



Spectral reconstruction of fundus images using retinex-based semantic spectral separation transformer, applied for retinal oximetry[☆]

Jicheng Liu^{a,b}, Wenteng Gao^c, Dehan Zhao^c, Lei Yang^c, Peng Liu^{a,b}, Ronald X. Xu^{a,b,**}, Mingzhai Sun^{a,b,*}

^a School of Biomedical Engineering, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, Anhui, 230026, PR China

^b Suzhou Institute for Advanced Research, University of Science and Technology of China, Suzhou, Jiangsu, 215123, PR China

^c Department of Precision Machinery and Precision Instruments, University of Science and Technology of China, Hefei 230026, China

ARTICLE INFO

Dataset link: <https://osf.io/8dcuw/>

Keywords:

Spectral reconstruction
Retinex theory
Retinal oxygen saturation
Transformer

ABSTRACT

Multispectral imaging, a non-invasive technique to measure retinal oxygen saturation levels, is limited due to its time-consuming and expensive nature. In this paper, we present R3ST (Retinex-based Semantic Spectral Separation Transformer), an innovative end-to-end method for RGB fundus image spectral reconstruction. Our proposed model leverages RGB images to reconstruct multispectral images, making retinal oximetry more accessible for diagnostics.

Existing spectral reconstruction methods face challenges in achieving high accuracy and generalizing across camera sources, affecting diagnostic results. R3ST addresses these issues through an unsupervised semantic spectral separation module that strives to encourage the separate reconstruction of pixels with different semantics as much as possible. Additionally, we introduce an oxygen saturation loss to improve reconstruction accuracy within the relevant regions. To overcome the limitations of generalization properties, our model incorporates a self-supervised Retinex reflectance extraction module. This simulates various imaging systems and extracts reflectance information from fundus images for reconstruction.

Extensive experiments demonstrate R3ST's superior performance in spectral reconstruction accuracy, making retinal oximetry a more viable, cost-effective diagnostic option that provides essential eye health and systemic health information. The code of R3ST is available at <https://github.com/telescOpe/R3ST>.

1. Introduction

Multispectral imaging is a powerful spectroscopy tool that captures the spectrum of each pixel in an image, enabling a detailed and precise analysis of objects' chemical and physical properties [1–5]. This technique has various potential applications, including bioscience [1], food science [2], remote sensing [3], biometrics [4], and medical imaging [5]. In the medical field, multispectral imaging facilitates the calculation of retinal oxygen saturation levels by measuring differential light absorption by oxyhemoglobin and deoxyhemoglobin in retinal blood vessels [6]. By illuminating the eye with different wavelengths of light and analyzing the reflected light, the amount and oxygenation state of hemoglobin can be determined, allowing for the calculation of blood vessel oxygen saturation levels [6]. This non-invasive method offers real-time measurements and serves as a valuable tool for assessing

retinal health and detecting issues such as retinal ischemia, as shown in Fig. 1(a).

Retinal oxygen saturation is a crucial parameter for evaluating eye health and function, as the retina is a highly active tissue requiring constant oxygen supply. Changes in retinal oxygen saturation levels may indicate eye diseases like diabetic retinopathy, retinal vein occlusion, and glaucoma, making accurate measurements essential for early detection, diagnosis, and management. Moreover, since the retina is an extension of the central nervous system and shares blood supply similarities with the brain, changes in retinal oxygen saturation levels can reflect cortical oxygen saturation level changes [7–11]. Accurate retinal oxygen saturation measurement can potentially detect changes in cortical oxygen saturation levels, aiding in the diagnosis and management of brain diseases like Alzheimer's and multiple sclerosis [7–11].

[☆] This work was funded by the National Key R&D Program of China (Grant No. 2021YFC2401402 and 2022YFA1104800).

* Corresponding author.

** Corresponding author at: School of Biomedical Engineering, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, Anhui, 230026, PR China.

E-mail addresses: xux@ustc.edu.cn (R.X. Xu), mingzhai@ustc.edu.cn (M. Sun).

<https://doi.org/10.1016/j.bspc.2024.106301>

Received 26 December 2023; Received in revised form 14 March 2024; Accepted 8 April 2024

Available online 12 April 2024

1746-8094/© 2024 Elsevier Ltd. All rights reserved.

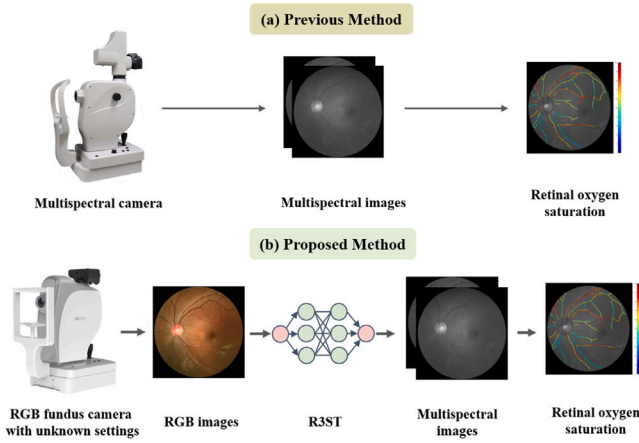


Fig. 1. Comparison of retinal oxygen saturation calculation methods. (a) Conventional method using a multi-spectral camera, which is costly. (b) Proposed method utilizing the Retinex-based Semantic Spectral Separation Transformer (R3ST) model for spectral reconstruction of retinal images, enabling oxygen saturation calculation based on RGB retinal images.

While multispectral images provide more accurate spectral information than RGB images, they require specialized hardware, which is not always feasible. We proposed a method for calculating retinal oxygen saturation using a cost-effective conventional fundus camera, as shown in Fig. 1(b). Spectral reconstruction can restore approximate multispectral images from a single RGB image, offering a convenient and affordable approach to obtain detailed spectral information [12–14].

Various spectral reconstruction methods based on RGB images exist, with accuracy depending on training data quality and model complexity [12–14]. This study presented a new model for multispectral reconstruction of RGB fundus images, addressing two key issues: insufficient accuracy in existing methods and poor generalization capabilities. To tackle the first issue, we designed a simple and effective unsupervised semantic spectral separation module that separates and recombines regions with different semantics in an unsupervised manner, encouraging reconstructed multispectral images to have distinct pixels for areas with varying semantics. To address the second issue, we designed a self-supervised Retinex reflectance extraction module that simulates different imaging systems and extracts the reflectance of fundus images for reconstruction.

In summary, we have made the following contributions:

- Proposing a self-supervised Retinex reflectance extraction module to enhance the generalizability of the spectral reconstruction model and enable better performance on cross-camera RGB images.
- Introducing a simple and effective unsupervised semantic spectral separation module that encourages reconstructed multispectral images to have distinct pixels for areas with different semantics in an unsupervised manner.
- Proposing a technique for calculating blood oxygen saturation using retinal RGB images, leveraging a spectral reconstruction network specifically designed for fundus images.
- Demonstrating through numerous experiments that the proposed spectral reconstruction network achieves state-of-the-art performance.

2. Related work

2.1. Spectral imaging techniques and retinal oximetry

Spectral imaging techniques involve capturing and analyzing images at various wavelengths [15], which enables the visualization of

structures and substances that are not visible with conventional imaging methods. Spectral imaging has revolutionized medical imaging and has found applications in diverse fields such as ophthalmology [16], dermatology [17], and gastroenterology [18]. In ophthalmology, spectral imaging has been used for retinal oximetry, a non-invasive method for measuring the oxygen saturation of retinal vessels. By analyzing the absorption of light by hemoglobin in the retinal vessels at different wavelengths, spectral imaging can determine the oxygen saturation levels of the vessels [19], providing valuable information on retinal blood flow and metabolism. Retinal oximetry has potential applications in the diagnosis and management of various ocular diseases, including glaucoma [20], diabetic retinopathy [21], and age-related macular degeneration [22]. The development of novel spectral imaging techniques holds great promise for the future of retinal oximetry and the diagnosis and treatment of retinal diseases. However, to the best of our knowledge, there is currently no work that utilizes spectral reconstruction to compute retinal oxygen saturation.

2.2. Spectral reconstruction from RGB images

Spectral reconstruction from RGB images is a challenging inverse problem. Spectral reconstruction methods can be divided into traditional methods and deep learning methods. Traditional methods typically utilize prior information, such as sparsity and spectral correlations, to improve the reconstruction performance. Arad and Ben-Shahar [13] used sparse coding to find a complete basis spectrum based on sparsity. Aeschbacher et al. [23] improved the reconstruction performance by introducing a shallow A+ method to establish an RGB-to-hyperspectral mapping in a local dictionary, rather than building a global mapping like sparse coding. Akhtar and Mian [24] modeled the basis spectra of different materials using Gaussian processes. However, these traditional methods have limitations in representing capacities and exhibit limited generalization ability, resulting in lower reconstruction accuracy.

Deep learning methods utilize paired data to find a data-driven mapping from RGB images to multispectral images. Xiong et al. [25] presented a unified deep learning framework to recover hyperspectral images from spectrally undersampled projections. Can and Timofte [26] designed a residual structure spectrum reconstruction network to efficiently complete the reconstruction task while reducing model complexity. Lore et al. [27] combined generative adversarial networks to perform both spectrum reconstruction and super-resolution tasks. Rangnekar et al. [28] used a conditional generative adversarial network for hyperspectral image reconstruction in remote sensing. Zhang et al. [29] proposed a pixel-perception hybrid network for spectrum reconstruction. Zhu et al. [30] fused semantic information into the reconstruction network to improve the reconstruction performance. Cai et al. [31] proposed the first Transformer-based framework, which is designed based on the sparsity and spectral self-similarity of spectral images and has low memory consumption. The performance of the spectral reconstruction model noticeably decreases when applied to RGB images captured by different cameras due to varying camera spectral sensitivities and post-processing techniques. To address this issue, several studies including [30,32,33] estimated the spectral sensitivity of the camera first and then reconstruct the spectral images accordingly. However, these methods achieve impressive results but exhibit limitations in capturing information unrelated to the imaging system and utilizing semantic information in an unsupervised manner.

2.3. Retinal reflectance and retinex theory

To potentially decrease errors in cross-camera spectral reconstruction, one can leverage properties of the image that are not influenced by the imaging system. The reflectance of an object is an inherent characteristic that remains unchanged throughout the imaging process. A viable approach for extracting reflectance information is based on

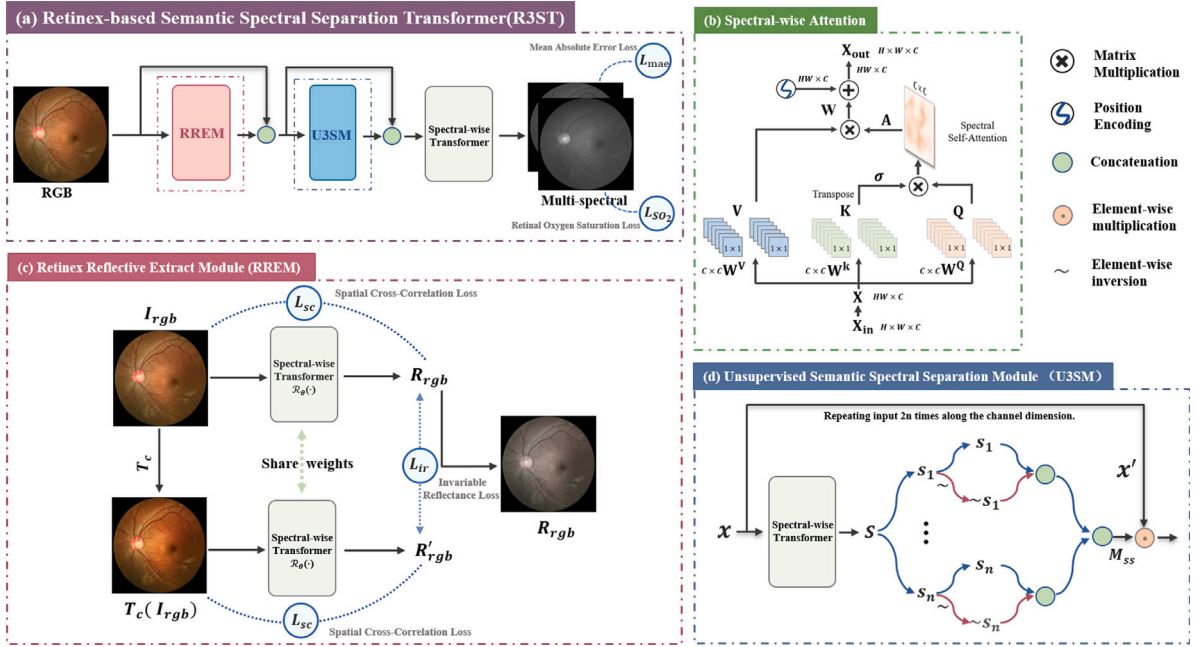


Fig. 2. Architecture of the proposed Retinex-based Semantic Spectral Separation Transformer (R3ST). (a) Overall architecture of R3ST, consisting of RREM, U3SM, and spectral-wise Transformer. (b) Spectral-wise Attention, a critical component in the Spectral-wise Transformer, is used to capture long-range dependencies and inter-spectra self-similarity. (c) Retinex reflective extract module, a self-supervised learning method for estimating reflectance in fundus images. (d) Structure of U3SM, used for input separation and subsequent spectral reconstruction. The network is constrained by L_{mae} , L_{so_2} , and L_{re} . The L_{re} is used to constrain RREM.

Retinex theory [34], which suggests that a color image can be decomposed into reflectance and illumination components. This theory has been widely used for various visual tasks. Several approaches have been proposed for visual enhancement and image processing. For example, Wang et al. [35] proposed an algorithm that preserves naturalness for enhancing non-uniform illumination images, while [36] presented a fusion-based method for improving weakly illuminated images. Furthermore, Ren et al. [37] proposed a Retinex-based technique for noise removal and low-light enhancement, and Wei et al. [38] introduced Retinex-Net, a method for estimating reflectance and illumination to enhance low-light images. Additionally, Shu et al. [39] extracted reflectance for neural face editing, while [40] proposed a weighted variational model for simultaneous estimation of reflectance and illumination. Moreover, Ma et al. [41] proposed a new method for Retinex theory that is effective in recovering reflectance from natural images and improving MRI data visual quality. Finally, Zhuang et al. [42] introduced a Bayesian retinex method for enhancing single underwater images. Based on the aforementioned study, the extraction of reflectance using Retinex theory has the potential to decrease errors in cross-camera spectral reconstruction. However, to the best of our knowledge, there has been no related work employing Retinex Theory to enhance the accuracy of reconstruction.

3. Method

3.1. Overall framework

We proposed the end-to-end Retinex-based Semantic Spectral Separation Transformer (R3ST) for multispectral reconstruction of fundus images. The overall framework of R3ST is shown in Fig. 2(a). The network includes three new components: the spectral-wise Transformer, the self-supervised Retinex Reflectance Extraction Module (RREM) and the Unsupervised Semantic Spectral Separation Module (U3SM). The RREM, based on the Retinex theory, extracts the reflectance map from the RGB fundus image, while the U3SM learns a semantic map in an unsupervised manner and uses it to guide the reconstruction process, enhancing the spectral reconstruction effect. The network is

constrained by proposed reflectance extraction loss L_{re} , the retinal oxygen saturation loss L_{so_2} , and the Mean Absolute Error (MAE) loss L_{mae} . The reflectance extraction loss comprises of two components: the invariant reflectivity term and the spatial correlation term. The retinal oxygen saturation loss is designed based on the calculation of the oxygen saturation level of the blood vessel and is used to improve the fidelity of the reconstruction.

3.2. Spectral-wise transformer

The Spectral-wise Transformer [31], a crucial component of R3ST, incorporates spectral-wise attention, convolution layers, and activation layers. The structure of spectral-wise attention is shown in Fig. 2(b). Given the input $X_{in} \in \mathbb{R}^{H \times W \times C}$ with height H , width W , and channels C , it is reshaped into C tokens, i.e., $X \in \mathbb{R}^{H \times W \times C}$:

$$Q = XW^Q, K = XW^K, V = XW^V \quad (1)$$

where W^Q , W^K , and $W^V \in \mathbb{R}^{C \times C}$ are learnable parameters, Q, K, V are equivalent to the results obtained by convolution with 1×1 convolutional kernels.

The spectral-wise attention splits Q, K , and V into N heads along the spectral channel dimension. The dimension of each head is $\frac{C}{N}$, Fig. 2(b) represents the case of $N = 1$. For each $head_j$ with $j \in [1, N]$, the output is calculated as:

$$A_j = \text{softmax}(\sigma_j K_j^T Q_j), \text{head}_j = V_j A_j \quad (2)$$

where σ_j represents a learnable scaling factor for each head, $A_j \in \mathbb{R}^{C \times C}$, K_j^T denotes the transposed matrix of K_j . Subsequently, the N heads are concatenated and multiplied by a learnable weight matrix $W \in \mathbb{R}^{C \times C}$. The final output, X_{out} , is obtained after adding a position embedding and reshaping it to $\mathbb{R}^{H \times W \times C}$:

$$H = \text{Concat}(\text{head}_1, \dots, \text{head}_N) \\ X_{out} = HW + f_{pe}(V) \quad (3)$$

$f_{pe}(\cdot)$ is responsible for generating position embeddings, which consists of convolution and activation function, $\text{Concat}(\cdot)$ denotes concatenation along the channel dimension.

3.3. Retinex reflectance extraction module

The RREM utilizes a Siamese network [43] architecture, where two networks share parameters, inspired by the Retinex theory [34], is designed to extract reflectance for spectral reconstruction. In this section, we briefly overview Retinex-Net [38], as its loss functions are utilized for comparison with our proposed method. Retinex-Net is a reflectance extraction method that takes different light images I_{rgb} as input and produces illumination T and reflectance R . The loss function comprises Invariable Reflectance Loss \mathcal{L}_{ir} , Reconstruction Loss \mathcal{L}_{recon} , and Illumination Structure-Aware Smoothness Loss \mathcal{L}_{is} :

$$\mathcal{L}_{ir} = \|R_{low} - R_{normal}\|_1 \quad (4)$$

$$\mathcal{L}_{recon} = \|R \odot T - I_{rgb}\|_1 \quad (5)$$

$$\mathcal{L}_{is} = \|\nabla T \odot \exp(-\lambda_g \nabla R)\|_1 \quad (6)$$

$$\mathcal{L}_{RetinexNet} = \lambda_i \mathcal{L}_{ir} + \lambda_r \mathcal{L}_{recon} + \lambda_s \mathcal{L}_{is} \quad (7)$$

where \odot denotes element-wise multiplication, ∇ represents the gradient, and λ_i , λ_r , λ_s , λ_g are weight coefficients, R_{low} and R_{normal} are the reflectance extracted from low brightness and normal brightness images, respectively.

Our proposed method extracts consistent reflectance from the same retina captured by different imaging systems as shown in Fig. 2(c). We employed a color transformation, T_c , to simulate the effects of varying imaging systems by altering hue and saturation.

$$H' = H \times (1 + \sigma_1) \quad (8)$$

$$S' = S \times (1 + \sigma_2) \quad (9)$$

where H and H' represents the hue of I_{rgb} before and after the change, and S and S' represents the saturation of I_{rgb} before and after the change, $\sigma_1 \sim \text{Uniform}(-a_1, a_1)$, $\sigma_2 \sim \text{Uniform}(-a_2, a_2)$. We empirically set a_1 and a_2 to 0.05 and 0.3, respectively.

The transformed image is denoted as $T_c(I_{rgb})$. Both the original image, I_{rgb} , and the transformed image, $T_c(I_{rgb})$, are used as inputs to a reflectance extraction network, $\mathcal{R}_\theta(\cdot)$, where θ represents learnable parameters. This network generates reflectance maps of R_{rgb} and R'_{rgb} , as denoted in Fig. 2(c). To enforce pixel-wise equality between $\mathcal{R}_\theta(I_{rgb})$ and $\mathcal{R}_\theta(T_c(I_{rgb}))$, we used the Invariable Reflectance loss, \mathcal{L}_{ir} , defined as:

$$\begin{aligned} R_{rgb} &= \mathcal{R}_\theta(I_{rgb}) \\ R'_{rgb} &= \mathcal{R}_\theta(T_c(I_{rgb})) \\ \mathcal{L}_{ir} &= \sum_{i,j} \|R_{rgb}(i,j) - R'_{rgb}(i,j)\|_1 \end{aligned} \quad (10)$$

where i and j denote the coordinates within the circular foreground region of the retinal image.

To ensure the spatial correlation between the reflectance map and the RGB image, we used local cross-correlation (\mathcal{LCC}) [44] defined as:

$$\alpha = \frac{\sum_{x,y} (f(x,y) - \hat{f}(x,y))(g(x,y) - \hat{g}(x,y))}{\left| \sum_{x,y} (f(x,y) - \hat{f}(x,y))(g(x,y) - \hat{g}(x,y)) \right|} \quad (11)$$

$$\mathcal{LCC}(f, g) = \frac{\alpha \left(\sum_{x,y} (f(x,y) - \hat{f}(x,y))(g(x,y) - \hat{g}(x,y)) \right)^2}{\sum_{x,y} (f(x,y) - \hat{f}(x,y))^2 \sum_{x,y} (g(x,y) - \hat{g}(x,y))^2} \quad (12)$$

where $\hat{f}(x,y)$ and $\hat{g}(x,y)$ are the local means of $f(x,y)$ and $g(x,y)$ over a window of size k , which is set to 7, α represents the sign of the numerator in Eq. (11), which is used to ensure that the extracted reflectance and RGB image are positively correlated rather than negatively correlated. The local cross-correlation computes the cross-correlation coefficient between two images at distinct positions, focusing on localized calculations within smaller overlapping regions. Unlike the global approach, which measures overall correlation, this method captures fine-grained local variations, enabling the effective

reconstruction of intricate details. We used reflectance spatial correlation loss \mathcal{L}_{sc} constrains the reflectance and RGB image to maintain the correlation using the \mathcal{LCC} .

$$\mathcal{L}_{sc} = \mathcal{LCC}(I_{rgb}, R_{rgb}) + \mathcal{LCC}(T_c(I_{rgb}), R'_{rgb}) \quad (13)$$

The reflectance extraction loss, \mathcal{L}_{re} , combines these two loss terms as a linear combination:

$$\mathcal{L}_{re} = \lambda_{ir} \mathcal{L}_{ir} - \lambda_{sc} \mathcal{L}_{sc} \quad (14)$$

where λ_{ir} and λ_{sc} are hyperparameters that control the relative weights of \mathcal{L}_{ir} and \mathcal{L}_{sc} , respectively.

The RREM is a self-supervised learning framework that is specifically designed to extract reflectance attributes conducive to spectral reconstruction in fundus images. This innovative approach initiates by simulating variations in image acquisition across different imaging systems through the application of color jittering techniques on the original images. Central to its methodology is the utilization of a weight-shared Siamese network, which operates to extract reflectance features concurrently from these simulated images, while enforcing an invariant reflectance constraint between them. Furthermore, our model capitalizes on the spatial correlation principle, acknowledging the direct association between the brightness levels in RGB images and their corresponding reflectance values. By recognizing that areas of higher luminosity generally signify greater reflectance, the RREM strategically calibrates this correlation. This calibration is executed without the necessity for exact alignment, notably refining the accuracy of our spectral reconstruction endeavors.

3.4. Unsupervised semantic spectral separation module

Different tissue structures exhibit distinct spectral characteristics. Separating and reconstructing tissues in various regions independently can notably enhance spectral reconstruction [30]. However, creating a comprehensive semantic segmentation map for a multitude of tissue types is challenging due to the need for extensive manual annotation.

To address this challenge, we proposed the Unsupervised Semantic Spectral Separation Module (U3SM). The U3SM module is designed to improve spectral reconstruction by semantically separating input information in an unsupervised manner, without the need for ground truth labels. As shown in Fig. 2(d), the U3SM comprises a network for semantic information extraction, which generates an n -channel semantic map denoted as S :

$$S = \{s_1, s_2, \dots, s_n\} \quad (15)$$

where s_i is the i th semantic channel. The element-wise inverse of each s_i is taken to obtain $\sim s_i$, it can be assumed that there is a foreground-background relationship between s_i and $\sim s_i$. The semantic spectral maps (M_{ss}) are defined as:

$$M_{ss} = \text{Concat}(s_1, \dots, s_n, \sim s_1, \dots, \sim s_n) \quad (16)$$

where Concat represents concatenation along the channel dimension. The resulting semantic separated maps serve as masks to segment the input information into different parts, thereby improving spectral reconstruction accuracy. Given an input x to the U3SM, x' is obtained by repeating input $2n$ times along the channel dimension. The output can be represented as:

$$\text{Output} = M_{ss} \odot x' \quad (17)$$

where \odot represents element-wise multiplication (Hadamard product). In the code implementation, $1 - s_i$ is used instead of $\sim s_i$. The effectiveness of the U3SM will be demonstrated in the experimental section.

The U3SM is a design similar to attention modules, encouraging reconstructed multispectral images to have distinct pixels for areas with different semantics in an unsupervised manner. The approach involves

treating each channel of the feature map as a semantic map at a certain level, where this semantic map can be considered as the foreground of certain semantics, and its element-wise inversion can be considered as the background of certain semantics. These separate feature maps are then concatenated by channel and element-wise multiplied with the original image. Through this simple method, notable improvements were observed in our experiments.

3.5. Loss functions

In this study, we formulated the objective loss function as a combination of mean absolute error loss (\mathcal{L}_{mae}), retinal oxygen saturation loss (\mathcal{L}_{so_2}), and reflectance extraction loss (\mathcal{L}_{re}). The \mathcal{L}_{re} loss incorporates terms for invariable reflectance and reflectance structural constraints.

The mean absolute error \mathcal{L}_{mae} is defined as:

$$\mathcal{L}_{mae}(y, \hat{y}) = \frac{1}{N} \sum_{i=1}^N \|y_i - \hat{y}_i\|_1 \quad (18)$$

where N represents the pixels belonging to the foreground region (circular) of the retinal image, $y \in \mathbb{R}^{H \times W \times 2}$ represents the groundtruth spectrum and the reconstructed spectrum is denoted as \hat{y} .

The retinal oxygen saturation loss \mathcal{L}_{so_2} is given by:

$$\mathcal{L}_{so_2}(y, \hat{y}) = \frac{1}{M} \sum_{i=1}^M \|y_i - \hat{y}_i\|_1 \quad (19)$$

where M represents the pixels belonging to the area for calculating retinal oxygen saturation, we used a 7×7 kernel for morphological dilation on the segmented blood vessel image to approximate this region.

The total loss is expressed as:

$$\mathcal{L}_{total} = \lambda_{mae} \mathcal{L}_{mae} + \lambda_{so_2} \mathcal{L}_{so_2} + \mathcal{L}_{re} \quad (20)$$

4. Experiments

4.1. Datasets and experimental setup

In this study, two datasets were employed. The first dataset utilized a dual-modal fundus camera (Hefei Orbis Biotech LTD, Hefei, China) to capture multi-spectral images, which provided the ground truth for oxygen saturation levels. To evaluate the generalizability of the proposed model, a second dataset comprising images acquired with a Canon fundus camera was used. Selection criteria for the dataset construction included images with distinct arterial and venous vessels. The four-fold cross-validation dataset includes 80 RGB images, along with their corresponding multi-modal images at specific wavelengths of 570 nm and 610 nm. Similarly, the second dataset, sourced from the Canon fundus camera, comprises 20 high-quality fundus RGB images and their respective multi-modal counterparts at the same wavelengths (570 nm, 610 nm).

All participants or their legal guardians provided written informed consent, and the study was approved by the ethics committee of the First Affiliated Hospital of USTC (2020-KYLS-167). The data collection procedures strictly adhered to the guidelines outlined in the Declaration of Helsinki.

The model was developed using Python and implemented on the PyTorch platform. All input images were standardized to a size of 480×480 pixels. The batch size was configured to 2, and the Adam optimizer was employed with a learning rate of $1e-4$. Data augmentation techniques, such as rotation, flip, and color jitter, were utilized with saturation and hue values set to 0.3 and 0.05, respectively, for the color jitter. The n of U3SM is set to 2. Parameter values for λ_{mae} , λ_{so_2} , λ_{lr} , and λ_{sc} were assigned as 1, 2, 0.1, and 0.01, respectively. Both training and testing procedures were executed on a GPU server equipped with 4 GTX 3090 units.

Table 1

The comparison with state-of-the-art spectral reconstruction models was performed using four-fold cross-validation.

Model	Param (M)	MAE-ROI	RMSE-ROI	PSNR
HSCNN+ [14]	4.65	4.14 ± 0.18	5.37 ± 0.24	35.53 ± 0.35
HRNet [12]	31.70	3.59 ± 0.21	4.55 ± 0.18	36.76 ± 0.27
HDNet [45]	2.66	3.85 ± 0.12	4.93 ± 0.07	36.18 ± 0.10
HINet [49]	5.21	4.19 ± 0.16	5.53 ± 0.23	35.28 ± 0.34
EDSR [47]	2.42	3.86 ± 0.17	4.99 ± 0.30	36.10 ± 0.45
MIRNet [48]	3.75	4.04 ± 0.09	5.19 ± 0.29	35.80 ± 0.43
MPRNet [46]	3.62	4.07 ± 0.23	5.33 ± 0.24	35.59 ± 0.36
Restormer [50]	15.11	3.46 ± 0.11	4.37 ± 0.23	37.03 ± 0.35
MST++ [31]	0.55	3.51 ± 0.07	4.33 ± 0.10	37.10 ± 0.16
Ours	1.64	3.36 ± 0.09	4.25 ± 0.11	37.22 ± 0.16

4.2. Evaluation metrics

Two evaluations were conducted in this study: pixel-wise evaluation of the reconstructed images compared to the ground truth images and comparison of the calculated oxygen saturation level (SO_2) based on the reconstructed images with measurements obtained using dual-modal fundus cameras.

To evaluate the image reconstruction performance, three metrics were employed: The Mean Absolute Error of the region of interest (MAE-ROI), The Root Mean Squared Error of the region of interest (RMSE-ROI), and Peak Signal-to-Noise Ratio (PSNR). The above ROI refers to the circular region that excluded the pure black background portion.

To assess the error in calculating oxygenation using the reconstructed multispectral images and measurements from dual-mode cameras, Mean Absolute Error (MAE- SO_2) was used. This metric provided a measure of the deviation between the calculated oxygen saturation levels and the ground truth measurements.

5. Results and discussion

In this section, we will present our results from two different perspectives: spectral image reconstruction and SO_2 evaluation accuracy. We have compared our results with previous studies and conducted ablation studies in both aspects. To assess the network's generalization property, we specifically evaluated its performance in terms of SO_2 evaluation accuracy using images captured with Canon cameras. We employed four-fold cross-validation, and the values for MAE-ROI, RMSE-ROI, PSNR and MAE- SO_2 are represented using mean \pm std.

5.1. Multi-spectral image reconstruction performance

5.1.1. Comparisons with state-of-the-arts

In this section, we will compare the proposed spectral reconstruction network, R3ST, with existing state-of-the-art methods, including HSCNN+ [14], HRNet [12], HDNet [45], MST++ [31], image recovery model, MPRNet [46], EDSR [47], MIRNet [48], HINet [49], and Restormer [50]. Our goal is to demonstrate the superiority of the R3ST method. The comparisons are evaluated on the quality of the reconstructed multi-spectral images compared to the ground truth images captured by the dual-modal fundus cameras.

Table 1 elucidates that the methodology introduced in this study markedly outperforms the comparative methods across various metrics, including MAE-ROI, RMSE-ROI, and PSNR. The proposed method shows superior performance, with the mean MAE-ROI recorded at 3.36, the mean RMSE-ROI at 4.25, and the mean PSNR at 37.22. Notably, there is a noticeable improvement in the mean MAE-ROI and PSNR when compared with the next leading technique, Restormer, which exhibits a mean MAE-ROI of 3.46 and a mean PSNR of 37.03. Moreover, the proposed methodology is distinguished by its reduced parameter count (1.64 million), in stark contrast to that of its counterparts

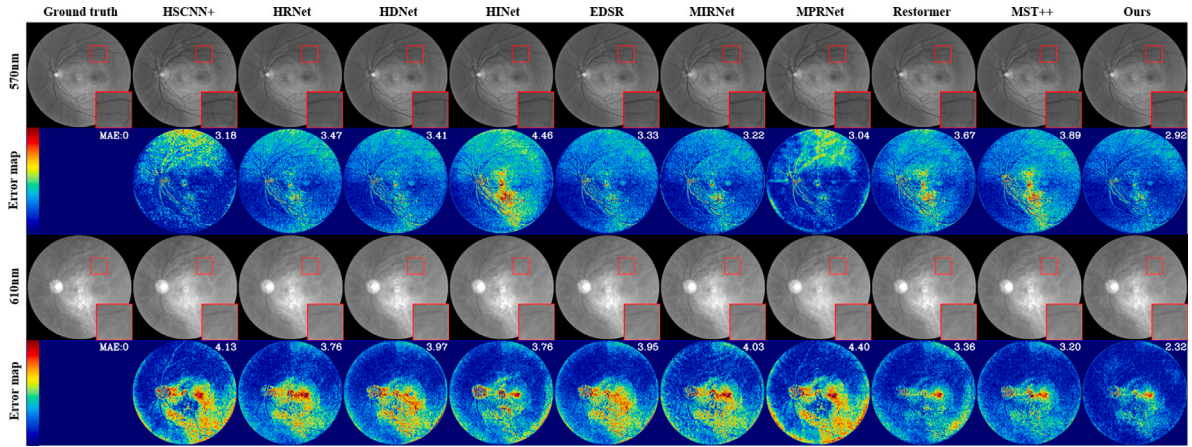


Fig. 3. Comparisons of reconstructed multispectral fundus images. The first and third rows display the visualizations of reconstructed multispectral fundus images, while the second and fourth rows show the error maps of different methods on a testing RGB image. The leftmost column represents the ground truth, and the rightmost column shows the results obtained with our method. The enlarged view in the bottom-right corner highlights specific regions. Compared to alternative approaches, our method demonstrates fewer locally bright regions. The image brightness has been enhanced for better detail observation.

Table 2

The results of the pixel-level ablation study demonstrate the effectiveness of the \mathcal{L}_{so_2} , RREM, and U3SM components of the R3ST in four-fold cross-validation.

Model	MAE-ROI	RMSE-ROI	PSNR
Baseline	3.50 ± 0.07	4.40 ± 0.17	36.88 ± 0.26
Baseline+ \mathcal{L}_{so_2}	3.52 ± 0.08	4.41 ± 0.07	36.87 ± 0.11
Baseline+ \mathcal{L}_{so_2} +RREM	3.41 ± 0.11	4.32 ± 0.19	36.96 ± 0.29
Baseline+\mathcal{L}_{so_2}+RREM+U3SM	3.36 ± 0.09	4.25 ± 0.11	37.22 ± 0.16

such as Restormer (15.11 million) and HRNet (31.70 million), thereby highlighting its efficiency and practicality in applicable contexts.

In Fig. 3, we present the reconstruction results of our method and compare them with other methods. The error maps corresponding to the 570 nm wavelength for HSCNN+ and MPRNet show that veins and arteries are relatively brighter than other regions, while other models typically exhibit darker error maps. This explains why these two models perform poorly in calculating blood oxygen (as shown in Table 5), while our method generally has lower errors in the vascular and surrounding areas, resulting in lower errors when calculating blood oxygen. Furthermore, our method has relatively dark error maps at both 570 nm and 610 nm wavelengths, whereas other models typically have larger errors at one wavelength compared to the other. We attribute this to the inclusion of U3SM, which distinguishes different tissue regions during the reconstruction process, thereby improving multi-spectral reconstruction performance.

5.1.2. Ablation study

The effectiveness of the RREM and the U3SM was demonstrated through ablative experiments, in which the baseline network was a spectral-wise Transformer. The retinal oxygen saturation loss, RREM, and U3SM were gradually introduced into the baseline to evaluate the effectiveness of the different modules. The same metrics used in previous experiments were employed to assess the performance of the various modules.

Improved model performance with RREM: The self-supervised Retinex Reflectance Extraction Module (RREM) notably enhanced the system's performance, as demonstrated in the results presented in Table 1 of the ablation study. The incorporation of RREM resulted in notable improvements in the mean MAE-ROI, mean RMSE-ROI, and mean PSNR metrics, each by approximately 0.1.

These improvements can be attributed to the successful extraction of reflectance that is favorable for spectral reconstruction, as well as the simulation of different imaging systems. Reflectance describes the intrinsic properties of captured objects that remain consistent under

Table 3

The comparison between the four-fold cross-validation results obtained by replacing the \mathcal{L}_{sc} and \mathcal{L}_{re} in RREM with the corresponding loss functions of Retinex-Net [38], as well as replacing RREM with RetinexNet.

Model	MAE-ROI	RMSE-ROI	PSNR
R3ST (replacing \mathcal{L}_{sc})	3.51 ± 0.12	4.61 ± 0.15	36.67 ± 0.22
R3ST (replacing \mathcal{L}_{re})	3.86 ± 0.24	5.06 ± 0.25	35.99 ± 0.37
R3ST (replacing RREM)	4.21 ± 0.23	5.76 ± 0.26	34.93 ± 0.38
R3ST	3.36 ± 0.09	4.25 ± 0.11	37.22 ± 0.16

different brightness conditions and imaging devices. The RREM simplifies the spectral reconstruction task and reduces spectral reconstruction errors caused by variations in imaging systems.

Furthermore, we tested the effectiveness of the \mathcal{L}_{sc} , \mathcal{L}_{re} , and RREM with RetinexNet. We replaced the \mathcal{L}_{sc} in RREM with the \mathcal{L}_{sc} from RetinexNet, the \mathcal{L}_{re} in RREM with the \mathcal{L}_{re} from RetinexNet, and directly replaced all of RREM with RetinexNet. The results, as shown in Table 3, demonstrated that when \mathcal{L}_{sc} , \mathcal{L}_{re} , and RREM were partially or completely replaced by RetinexNet, there was a notable deterioration in the MAE-ROI, RMSE-ROI, and PSNR metrics. Among these, replacing RREM entirely with RetinexNet resulted in the worst performance, which might be due to RetinexNet's network structure being unsuitable for the current task. The experiment proved that our specifically designed RREM performs better than RetinexNet for the current task. Additionally, the RREM focuses solely on the extraction of the image reflectance R and omits the illumination map. Therefore, our method requires only two hyperparameters for reflectance extraction, making it simpler to adjust than the four hyperparameters used in Retinex-Net.

Overall, these findings highlight the importance of the RREM in improving the performance of the proposed model and demonstrate the superiority of our \mathcal{L}_{re} loss function over the one used in Retinex-Net for reflectance extraction.

Enhanced model performance with U3SM: The Unsupervised Semantic Spectral Separation Module (U3SM) underwent testing, and the results from the ablation studies demonstrated notable enhancements across all performance indicators following its incorporation, as depicted in Tables 2 and 6. The implementation of U3SM led to a reduction in the mean MAE-ROI and mean RMSE-ROI by approximately 0.05 and 0.07, respectively, along with an improvement in the mean PSNR by about 0.2.

The proposed U3SM employs unsupervised learning to obtain semantic information, which is subsequently integrated into the spectral reconstruction process. The experiments demonstrate U3SM's efficacy in enhancing multiple metrics of retinal image spectral reconstruction.

Table 4

The comparison between the four-fold cross-validation results obtained by replacing different plug-in type attention modules.

Model	MAE-ROI	RMSE-ROI	PSNR
SE-Net [51]	3.41 ± 0.07	4.43 ± 0.21	36.95 ± 0.31
SK-Net [52]	3.39 ± 0.12	4.40 ± 0.17	36.99 ± 0.25
CBAM [53]	3.42 ± 0.09	4.39 ± 0.13	37.00 ± 0.19
DANet [54]	3.38 ± 0.10	4.26 ± 0.18	37.19 ± 0.26
U3SM	3.36 ± 0.09	4.25 ± 0.11	37.22 ± 0.16

This highlights the importance of semantic information in refining spectral reconstruction, as corroborated by previous studies [30]. The results suggest that U3SM can effectively acquire valuable semantic information for spectral reconstruction without requiring additional semantic labels during training, making it a practical and beneficial solution for enhancing spectral reconstruction quality in clinical settings (see Table 4).

While the introduction of the retinal oxygen saturation loss (\mathcal{L}_{so_2}) did not improve the network performance in terms of image reconstruction, it plays a crucial role in accurately evaluating the calculated oxygen saturation (SO_2). This is highlighted in Table 2, where the network performance slightly decreased with the inclusion of \mathcal{L}_{so_2} . However, the inclusion of this loss function is essential for the accurate evaluation of SO_2 , as discussed in the following sections. Additionally, we compared the U3SM with other plug-in type attention module methods, including SE-Net [51], SK-Net [52], CBAM [53], and DANet [54], by replacing the U3SM in R3ST with these modules. Through this comparison, we verified that U3SM improves the mean MAE-ROI, mean RMSE-ROI, and mean PSNR, and it also has the smallest standard deviation among all the metrics. These experimental results demonstrate the superiority and stability of U3SM.

In summary, our proposed network architecture, R3ST, has demonstrated remarkable improvements in the accuracy of retinal image spectral reconstruction. While \mathcal{L}_{so_2} may not directly enhance image reconstruction performance, it is a vital component for precise SO_2 evaluation.

5.2. SO_2 accuracy and network generalization property

To evaluate the accuracy of SO_2 calculation, the Mean Absolute Error (MAE- SO_2) metric was used to compare the calculated SO_2 values based on the reconstructed images with the measurements obtained using dual-modal fundus cameras. This approach is grounded on the principle that, in calculating retinal oxygen saturation using multispectral images, we analyze not only the pixel values within the blood vessels but also those in the surrounding areas. This method relies on a critical assumption: the spectral differences between the blood vessels and their surrounding areas can reveal changes in blood oxygen saturation [6]. Therefore, when comparing the ground truth spectrum to the spectrum reconstructed by our model, the differences between them reflect changes in oxygen saturation within the retinal area. Specifically, these differences reveal changes in the spectral absorption characteristics caused by variations in blood oxygen levels, thereby indirectly indicating the loss of retinal oxygen saturation.

The network's generalization property was assessed by testing its performance on a different dataset captured by a Canon camera. This evaluation aimed to evaluate the model's ability to adapt to different imaging conditions and maintain accurate SO_2 calculations.

The comparisons are evaluated on the accuracy of the calculated retinal oxygen saturation levels derived from the reconstructed images, compared with the values obtained from the captured images.

Table 5

Comparisons of blood oxygen error calculations with state-of-the-art spectral reconstruction models were conducted using four-fold cross-validation and the Canon camera test dataset.

Model	MAE- SO_2		MAE- SO_2 (Canon)	
	Artery	Vein	Artery	Vein
HSCNN+ [14]	7.80 ± 0.24	5.42 ± 0.40	12.63 ± 1.60	8.53 ± 0.65
HRNet [12]	6.92 ± 0.49	4.12 ± 0.46	9.68 ± 1.18	7.36 ± 0.71
HDNet [45]	7.52 ± 0.19	4.88 ± 0.20	11.84 ± 1.54	7.33 ± 0.64
HINet [49]	8.52 ± 0.47	5.74 ± 0.35	13.06 ± 1.81	9.74 ± 0.89
EDSR [47]	7.02 ± 0.93	4.80 ± 0.38	11.89 ± 1.98	7.57 ± 0.60
MIRNet [48]	7.72 ± 0.65	5.43 ± 0.21	11.92 ± 1.89	8.78 ± 0.63
MPRNet [46]	8.10 ± 0.67	5.64 ± 0.50	12.16 ± 1.93	9.10 ± 0.76
Restormer [50]	6.28 ± 0.24	4.17 ± 0.21	10.11 ± 1.14	6.94 ± 0.43
MST++ [31]	6.39 ± 0.52	3.99 ± 0.16	9.74 ± 1.29	6.87 ± 0.58
Ours	5.89 ± 0.35	3.76 ± 0.21	8.15 ± 1.71	6.21 ± 0.44

Table 6

The results of the ablation study on retinal blood oxygen saturation demonstrate the effectiveness of \mathcal{L}_{so_2} , RREM, and U3SM of the R3ST on both the four-fold cross-validation and the Canon camera test dataset.

Model	MAE- SO_2		MAE- SO_2 (Canon)	
	Artery	Vein	Artery	Vein
Baseline	6.28 ± 0.38	4.02 ± 0.20	9.69 ± 1.12	6.56 ± 0.45
Baseline+ \mathcal{L}_{so_2}	5.97 ± 0.39	3.89 ± 0.46	8.80 ± 0.73	6.28 ± 0.51
Baseline+ \mathcal{L}_{so_2} +RREM	5.92 ± 0.30	3.85 ± 0.35	8.76 ± 1.29	6.30 ± 0.39
Baseline+\mathcal{L}_{so_2}+RREM+U3SM	5.89 ± 0.35	3.76 ± 0.21	8.15 ± 1.71	6.21 ± 0.44

5.2.1. Comparisons with state-of-the-arts

Table 5 presents the results of comparing our model with state-of-the-art spectral reconstruction models regarding the calculation of blood oxygen errors on both the test dataset and the Canon camera test dataset. Our model surpasses all others in performance across both datasets. Specifically, it attains a mean MAE- SO_2 of 5.89 for arteries and 3.76 for veins on the test dataset. These results exceed those of the previously leading models, Restormer and MST++, by margins of 0.5 for arteries and 0.23 for veins, respectively.

When evaluated on the Canon camera dataset, our model demonstrated a mean MAE- SO_2 of 8.15 and 6.21 for arteries and veins, respectively. These results surpass the previous best model, MST++, by margins of 1.6 and 0.66 for arteries and veins, respectively. This performance underscores our model's robustness to various camera configurations and its ability to accurately estimate blood oxygen errors across diverse datasets.

Overall, the results show that our model achieved state-of-the-art performance in terms of the calculation of blood oxygen errors, demonstrating its effectiveness in accurately reconstructing spectral information and facilitating the diagnosis of retinal diseases.

5.2.2. Ablation study

In this study, we utilized the accuracy of SO_2 evaluation as the criterion to assess the performance of the network with various designed modules. This evaluation metric allows us to gauge the effectiveness of the different components in improving the network's ability to calculate retinal oxygen saturation levels accurately.

Improved calculation accuracy with oxygen saturation loss: The introduction of the retinal oxygen saturation loss resulted in a noticeable improvement in the calculation of blood oxygen saturation, as shown in Table 6. However, incorporating this loss led to a slight increase in the MAE- SO_2 , and a slight decrease in MAE-ROI metrics for image reconstruction, as demonstrated in Table 2. This occurs because the retinal oxygen saturation loss enhances the reconstruction of local regions of the image, which are closely related to the blood oxygen calculation region. The results of the experiment emphasize the notable and effective approach of incorporating the retinal oxygen saturation loss in improving blood oxygen calculation.

Improved calculation accuracy with RREM: Incorporating the RREM led to improvements in image quality, blood oxygen calculation accuracy, and cross-camera blood oxygen calculation accuracy. The mean arterial and venous blood oxygen errors were also reduced by approximately 0.05 and 0.04, respectively. Cross-camera testing also demonstrated an improvement, with the arterial error decreasing by approximately 0.04.

Enhanced calculation accuracy with U3SM: The test set results also showed reduced errors in calculating arterial and venous oxygenation by about 0.03 and 0.09, respectively. For the test set derived from different cameras, arterial errors decreased by approximately 0.6, while venous errors were reduced by around 0.1.

5.2.3. Enhanced generalization property

The enhanced generalization property of our proposed model was thoroughly evaluated and demonstrated through comparative analyses.

In Table 5, the performance of our model was compared with other existing models in terms of generalization across different datasets. Our model consistently outperformed the state-of-the-art methods, indicating its robustness and adaptability to diverse imaging conditions. Furthermore, in Table 6, an ablation study was conducted to evaluate the necessity of the designed components in achieving the improved generalization property. The results clearly indicate that the designed components, RREM, U3SM, and \mathcal{L}_{SO_2} , were crucial for enhancing the generalization capabilities of the model. The inclusion of these components resulted in notable improvements in generalization performance.

These findings highlight the importance of the designed components and their contributions to the enhanced generalization property of our model. By incorporating the RREM and U3SM, our model achieved superior generalization across different datasets, making it more reliable and versatile for real-world applications.

6. Conclusion

In this work, we proposed a multispectral reconstruction network R3ST specifically designed for RGB fundus images. The R3ST is designed to enhance the accuracy and generalization of spectral reconstruction across different cameras. R3ST incorporated an unsupervised semantic spectral separation module (U3SM) which leverages unsupervised learning to obtain semantic information. This effectively separates input images based on their tissue structures, leading to improvements in the reconstruction process. Additionally, we employed a retinal oxygen saturation loss function to reduce the error in calculating blood oxygen saturation, thereby enhancing the reconstruction in areas crucial for blood oxygen calculation.

To address the challenge of generalizing spectral reconstruction algorithms to different cameras, we designed the self-supervised Retinex-based reflectance extraction module (RREM). RREM is based on the Retinex algorithm and utilizes random color transformation to simulate the effects of various imaging systems. By extracting reflectance suitable for spectral reconstruction, RREM improved the reconstruction effect across different fundus cameras. Our approach demonstrated improvements in all metrics, achieving notable reductions in errors for retinal blood oxygen saturation calculation in both arteries and veins. The results showcased the effectiveness of our method, with potential implications for reducing costs and improving accessibility to retinal blood oxygen saturation calculation in clinical settings.

Our research marks a notable advancement in the field but is not devoid of limitations, especially when considered for practical applications in clinical environments. Notably, challenges such as suboptimal imaging quality and the variance in image data outside our training set's distribution notably impact the accuracy of our reconstruction technique. Moreover, adapting our method to various camera systems and ensuring precise retinal oxygen level determinations have posed substantial hurdles, highlighting our approach's current limitations in

these aspects. To address these concerns, we are exploring the refinement of our simulation techniques tailored to different imaging systems, aiming to enhance the accuracy of reflectance extraction. Such improvements are anticipated to not only mitigate the identified issues but also elevate our methodology's adaptability and reliability for clinical applications. Furthermore, we are actively pursuing research collaborations and pilot studies in clinical settings to validate and refine our approach, ensuring its effectiveness and utility in real-world scenarios.

CRedit authorship contribution statement

Jicheng Liu: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Methodology, Investigation, Conceptualization. **Wenteng Gao:** Formal analysis, Data curation. **Dehan Zhao:** Investigation. **Lei Yang:** Investigation, Validation. **Peng Liu:** Investigation, Project administration. **Ronald X. Xu:** Project administration, Supervision. **Mingzhai Sun:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Deidentified data, which were gathered for this research investigation, are accessible for academically use at <https://osf.io/8dcuw/>.

References

- [1] M. Wu, M.Y. Lee, V. Bahl, D. Traum, J. Schug, I. Kusmartseva, M.A. Atkinson, G. Fan, K.H. Kaestner, Single-cell analysis of the human pancreas in type 2 diabetes using multi-spectral imaging mass cytometry, *Cell Rep.* 37 (5) (2021).
- [2] T. Antequera, D. Caballero, S. Grassi, B. Uttaro, T. Perez-Palacios, Evaluation of fresh meat quality by hyperspectral imaging (HSI), nuclear magnetic resonance (NMR) and magnetic resonance imaging (MRI): a review, *Meat Sci.* 172 (2021) 108340.
- [3] Y. Zhao, W. Zheng, W. Xiao, S. Zhang, X. Lv, J. Zhang, Rapid monitoring of reclaimed farmland effects in coal mining subsidence area using a multi-spectral UAV platform, *Environ. Monit. Assess.* 192 (2020) 1–19.
- [4] P. Gupta, P. Gupta, Multibiometric authentication system using slap fingerprints, palm dorsal vein, and hand geometry, *IEEE Trans. Ind. Electron.* 65 (12) (2018) 9777–9784.
- [5] C. Li, S. Balla-Arabé, F. Yang, Embedded multi-spectral image processing for real-time medical application, *J. Syst. Archit.* 64 (2016) 26–36.
- [6] S.H. Hardarson, Retinal oximetry, *Acta Ophthalmol.* 91 (thesis2) (2013) 1–47.
- [7] A.B. Einarsson, S.H. Hardarson, J.V. Kristjansson, D.T. Bragason, J. Snaedal, E. Stefánsson, Retinal oximetry imaging in Alzheimer's disease, *J. Alzheimer's Dis.* 49 (1) (2016) 79–83.
- [8] O.B. Olafsdottir, H.S. Saevarsdottir, S.H. Hardarson, K.H. Hannesdottir, V.D. Traustadottir, R.A. Karlsson, A.B. Einarsson, K.D. Jonsdottir, E. Stefansson, J. Snaedal, Retinal oxygen metabolism in patients with mild cognitive impairment, *Alzheimer's Dementia Diagnosis Assess. Dis. Monit.* 10 (2018) 340–345.
- [9] S. Szegedi, P. Dal-Bianco, E. Stögmann, T. Traub-Weidinger, M. Rainer, A. Masching, D. Schmid, R.M. Werkmeister, J. Chua, L. Schmetterer, et al., Anatomical and functional changes in the retina in patients with Alzheimer's disease and mild cognitive impairment, *Acta Ophthalmol.* 98 (7) (2020) e914–e921.
- [10] A.B. Einarsson, O.B. Olafsdottir, H. Hjaltason, S.H. Hardarson, Retinal oximetry is affected in multiple sclerosis, *Acta Ophthalmol.* 96 (5) (2018) 528–530.
- [11] T. Svrčinová, P. Hok, I. Šínová, T. Doriák, M. Král, P. Hübnarová, J. Mareš, P. Kaňovský, M. Šín, Changes in oxygen saturation and the retinal nerve fibre layer in patients with optic neuritis associated with multiple sclerosis in a 6-month follow-up, *Acta Ophthalmol.* 98 (8) (2020) 841–847.
- [12] Y. Zhao, L.M. Po, Q. Yan, W. Liu, T. Lin, Hierarchical regression network for spectral reconstruction from RGB images, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 422–423.

- [13] B. Arad, O. Ben-Shahar, Sparse recovery of hyperspectral signal from natural RGB images, in: *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, the Netherlands, October 11–14, 2016, Proceedings, Part VII 14*, Springer, 2016, pp. 19–34.
- [14] Z. Shi, C. Chen, Z. Xiong, D. Liu, F. Wu, HSCNN+: Advanced cnn-based hyperspectral recovery from RGB images, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 939–947.
- [15] G. ElMasry, D.W. Sun, Principles of hyperspectral imaging technology, in: *Hyperspectral Imaging for Food Quality Analysis and Control*, Elsevier, 2010, pp. 3–43.
- [16] W.R. Johnson, D.W. Wilson, W. Fink, M. Humayun, G. Bearman, Snaps nmshot hyperspectral imaging in ophthalmology, *J. Biomed. Opt.* 12 (1) (2007) 014036.
- [17] H.J. Noordmans, R. De Roode, R. Verdaasdonk, Compact multi-spectral imaging system for dermatology and neurosurgery, in: *Medical Imaging 2007: Physics of Medical Imaging*, vol. 6510, SPIE, 2007, pp. 213–220.
- [18] R. Miyaki, S. Yoshida, S. Tanaka, Y. Kominami, Y. Sanomura, T. Matsuo, S. Oka, B. Raychev, T. Tamaki, T. Koide, et al., Quantitative identification of mucosal gastric cancer under magnifying endoscopy with flexible spectral imaging color enhancement, *J. Gastroenterol. Hepatol.* 28 (5) (2013) 841–847.
- [19] L.E. MacKenzie, A.R. Harvey, Oximetry using multispectral imaging: theory and application, *J. Opt.* 20 (6) (2018) 063501.
- [20] O.B. Olafsdottir, S.H. Hardarson, M.S. Gottfredsdottir, A. Harris, E. Stefánsson, Retinal oximetry in primary open-angle glaucoma, *Invest. Ophthalmol. Vis. Sci.* 52 (9) (2011) 6409–6413.
- [21] S.H. Hardarson, E. Stefánsson, Retinal oxygen saturation is altered in diabetic retinopathy, *Br. J. Ophthalmol.* 96 (4) (2012) 560–563.
- [22] A. Geirsdottir, S.H. Hardarson, O.B. Olafsdottir, E. Stefánsson, Retinal oxygen metabolism in exudative age-related macular degeneration, *Acta Ophthalmol.* 92 (1) (2014) 27–33.
- [23] J. Aeschbacher, J. Wu, R. Timofte, In defense of shallow learned spectral reconstruction from RGB images, in: *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2017, pp. 471–479.
- [24] N. Akhtar, A. Mian, Hyperspectral recovery from RGB images using Gaussian processes, *IEEE Trans. Pattern Anal. Mach. Intell.* 42 (1) (2018) 100–113.
- [25] Z. Xiong, Z. Shi, H. Li, L. Wang, D. Liu, F. Wu, HSCNN: Cnn-based hyperspectral image recovery from spectrally undersampled projections, in: *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2017, pp. 518–525.
- [26] Y.B. Can, R. Timofte, An efficient CNN for spectral reconstruction from RGB images, 2018, arXiv preprint arXiv:1804.04647.
- [27] K.G. Lore, K.K. Reddy, M. Giering, E.A. Bernal, Generative adversarial networks for spectral super-resolution and bidirectional RGB-to-multispectral mapping, in: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, CVPRW, IEEE*, 2019, pp. 926–933.
- [28] A. Rangnekar, N. Mokashi, E. Ientilucci, C. Kanan, M. Hoffman, Aerial spectral super-resolution using conditional adversarial networks, 2017, arXiv preprint arXiv:1712.08690.
- [29] L. Zhang, Z. Lang, P. Wang, W. Wei, S. Liao, L. Shao, Y. Zhang, Pixel-aware deep function-mixture network for spectral super-resolution, in: *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34, No. 07, 2020, pp. 12821–12828.
- [30] Z. Zhu, H. Liu, J. Hou, H. Zeng, Q. Zhang, Semantic-embedded unsupervised spectral reconstruction from single RGB images in the wild, in: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 2279–2288.
- [31] Y. Cai, J. Lin, Z. Lin, H. Wang, Y. Zhang, H. Pfister, R. Timofte, L. Van Gool, MST++: Multi-stage spectral-wise transformer for efficient spectral reconstruction, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 745–755.
- [32] Y. Fu, T. Zhang, Y. Zheng, D. Zhang, H. Huang, Joint camera spectral sensitivity selection and hyperspectral image recovery, in: *Proceedings of the European Conference on Computer Vision, ECCV*, 2018, pp. 788–804.
- [33] B. Kaya, Y.B. Can, R. Timofte, Towards spectral estimation from a single RGB image in the wild, in: *2019 IEEE/CVF International Conference on Computer Vision Workshop, ICCVW, IEEE*, 2019, pp. 3546–3555.
- [34] E.H. Land, The retinex theory of color vision, *Sci. Am.* 237 (6) (1977) 108–129.
- [35] S. Wang, J. Zheng, H.M. Hu, B. Li, Naturalness preserved enhancement algorithm for non-uniform illumination images, *IEEE Trans. Image Process.* 22 (9) (2013) 3538–3548.
- [36] X. Fu, D. Zeng, Y. Huang, Y. Liao, X. Ding, J. Paisley, A fusion-based enhancing method for weakly illuminated images, *Signal Process.* 129 (2016) 82–96.
- [37] X. Ren, M. Li, W.H. Cheng, J. Liu, Joint enhancement and denoising method via sequential decomposition, in: *2018 IEEE International Symposium on Circuits and Systems, ISCAS, IEEE*, 2018, pp. 1–5.
- [38] C. Wei, W. Wang, W. Yang, J. Liu, Deep retinex decomposition for low-light enhancement, 2018, arXiv preprint arXiv:1808.04560.
- [39] Z. Shu, E. Yumer, S. Hadap, K. Sunkavalli, E. Shechtman, D. Samaras, Neural face editing with intrinsic image disentangling, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 5541–5550.
- [40] X. Fu, D. Zeng, Y. Huang, X.P. Zhang, X. Ding, A weighted variational model for simultaneous reflectance and illumination estimation, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2782–2790.
- [41] W. Ma, J.M. Morel, S. Osher, A. Chien, An L1-based variational model for Retinex theory and its application to medical images, in: *CVPR 2011, IEEE*, 2011, pp. 153–160.
- [42] P. Zhuang, C. Li, J. Wu, Bayesian retinex underwater image enhancement, *Eng. Appl. Artif. Intell.* 101 (2021) 104171.
- [43] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, R. Shah, Signature verification using a "siamese" time delay neural network, *Adv. Neural Inf. Process. Syst.* 6 (1993).
- [44] B.B. Avants, C.L. Epstein, M. Grossman, J.C. Gee, Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain, *Med. Image Anal.* 12 (1) (2008) 26–41.
- [45] X. Hu, Y. Cai, J. Lin, H. Wang, X. Yuan, Y. Zhang, R. Timofte, L. Van Gool, HDNET: High-resolution dual-domain learning for spectral compressive imaging, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 17542–17551.
- [46] S.W. Zamir, A. Arora, S. Khan, M. Hayat, F.S. Khan, M.H. Yang, L. Shao, Multi-stage progressive image restoration, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 14821–14831.
- [47] B. Lim, S. Son, H. Kim, S. Nah, K. Mu Lee, Enhanced deep residual networks for single image super-resolution, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 136–144.
- [48] S.W. Zamir, A. Arora, S. Khan, M. Hayat, F.S. Khan, M.H. Yang, L. Shao, Learning enriched features for real image restoration and enhancement, in: *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16*, Springer, 2020, pp. 492–511.
- [49] L. Chen, X. Lu, J. Zhang, X. Chu, C. Chen, HINet: Half instance normalization network for image restoration, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 182–192.
- [50] S.W. Zamir, A. Arora, S. Khan, M. Hayat, F.S. Khan, M.H. Yang, Restormer: Efficient transformer for high-resolution image restoration, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 5728–5739.
- [51] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7132–7141.
- [52] X. Li, W. Wang, X. Hu, J. Yang, Selective kernel networks, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 510–519.
- [53] S. Woo, J. Park, J.Y. Lee, I.S. Kweon, CBAM: Convolutional block attention module, in: *Proceedings of the European Conference on Computer Vision, ECCV*, 2018, pp. 3–19.
- [54] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, H. Lu, Dual attention network for scene segmentation, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 3146–3154.