation Studies on Different Color Spaces:* The selection and construction of a color space are crucial factors for low-quality medical image enhancement. To investigate the effects of different color spaces on MSQNet, we experiment with various color spaces to observe their effects on MSQNet's enhancement performance, as reported in Table VII. These results reveal the following several interesting points:

TABLE VII  
EFFECTS OF THE DIFFERENT COLOR SPACES FOR IMAGE QUALITY ENHANCEMENT ON RETINAL FUNDUS DATASET.

Method	Color Space	PSNR $\uparrow$	SSIM $\uparrow$	IQAS $\uparrow$
MSQNet	only RGB	22.46	0.856	0.584
MSQNet	RGB+HSV	25.46	0.878	0.642
MSQNet	RGB+LAB	24.57	0.842	0.625
MSQNet	RGB+V	26.42	0.893	0.651
MSQNet	RGB+VAB	<b>30.47</b>	<b>0.937</b>	<b>0.698</b>

1) When considering the RGB+VAB color space, MSQNet achieves remarkable quality enhancement performance, with a PSNR of 30.47, an SSIM of 0.937, and an IQAS of 0.698. These results verify the effectiveness of incorporating the VAB color space for performance improvement over using only the RGB color space, which yields a PSNR of 22.46, an SSIM of 0.856, and an IQAS of 0.584.

2) Compared with using only the RGB color space, incorporating an additional HSV or LAB color space enables MSQNet to achieve varying degrees of performance improvement. With RGB+HSV, MSQNet achieves a PSNR of 25.46, an SSIM of 0.878, and an IQAS of 0.642, whereas RGB+LAB results in a PSNR of 24.57, an SSIM of 0.842, and an IQAS of 0.625. The HSV and LAB color spaces can guide low-quality image enhancement from different aspects. LAB can provide richer color information for the main branch, which has a better effect on the color balance of medical images. In contrast, HSV supplements richer structural information for the main branch, enabling MSQNet to effectively eliminate artifacts and blurring in low-quality images and preserve fine-grained details of lesions and anatomical structures.

3) RGB+V outperforms RGB+HSV by margins of 0.96 in PSNR, 0.015 in SSIM, and 0.009 in IQAS. This suggests that the V channel in the HSV color space contributes more to image quality enhancement than the H and S channels do. The V channel represents brightness information, which is crucial for improving the overall clarity and contrast of low-quality images, especially those with uneven illumination. In contrast, the H and S channels do not play a positive role in enhancing image quality.

4) MSQNet with RGB+V performs worse than MSQNet with RGB+VAB dose, achieving a PSNR of 26.42, an SSIM of 0.893, and an IQAS of 0.651 compared with the RGB+VAB results of 30.47, 0.937, and 0.698, respectively. The results further indicate that the A and B channels are critical for enhancing image quality, especially in color balancing for low-quality images, as the performance gap between RGB+V and RGB+VAB is still significant, with differences of 4.05 in PSNR, 0.044 in SSIM, and 0.047 in IQAS.

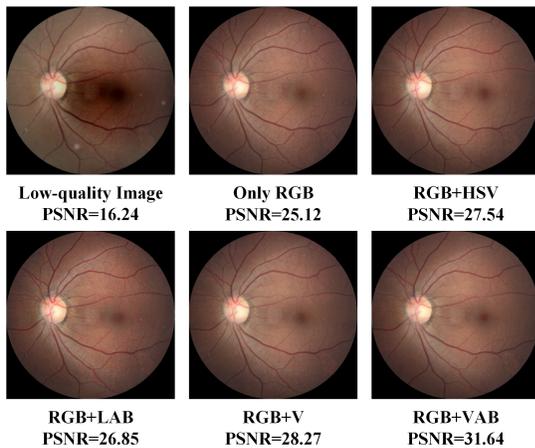


Fig. 7. Qualitative comparison of the enhancement results obtained by MSQNet under different color spaces.

As shown in Figure 7, we also conduct a qualitative comparison and analysis of the enhancement results obtained by the MSQNet framework under different color spaces. Figure 7 shows that, compared with a single color space, multiple color spaces can assist the MSQNet framework in achieving superior quality enhancement effects. Specifically, the combination of the RGB and HSV color spaces achieves the best visual results. The HSV color space can effectively assist MSQNet in enhancing the details of lesions and anatomical structures in low-quality fundus images. Furthermore, the combination of RGB and HSV can also effectively enhance and optimize the color information in low-quality images.

### E. Applications

The primary objectives of medical image quality enhancement methods are to facilitate downstream clinical tasks and improve the reliability of clinical diagnosis. To evaluate the effect of MSQNet, we employ it as a pre-processing step for various downstream image analysis tasks, such as vessel segmentation, skin lesion segmentation, and polyp segmentation. For vessel segmentation, we utilize CE-Net [44] as the baseline model, whereas UNet3+ [45] is used as the baseline for both the skin lesion and polyp segmentation tasks.

1) *Vessel Segmentation*: To validate the effects of the enhancement methods on the vessel segmentation task, we train vessel segmentation models using the same CE-Net architecture on both low-quality fundus images and enhanced fundus images obtained via different enhancement methods. The trained models are then evaluated on the testing set and corresponding enhanced images. As shown in Figure 8, the top row displays a low-quality fundus image and its corresponding quality enhancement results obtained via comparable methods and MSQNet. The middle and bottom rows illustrate the heatmaps and segmentation results of the vessel segmentation models, respectively. With the assistance of MSQNet, the vessel segmentation model is able to identify more complete and fine-grained vessel structures. Furthermore, the quantitative vessel segmentation results are reported in the first column of Table VIII. MSQNet achieves the highest DSC of

0.592 and IoU of 0.785, outperforming other enhancement methods and demonstrating its effectiveness in improving vessel segmentation performance.

2) *Skin Lesion Segmentation*: To evaluate the effects of the enhancement methods on the skin lesion segmentation task, we train skin lesion segmentation models using the same UNet3+ architecture on both low-quality skin lesion images and enhanced skin lesion images. The trained models are then evaluated on the original low-quality testing set and the enhanced testing sets. As shown in Figure 8 and the second column of Table VIII, the lesion segmentation model captures more specific lesion boundaries and achieves better segmentation results for the enhanced images obtained by MSQNet. Quantitatively, MSQNet yields the highest DSC of 0.597 and IoU of 0.712, surpassing other enhancement methods and showing its ability to enhance skin lesion images for improved segmentation performance.

TABLE VIII  
EFFECTS OF THE MSQNET FRAMEWORK ON MULTIPLE DOWNSTREAM MEDICAL IMAGE ANALYSIS TASKS.

Methods	<i>Vessel Seg.</i>		<i>Skin lesion Seg.</i>		<i>Polyp Seg.</i>	
	DSC $\uparrow$	IoU $\uparrow$	DSC $\uparrow$	IoU $\uparrow$	DSC $\uparrow$	IoU $\uparrow$
Low-quality image	0.483	0.678	0.462	0.573	0.634	0.712
SCR-Net [15]	0.564	0.760	0.576	0.692	0.723	0.796
PCE-Net [34]	0.556	0.748	0.558	0.662	0.703	0.774
ArcNet [42]	0.551	0.743	0.548	0.651	0.688	0.762
cGAN [43]	0.568	0.762	0.567	0.675	0.733	0.814
RFormer [14]	0.561	0.752	0.559	0.672	0.714	0.787
StillGAN [16]	0.562	0.758	0.563	0.677	0.718	0.792
CutGAN [30]	0.558	0.746	0.546	0.648	0.702	0.781
LED [17]	0.565	0.766	0.565	0.672	0.726	0.803
Cycle-CBAM [29]	0.553	0.747	0.552	0.653	0.694	0.769
I-SECRET [31]	0.573	0.764	0.587	0.702	0.728	0.805
MSQNet(Ours)	<b>0.592</b>	<b>0.785</b>	<b>0.597</b>	<b>0.712</b>	<b>0.749</b>	<b>0.837</b>

3) *Polyp Segmentation*: To demonstrate the benefit of the enhancement methods for the polyp segmentation task, we train polyp segmentation models using the same UNet3+ architecture on both low-quality endoscopy images and enhanced endoscopy images. The trained models are then evaluated on the low-quality endoscopy test images and the corresponding enhanced images. As shown in Figure 8 and the third column of Table VIII, MSQNet improves the quality of the endoscopic images by reducing blurring and uneven illumination. Consequently, the large number of false positives produced by the polyp segmentation model is alleviated, and the best performance for the polyp segmentation task is achieved. MSQNet (Ours) obtains the highest DSC of 0.749 and IoU of 0.837, outperforming other enhancement methods and validating its effectiveness in enhancing endoscopy images for polyp segmentation.

## VI. DISCUSSION AND LIMITATIONS

### A. Discussion

Although medical image quality enhancement methods have made remarkable progress in recent years, they do not fully exploit all available information due to the limited number of high-quality medical images and feature degradation. We

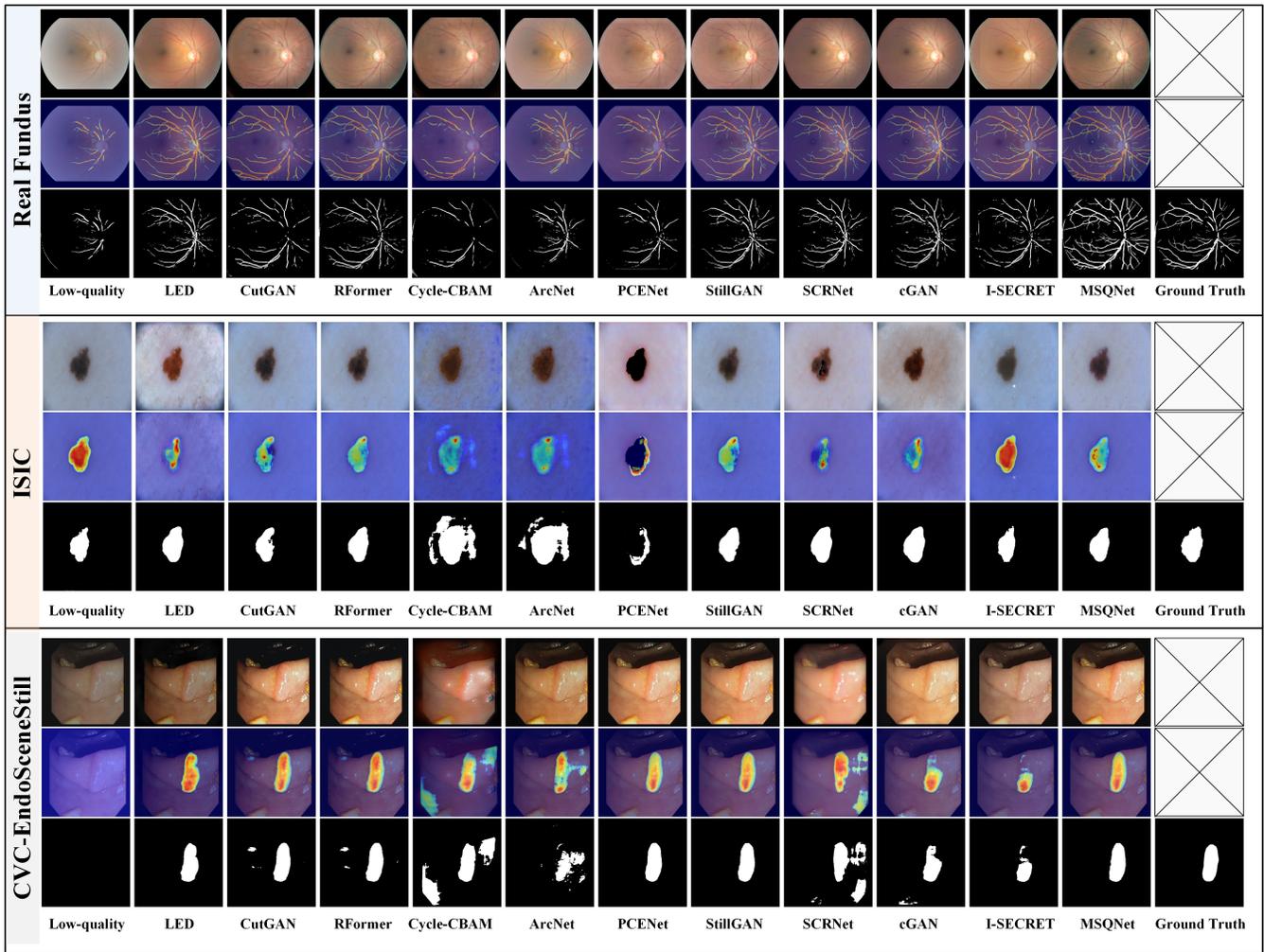


Fig. 8. Visual comparison of vessel, skin lesion, and polyp segmentation results on low-quality images and enhanced images obtained via different enhancement methods. The top row shows the low-quality images and their corresponding enhanced results. The middle and bottom rows display the heatmaps and segmentation results of the segmentation models, respectively.

highlight three critical questions that have been largely overlooked in the context of medical image quality enhancement, and we hope that this study inspires further research in the field of medical image enhancement.

1) How can we enhance the interpretability of the medical image quality enhancement model?

The advantage of MSQNet is that it simultaneously enhances the quality of medical images while avoiding the degradation of lesions and anatomical structures. Preserving fine-grained lesions and anatomical structures is highly importance for enhanced medical images because they provide intuitive illustrations for diagnosis by radiologists. As shown in Figure 9 (top row), despite the visual enhancement achieved by training MSQNet using only the RGB color space, a major limitation is that the network struggles to preserve and highlight fine-grained structural and lesion details in the deeper layers. As the network depth increases, the fine-grained lesion and structural details in the feature maps become progressively more degraded. This makes it challenging to identify the precise pathological markers needed for diagnosis, such as microaneurysms in fundus images and polyp borders in

endoscopic images. The previous quality enhancement models do not provide interpretability for the enhanced results.

In contrast, as illustrated in Figure 9 (bottom row), when MSQNet is trained jointly with both the RGB and VAB color spaces, it is able to much more effectively preserve and highlight anatomical structures and lesion features in the feature maps, even in the deepest layers. Guided by  $P^2Trans$ , MSQNet integrates relevant multi-scale features from the VAB color space. Consequently, discriminative features such as edges, textures, and color variations of key diagnostic markers are retained and highlighted.

2) What is the relationship between image quality and image analysis tasks?

To investigate the relationship between image quality and image analysis tasks, as well as the effects of various enhancement methods on downstream model training, we train segmentation models on enhanced images obtained via various enhancement methods. The experimental results indicate that higher image quality is beneficial for downstream tasks by removing artifacts, unbalanced illumination, and other interference while preserving lesion characteristics for low-

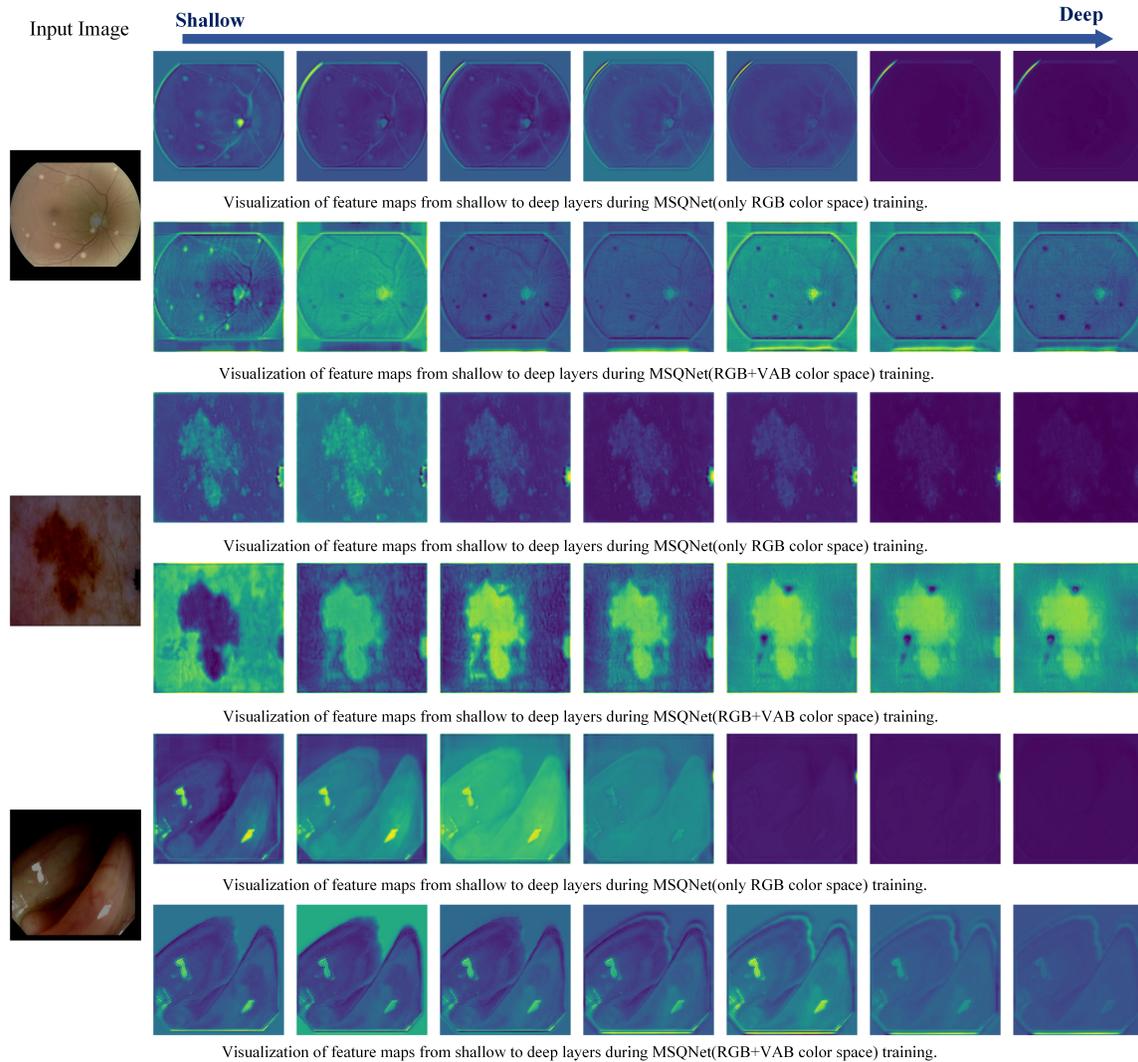


Fig. 9. Visualization of feature maps from shallow to deep layers during MSQNet training, comparing the use of only the RGB color space (top row) versus the RGB+VAB color spaces (bottom row). Training with RGB+VAB more effectively preserves and highlights fine-grained structural and lesion details across all layers, providing enhanced feature interpretability for improved medical image analysis and diagnosis.

quality images. As shown in Figure 10, taking the skin lesion segmentation task as an example, the enhanced image obtained by MSQNet enables the segmentation model to converge faster and achieve a lower loss during training. This result indicates that high-quality data essentially facilitate the downstream model’s training as well, such that the downstream model realizes better segmentation performance on unseen images.

3) How can information from unpaired images be fully leveraged?

The existing quality enhancement methods suffer from a major limitation: the insufficient availability of paired images. To address this issue, an increasing number of unsupervised quality enhancement methods leverage unpaired images to improve image quality. However, training enhancement methods that are based on unpaired images may introduce undesirable artifacts or distortions. In our study, we construct the VAB color space of low-quality images to leverage information fully from these unpaired images. As shown in Figure 11, compared with the feature map of the RGB color space, the feature map

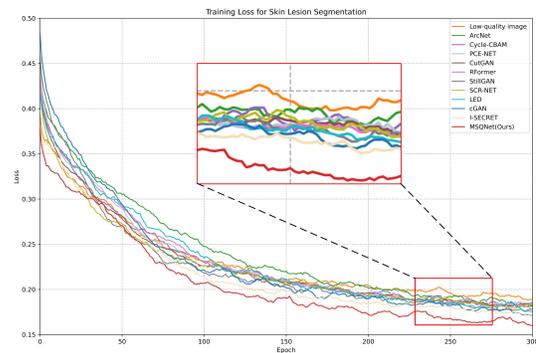


Fig. 10. Effects of various enhancement methods on the training of skin lesion segmentation model.

of the VAB color space highlights more fine-grained details and structures, such as the optic disc and blood vessels in fundus images, which are essential for accurate diagnosis and

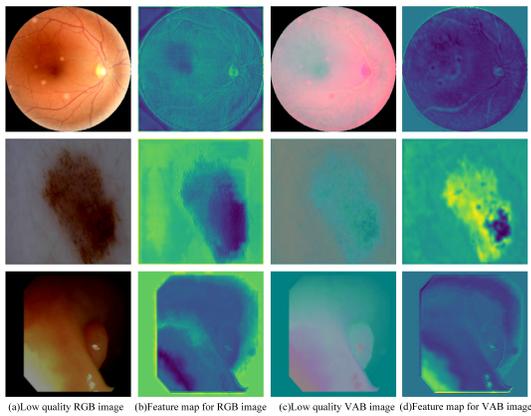


Fig. 11. Comparison of low-quality RGB and VAB images (a, c) and their corresponding feature maps (b, d). Compared with the RGB feature map (b), the feature map of the VAB color space (d) captures more fine-grained details, demonstrating the potential of the VAB color space to assist in enhancing RGB medical images for improved diagnosis and analysis.

analysis.

### B. Limitations

Although MSQNet demonstrates significant advantages in improving the usability and diagnostic performance of low-quality RGB medical images, it is still subject to the following two limitations in practical applications.

1) Adaptability to specific types of medical images: MSQNet demonstrates promising performance in enhancing low-quality RGB medical images by integrating multiple color spaces, such as RGB, LAB, and HSV. However, its applicability may be constrained by the inherent properties of color space representations. For other types of medical images, such as X-ray, CT, and MRI, the performance and generalization of MSQNet require further investigation due to their grayscale or modality-specific characteristics. For example, X-ray, CT, and MRI images exhibit notable differences in texture and contrast compared to RGB medical images. In contrast, relying solely on designed VAB color space and  $P^2Trans$  may not adequately capture the pathological features in these grayscale images.

2) MSQNet processes multiple color spaces in parallel, which increases both the number of parameters and the computational complexity compared to existing lightweight enhancement methods. In resource-constrained scenarios, such as portable medical devices or low-cost diagnostic equipment, the increased parameter size and computational complexity may hinder its practical applicability.

Considering that MSQNet still relies on the supervised guidance of paired medical images and the aforementioned limitations, future research will focus on the following directions:

1) Exploring unsupervised frameworks for medical image quality enhancement: Given that MSQNet relies on paired high- and low-quality training images, future research could focus on developing unsupervised enhancement frameworks to mitigate dependence on paired datasets. Techniques such as

adversarial learning (e.g., GANs, diffusion models) and self-supervised learning can be explored, leveraging the collaborative optimization process between generators and discriminators to enhance medical image quality.

2) Extending quality enhancement to multi-modal medical images: MSQNet is primarily designed for optimizing the quality of RGB medical images. However, other imaging modalities (e.g., X-ray, CT, MRI, OCT) are equally critical for clinical diagnosis. Future research should focus on developing a more universal enhancement framework that is applicable across various imaging modalities, thereby improving adaptability and effectiveness in diverse clinical scenarios.

3) Development of lightweight enhancement frameworks: To reduce the model parameters of MSQNet, future research could focus on developing lightweight variants specifically tailored for resource-constrained scenarios. By exploring techniques such as network pruning and knowledge distillation, the MSQNet could be further optimized to achieve a trade-off between performance and computational efficiency.

## VII. CONCLUSION

In this study, we propose MSQNet, a novel framework designed for enhancing low-quality medical images by leveraging multi-color spaces. By combining the advantages of the HSV and LAB color spaces, we introduce the VAB color space, which broadens the scope of traditional image enhancement techniques. Considering the potential risk of degrading pathological features during the quality enhancement process, we propose the pathology-preserving transformer ( $P^2Trans$ ). This component extracts and integrates multi-level global information from the VAB color space, effectively guiding the enhancement of RGB medical images while preserving essential pathological features.  $P^2Trans$  ensures that image quality enhancement aligns with the clinical diagnostic value of the medical images. The experimental results demonstrate the robustness of our method, which produces high-quality enhancement results while preserving lesions and structural integrity at a state-of-the-art level. Moreover, our method positively affects various downstream tasks. With respect to performance improvements in these tasks, our method exhibits strong interpretability, enabling us to explain how quality enhancement contributes to task performance.

## ACKNOWLEDGMENTS

This research was supported by the National Natural Science Foundation of China (No.62076059), the Science and Technology Joint Project of Liaoning province (2023JH2/101700367), the Fundamental Research Funds for the Central Universities (No. N2424010-7) and China Scholarship Council(202306080125). The authors declare no known competing financial interests or personal relationships that could have potentially influenced the work presented in this paper.

## REFERENCES

- [1] Y. Yang, J. Yu, J. Zhang, W. Han, H. Jiang, and Q. Huang, "Joint embedding of deep visual and semantic features for medical image report generation," *IEEE Transactions on Multimedia*, vol. 25, pp. 167–178, 2023.

- [2] J. Lou, H. Lin, P. Young, R. White, Z. Yang, S. Shelmerdine, D. Marshall, E. Spezi, M. Palombo, and H. Liu, "Predicting radiologists' gaze with computational saliency models in mammogram reading," *IEEE Transactions on Multimedia*, vol. 26, pp. 256–269, 2024.
- [3] X. Liu, L. Faes, A. U. Kale, S. K. Wagner, D. J. Fu, A. Bruynseels, T. Mahendiran, G. Moraes, M. Shamdass, C. Kern, *et al.*, "A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis," *The lancet digital health*, vol. 1, no. 6, pp. e271–e297, 2019.
- [4] Q. Hou, Y. Wang, P. Cao, S. Cheng, L. Lan, J. Yang, X. Liu, and O. R. Zaiane, "A collaborative self-supervised domain adaptation for low-quality medical image enhancement," *IEEE Transactions on Medical Imaging*, 2024.
- [5] I. C. Holmen, S. M. Konda, J. W. Pak, K. W. McDaniel, B. Blodi, K. E. Stepien, and A. Domalpally, "Prevalence and severity of artifacts in optical coherence tomographic angiograms," *JAMA ophthalmology*, vol. 138, no. 2, pp. 119–126, 2020.
- [6] H. Liu, H. Li, X. Wang, H. Li, M. Ou, L. Hao, Y. Hu, and J. Liu, "Understanding how fundus image quality degradation affects cnn-based diagnosis," in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 438–442, IEEE, 2022.
- [7] J. Ee, E. Baylor, F. Hersch, A. Iurchenko, L. Wilcox, P. Ruamviboonsuk, and L. M. Vardoulakis, "A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy," in *Proceedings of the 2020 CHI conference on human factors in computing systems*, pp. 1–12, 2020.
- [8] Q. Hou, P. Cao, L. Jia, L. Chen, J. Yang, and O. R. Zaiane, "Image quality assessment guided collaborative learning of image enhancement and classification for diabetic retinopathy grading," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 3, pp. 1455–1466, 2022.
- [9] Z. Pan, F. Yuan, J. Lei, W. Li, N. Ling, and S. Kwong, "Miegan: Mobile image enhancement via a multi-module cascade neural network," *IEEE Transactions on Multimedia*, vol. 24, pp. 519–533, 2022.
- [10] H. Li, Z. Lin, Z. Qiu, Z. Li, K. Niu, N. Guo, H. Fu, Y. Hu, and J. Liu, "Enhancing and adapting in the clinic: Source-free unsupervised domain adaptation for medical image enhancement," *IEEE Transactions on Medical Imaging*, 2023.
- [11] B. Ding, R. Zhang, L. Xu, G. Liu, S. Yang, Y. Liu, and Q. Zhang, "U2d2net: Unsupervised unified image dehazing and denoising network for single hazy image enhancement," *IEEE Transactions on Multimedia*, vol. 26, pp. 202–217, 2024.
- [12] F. Shamshad, S. Khan, S. W. Zamir, M. H. Khan, M. Hayat, F. S. Khan, and H. Fu, "Transformers in medical imaging: A survey," *Medical Image Analysis*, p. 102802, 2023.
- [13] Q. Ma, Y. Wang, and T. Zeng, "Retinex-based variational framework for low-light image enhancement and denoising," *IEEE Transactions on Multimedia*, vol. 25, pp. 5580–5588, 2023.
- [14] Z. Deng, Y. Cai, L. Chen, Z. Gong, Q. Bao, X. Yao, D. Fang, W. Yang, S. Zhang, and L. Ma, "Rformer: Transformer-based generative adversarial network for real fundus image restoration on a new clinical benchmark," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 9, pp. 4645–4655, 2022.
- [15] H. Li, H. Liu, H. Fu, H. Shu, Y. Zhao, X. Luo, Y. Hu, and J. Liu, "Structure-consistent restoration network for cataract fundus image enhancement," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 487–496, Springer, 2022.
- [16] Y. Ma, J. Liu, Y. Liu, H. Fu, Y. Hu, J. Cheng, H. Qi, Y. Wu, J. Zhang, and Y. Zhao, "Structure and illumination constrained gan for medical image enhancement," *IEEE Transactions on Medical Imaging*, vol. 40, no. 12, pp. 3955–3967, 2021.
- [17] P. Cheng, L. Lin, Y. Huang, H. He, W. Luo, and X. Tang, "Learning enhancement from degradation: A diffusion model for fundus image enhancement," *arXiv preprint arXiv:2303.04603*, 2023.
- [18] W. Tang, F. He, Y. Liu, and Y. Duan, "Matr: Multimodal medical image fusion via multiscale adaptive transformer," *IEEE Transactions on Image Processing*, vol. 31, pp. 5134–5149, 2022.
- [19] W. Tang, F. He, and Y. Liu, "Ydtr: Infrared and visible image fusion via y-shape dynamic transformer," *IEEE Transactions on Multimedia*, vol. 25, pp. 5413–5428, 2022.
- [20] H. Yang, G. Nan, M. Lin, F. Chao, Y. Shen, K. Li, and R. Ji, "Lab-net: Lab color-space oriented lightweight network for shadow removal," *arXiv preprint arXiv:2208.13039*, 2022.
- [21] Y. Wang, J. Guo, H. Gao, and H. Yue, "Uiecc<sup>2</sup>-net: Cnn-based underwater image enhancement using two color space," *Signal Processing: Image Communication*, vol. 96, p. 116250, 2021.
- [22] K. Zuiderveld, "Contrast limited adaptive histogram equalization," *Graphics gems*, pp. 474–485, 1994.
- [23] A. Mitra, S. Roy, S. Roy, and S. K. Setua, "Enhancement and restoration of non-uniform illuminated fundus image of retina obtained through thin layer of cataract," *Computer methods and programs in biomedicine*, vol. 156, pp. 169–178, 2018.
- [24] L. Cao, H. Li, and Y. Zhang, "Retinal image enhancement using low-pass filtering and  $\alpha$ -rooting," *Signal Processing*, vol. 170, p. 107445, 2020.
- [25] J. Cheng, Z. Li, Z. Gu, H. Fu, D. W. K. Wong, and J. Liu, "Structure-preserving guided retinal image filtering and its application for optic disk analysis," *IEEE Transactions on Medical Imaging*, vol. 37, no. 11, pp. 2536–2546, 2018.
- [26] W. Wang, D. Yan, X. Wu, W. He, Z. Chen, X. Yuan, and L. Li, "Low-light image enhancement based on virtual exposure," *Signal Processing: Image Communication*, vol. 118, p. 117016, 2023.
- [27] Z. Fu, Y. Yang, X. Tu, Y. Huang, X. Ding, and K.-K. Ma, "Learning a simple low-light image enhancer from paired low-light instances," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 22252–22261, 2023.
- [28] M. Wang, J. Li, and C. Zhang, "Low-light image enhancement by deep learning network for improved illumination map," *Computer Vision and Image Understanding*, vol. 232, p. 103681, 2023.
- [29] Q. You, C. Wan, J. Sun, J. Shen, H. Ye, and Q. Yu, "Fundus image enhancement method based on cyclegan," in *2019 41st annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, pp. 4500–4503, IEEE, 2019.
- [30] T. Park, A. A. Efros, R. Zhang, and J.-Y. Zhu, "Contrastive learning for unpaired image-to-image translation," in *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX 16*, pp. 319–345, Springer, 2020.
- [31] P. Cheng, L. Lin, Y. Huang, J. Lyu, and X. Tang, "I-secret: Importance-guided fundus image enhancement via semi-supervised contrastive constraining," in *Medical Image Computing and Computer Assisted Intervention—MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VIII 24*, pp. 87–96, Springer, 2021.
- [32] A. Raj, N. A. Shah, and A. K. Tiwari, "A novel approach for fundus image enhancement," *Biomedical signal processing and control*, vol. 71, p. 103208, 2022.
- [33] Z. Shen, H. Fu, J. Shen, and L. Shao, "Modeling and enhancing low-quality retinal fundus images," *IEEE transactions on medical imaging*, vol. 40, no. 3, pp. 996–1006, 2020.
- [34] H. Liu, H. Li, H. Fu, R. Xiao, Y. Gao, Y. Hu, and J. Liu, "Degradation-invariant enhancement of fundus images via pyramid constraint network," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 507–516, Springer, 2022.
- [35] E. Decencière, X. Zhang, G. Cazuguel, B. Lay, B. Cochener, C. Trone, P. Gain, R. Ordonez, P. Massin, A. Erginay, B. Charton, and J.-C. Klein, "Feedback on a publicly distributed image database: the messidor database," *Image Analysis & Stereology*, vol. 33, no. 3, pp. 231–234, 2014.
- [36] N. C. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler, *et al.*, "Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic)," in *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*, pp. 168–172, IEEE, 2018.
- [37] J. Bernal, J. Sánchez, and F. Vilarino, "Impact of image preprocessing methods on polyp localization in colonoscopy frames," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 7350–7354, IEEE, 2013.
- [38] D. Vázquez, J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, A. M. López, A. Romero, M. Drozdal, A. Courville, *et al.*, "A benchmark for endoluminal scene segmentation of colonoscopy images," *Journal of healthcare engineering*, vol. 2017, 2017.
- [39] X. Guo, Y. Li, and H. Ling, "Lime: Low-light image enhancement via illumination map estimation," *IEEE Transactions on image processing*, vol. 26, no. 2, pp. 982–993, 2016.
- [40] X. Fu, P. Zhuang, Y. Huang, Y. Liao, X.-P. Zhang, and X. Ding, "A retinex-based enhancing approach for single underwater image," in *2014 IEEE international conference on image processing (ICIP)*, pp. 4572–4576, IEEE, 2014.
- [41] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 12, pp. 2341–2353, 2010.

- [42] H. Li, H. Liu, Y. Hu, H. Fu, Y. Zhao, H. Miao, and J. Liu, "An annotation-free restoration network for cataractous fundus images," *IEEE Transactions on Medical Imaging*, vol. 41, no. 7, pp. 1699–1710, 2022.
- [43] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1125–1134, 2017.
- [44] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, "Ce-net: Context encoder network for 2d medical image segmentation," *IEEE transactions on medical imaging*, vol. 38, no. 10, pp. 2281–2292, 2019.
- [45] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y.-W. Chen, and J. Wu, "Unet 3+: A full-scale connected unet for medical image segmentation," in *ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pp. 1055–1059, IEEE, 2020.

VIII. BIOGRAPHY SECTION



Assisted Intervention (MICCAI), and ACM Multimedia (ACM MM).

**Qingshan Hou** is currently a Ph.D. candidate in the Key Laboratory of Medical Image Computing of Ministry of Education, College of Computer Science and Engineering, Northeastern University, China. His research interests lie in the fields of computer vision, image quality enhancement, and medical image analysis. He has published his research findings in prestigious journals and conferences, including IEEE Journal of Biomedical and Health Informatics (JBHI), IEEE Transactions on Medical Imaging (TMI), Medical Image Computing and Computer Assisted Intervention (MICCAI), and ACM Multimedia (ACM MM).



**Yaqi Wang** is currently pursuing the master's degree in Key Laboratory of Medical Image Computing of Ministry of Education, College of Computer Science and Engineering, Northeastern University, China. The objectives of his research are knowledge distillation, medical image segmentation, and medical image enhancement. He has published her research findings in prestigious journals, such as IEEE Transactions on Medical Imaging (TMI) and Applied Intelligence.

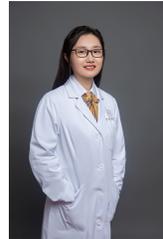


Methods and Programs in Biomedicine, Medical Image Computing and Computer Assisted Intervention (MICCAI), and AAAI Conference on Artificial Intelligence (AAAI).

**Peng Cao** is an associate professor of Key Laboratory of Intelligent Computing in Medical Image, Ministry of Education. He earned his Ph.D. degree in computer application in 2014 at Northeastern University, China. His research interests include image processing, brain network analysis and machine learning for medical data. His research papers have been published in or accepted by journals or conference including Pattern Recognition, IEEE Journal of Biomedical and Health Informatics (JBHI), IEEE Transactions on Medical Imaging (TMI), Computer Assisted Intervention (MICCAI), and AAAI Conference on Artificial Intelligence (AAAI).



**Jianguo Ju** is currently pursuing a doctorate in the School of Information Science and Technology at Northwest University. His research interests include computer vision, medical image process, and weakly supervised learning. He has published his research findings in prestigious journals including IEEE Journal of Biomedical and Health Informatics (JBHI), Pattern Recognition Letters (PRL), BMC Oral Health, Alexandria Engineering Journal (AEJ).



**Huijuan Tu** is an associate chief radiologist of Kunshan Hospital of Chinese Medicine. She is currently pursuing a doctorate at Soochow University. She received her bachelor's degree from Southeast University in 2009 and master's degree from Nanjing Medical University in 2012 respectively. She also was an observer of University Hospital Coventry & Warwickshire and Moorfields Eye Hospital. She specializes in digestive system imaging, especially small intestine and pancreas imaging.



Computing and Computer Assisted Intervention (MICCAI).

**Xiaoli Liu** received her Ph.D. in computer application from Northeastern University, China, in 2018. She is an algorithm expert for an internet company and a former postdoctoral research fellow at the National University of Singapore. Her research interest focuses on computer graphics, 3D reconstruction, medical imaging, bioinformatics and text analysis. Her research papers have been published in or accepted by journals or conferences including Nature Machine Intelligence, ACM TKDD, Advanced Science, Pattern Recognition and Medical Image Computing and Computer Assisted Intervention (MICCAI).



of Sciences of the United States of America(PNAS).

**Jinzhu Yang** received the Ph.D. degree in pattern recognition and intelligent system from the Northeastern University, Shenyang, China, in 2007. He is a professor and director of the Key Laboratory of Intelligent Computing in Medical Image, Ministry of Education. His research interest focuses on image processing and analysis, artificial intelligence pattern recognition, and data analysis. His research papers have been published in or accepted by journals including Magnetic Resonance Imaging, BMC Bioinformatics, and Proceedings of the National Academy of Sciences of the United States of America(PNAS).



vision, AI in healthcare, and trustworthy AI.

**Huazhu Fu (Senior Member, IEEE)** received the Ph.D. degree from Tianjin University, Tianjin, China, in 2013. He is currently a senior Scientist with IHPC, A\*STAR, Singapore. He was a Research Fellow with Nanyang Technological University (NTU), Singapore, during 2013–2015, a Research Scientist with the Institute for Infocomm Research (I2R), A\*STAR, Singapore, during 2015–2018, and a Senior Scientist with Inception Institute of Artificial Intelligence (IIAI), UAE, during 2018–2021. His research interests include computer vision, AI in healthcare, and trustworthy AI.



ture Aging, Ophthalmology, JAMA Ophthalmology, British Journal of Ophthalmology etc.

**Yih Chung Tham** is a clinically trained-optometrist with PhD training in ocular epidemiology from the National University of Singapore (NUS). He is currently an Assistant Professor (tenure-track) at the Department of Ophthalmology, Yong Loo Lin School of Medicine, NUS. His research scope focuses on visual impairment, age-related eye diseases, ocular imaging and deep learning. To date, Dr Tham has published over 200 peer-reviewed scientific papers in notable journals such as Lancet Digital Health, Nature Biomedical Engineering, Nature Aging, Ophthalmology, JAMA Ophthalmology, British Journal of Ophthalmology etc.



**Osmar R. Zaiane(Senior Member, IEEE)** is a professor in Computing Science at the University of Alberta, Canada, and Scientific Director of the Alberta Innovates Centre for Machine Learning. He is an associate Editor of many International Journals on data mining and data analytics and served as program chair and general chair for scores of international conferences in the field of knowledge discovery and data mining. His current research interests include deep learning, data mining and healthcare informatics.