symmetry

MDPI

Article

Deep Learning-Based Exposure Asymmetry Multispectral Reconstruction from Digital RGB Images

Jinxing Liang [1,2], Xin Hu [1], Wensen Zhou [1], Kaida Xiao [3] and Zhaojing Wang [1,2,*]

[1] School of Computer Science and Artificial Intelligence, Wuhan Textile University, Wuhan 430200, China; jxliang@wtu.edu.cn (J.L.); 2315363120@wtu.edu.cn (X.H.); 2215363075@wtu.edu.cn (W.Z.)
[2] Engineering Research Center of Hubei Province for Clothing Information, Wuhan 430200, China
[3] School of Design, University of Leeds, Leeds LS2 9JT, UK; k.xiao1@leeds.ac.uk
* Correspondence: zjwang@wtu.edu.cn

Abstract: Multispectral reconstruction is an important way to acquire spectral images with a high spatial resolution as snapshots. Current deep learning-based multispectral reconstruction models perform well under symmetric conditions, where the exposure of training and testing images is consistent. However, further research has shown that these models are sensitive to exposure changes. When the exposure symmetry is not maintained and testing images are input into the multispectral reconstruction model under different exposure conditions, the reconstructed multispectral images tend to deviate from the real ground truth to varying degrees. This limitation restricts the robustness and applicability of the model in practical scenarios. To address this challenge, we propose an exposure estimation multispectral reconstruction model of EFMST++ with data augmentation and optimized deep learning architecture, where Retinex decomposition and a wavelet transform are introduced into the proposed model. Based on the currently available dataset in this field, a comprehensive comparison is made between the proposed and existing models. The results show that after the current multispectral reconstruction models are retrained using the augmented datasets, the average MRAE and RMSE of the current most advanced model of MST++ are reduced from 0.570 and 0.064 to 0.236 and 0.040, respectively. The proposed method further reduces the average MRAE and RMSE to 0.229 and 0.037, with the average PSNR increasing from 27.94 to 31.43. The proposed model supports the use of multispectral reconstruction in open environments.

Keywords: digital imaging; RGB images; multispectral reconstruction; deep learning; exposure asymmetry

check for updates

Academic Editor: Sergei Odintsov

Received: 29 December 2024
Revised: 6 February 2025
Accepted: 12 February 2025
Published: 13 February 2025

Citation: Liang, J.; Hu, X.; Zhou, W.; Xiao, K.; Wang, Z. Deep Learning-Based Exposure Asymmetry Multispectral Reconstruction from Digital RGB Images. *Symmetry* **2025**, *17*, 286. https://doi.org/10.3390/sym17020286

1. Introduction

Compared to RGB images, multispectral images encompass more bands, providing richer information that allows for characterizing the physical and chemical properties of materials in detail, while overcoming the metamerism problem in colorimetry. Therefore, multispectral imaging is widely used in material identification, high-fidelity color reproduction, agriculture monitoring, remote sensing, and so on [1–3]. Regarding spectral acquisition, current multispectral imaging devices still face challenges such as having insufficient spatial resolution and being time-consuming, limiting their applications. To improve the efficiency of acquiring multispectral images and reduce costs, there have been widespread investigations into multispectral reconstruction (SR) from RGB images in the past twenty years [4–12]. The principle is to construct a mapping model from response to spectral reflectance using a spectral reconstruction algorithm under specified imaging conditions using training samples with known spectral reflectance and RGB response, which can be applied to test samples under the same imaging conditions for spectral reconstruction to generate a multispectral image of the test target.

The key to this technique lies in multispectral reconstruction methods, which can be broadly categorized into machine learning-based and deep learning-based methods. The former category mainly includes pseudo-inverse, interpolation, principal component analysis (PCA), and kernel-based methods [4–6]. With the advancement of deep learning, numerous network models have also been proposed to address multispectral reconstruction problems, achieving remarkable results. Notably, the NTIRE Spectral Recovery Challenge, organized by the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), has showcased many impressive reconstruction models, significantly boosting the development of deep learning-based multispectral reconstruction models.

Shi et al. proposed the HSCNN+ network based on HSCNN by replacing conventional convolutional layers with residual blocks and deepening the network structure, which ranked first and second in the 'clean' and 'real' tracks of the NTIRE 2018 Spectral Reconstruction Challenge, respectively [7]. Li et al. introduced a new weighted adaptive network (AWAN) by considering the interdependence between the camera spectral sensitivity (CSS) and intermediate features [8]. Zhao et al. proposed a four-level hierarchical regression network (HRNet), utilizing residual dense blocks to remove artifacts from real-world RGB images and employing residual global blocks to construct an attention mechanism that extends the perceptual domain [9]. These two networks were ranked first in the 'clean' and 'real' tracks of the NTIRE 2020 Spectral Reconstruction Challenge. In recent years, the Transformer has gained widespread attention for its advantages in capturing long-range correlations between spatial regions [10], which has opened new opportunities for multispectral reconstruction algorithms. Cai et al. improved the MSA in Transformer by adopting a U-shaped structure in tandem to extract multi-scale contextual information, and proposed the MST++ network, which was ranked first in the 2022 Spectral Reconstruction Challenge [11], further enhancing the accuracy of spectral reconstruction algorithms. Wang et al. proposed the SSTU network, which combines the Swin Transformer with U-net to overcome the high computational cost of the Transformer architecture while preserving spatial features, thereby providing high-quality data with both high spatial and spectral resolution for tissue pathology [12].

However, these algorithms or models have focused on improving the accuracy of multispectral reconstruction while neglecting the robustness of these models to expose changes in practical applications. Lin and Finlayson et al. found that existing multispectral reconstruction algorithms are sensitive to exposure [13]. As shown in Figure 1, even under the same light source, using a model constructed at one specific exposure level (such as Exposure*1) to reconstruct the spectral reflectance of objects at other exposure levels (such as Exposure*K) can cause the reconstructed spectral curves to deviate from the truth spectral curves to varying degrees when the exposure conditions are asymmetric.

This deviation prevents the accurate application of reconstructed spectra for object analysis and identification. A proposed multispectral reconstruction model that can counteract the effects of exposure asymmetry should satisfy Equations (1) and (2):

$$SR(\mathbf{d}) \approx \mathbf{r}_{rec}, \tag{1}$$

$$SR(K \cdot \mathbf{d}) \approx K \cdot \mathbf{r}_{rec}, \tag{2}$$

where $SR()$ is the reconstruction model, $\mathbf{d}$ is the camera response value, $\mathbf{r}_{rec}$ is the reconstructed multispectral data, and K is the exposure adjustment factor. Lin et al. tested the traditional algorithm and the HSCNN++ model, and they found that the highly nonlinear network model's performance is dependent on the exposure. They then introduced the exposure factor scaling to the proposed HSCNN-R_{pd} model, which sacrifices spectral reconstruction accuracy for improved ability to cope with exposure inconsistency [14].

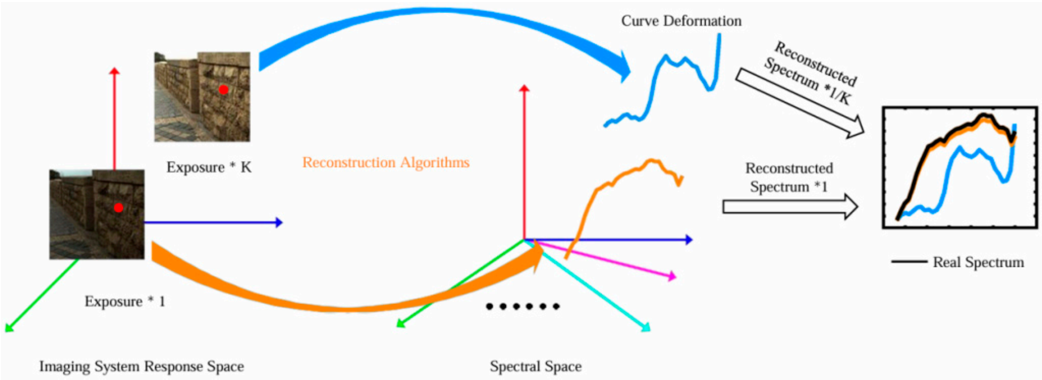

Figure 1. Schematic diagram of the multispectral reconstruction algorithms' sensitivity to changes in exposure level. Using a multispectral reconstruction algorithm of exposure asymmetry to establish a reconstruction model at one exposure level (Exposure*1) to reconstruct the spectral reflectance of objects at different exposure levels (Exposure*1, Exposure*K) can cause the reconstructed spectral curves (orange curves for Exposure*1, blue curves for Exposure*K) to deviate from the ground truth (black curve) to varying degrees.

Liang et al. enhanced Zhang's model by augmenting the training data and adding the attention mechanism to realize the exposure invariant and to improve its performance [15,16]. However, their study was based on the linear synthesized raw images and cannot fully represent the nonlinear real captured images. Additionally, in their study, the reconstructed multispectral images should be multiplied by $1/K$ to achieve the right results regarding the ground truth. This is difficult to implement in reality as the value of K is often unknown. Some unsupervised learning models have also been investigated to enhance the practical feasibility of the SR model [17,18], but this requires more consistency in the RGB images and is difficult to adapt to situations such as exposure changes.

In image processing, some methods have been used to solve image quality problems caused by over- or under-exposure. Histogram equalization improves the visual effect by redistributing the gray levels of an image to enhance contrast and brightness [19,20]. However, these methods usually ignore structural information in the image, which can lead to a loss of detail. High dynamic range (HDR) restoration [21,22] and image fusion techniques [23,24] recover exposure levels by fusing multiple images under different exposure conditions, but the high data requirements limit their popularity in practical applications. Techniques based on Retinex theory [25,26] decompose images into reflectance and illumination components to improve the visual quality of images. However, these methods often fail to adequately account for the corruption of exposure error images, leading to severe noise and color distortion in the enhancement process. There are likewise many neural network models [27–29] active in exposure correction tasks, but these models usually focus on dealing with low-light image enhancement and are applied to the field of spectral reconstruction with unknown effectiveness.

In this study, we first tested the current advanced deep learning-based multispectral reconstruction models and found that they are all sensitive to exposure changes. Therefore, we augmented the dataset and used it to retrain the current advanced models. Although there is still some degradation in the models' performance in tests on their exposure to their original version, MST++ performs within an acceptable range. After that, by using data augmentation and introducing a Transformer-based Retinex decomposition mechanism to reduce the impact caused by exposure changes and using wavelet transform convolution to emphasize the low-frequency information, we propose a new end-to-end network to improve its exposure symmetry performance.

2. Methodology

Unlike Lin's assumptions in Equations (1) and (2), we define the ability to combat exposure variations as Equation (3) using deep learning's advanced data representation and processing capabilities. In other words, our objective, similar to the image enhancement task [30], is to directly reconstruct the multispectral image under different exposure settings.

$$SR(K \cdot \mathbf{d}) \approx \mathbf{r}_{rec}, \tag{3}$$

Our model architecture is shown in Figure 2, consisting of an exposure estimator (EE, Figure 2a) and a spectral reconstruction module (EFMST++, Figure 2b). The EE module is composed of three convolutional blocks. The core components of EFMST++ are the EGT and SST modules, as detailed in Figure 2c,d.

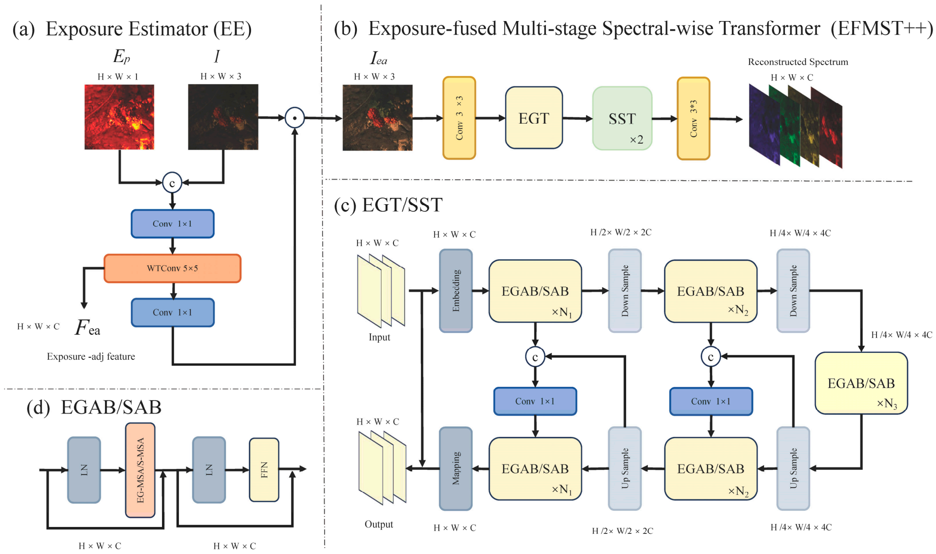

Figure 2. Architecture of our model: (**a**) exposure estimator; (**b**) spectral reconstructor; (**c**) EGT and SST; (**d**) EGAB and SAB. The entire network consists of two components, (**a**,**b**). The input image ***I*** undergoes exposure correction through (**a**) to obtain the corrected image ***Iea***, which is then processed by (**b**) to generate the final multispectral image. (**c**,**d**) provide detailed structures of the module in (**b**).

Our model aims to first recover the image from the exposure and then reconstruct the multispectral image from the recovered image. For this purpose, we try to integrate the Retina theory [31,32], which is commonly used in the field of image enhancement, into our reconstruction model. The traditional Retinex algorithm simulates the human visual perception of luminance and color, decomposing the image $\boldsymbol{I} \in \mathrm{R}^{\mathrm{H}\times\mathrm{W}\times 3}$ into an indication of the reflectance $\boldsymbol{R} \in \mathrm{R}^{\mathrm{H}\times\mathrm{W}\times 3}$ and the exposure component $\boldsymbol{E} \in \mathrm{R}^{\mathrm{H}\times\mathrm{W}}$ as in Equation (4):

$$I = R \odot E \tag{4}$$

where $\odot$ denotes element-wise multiplication, R is the intrinsic property of the object, and E denotes the illumination condition. As shown in Equation (4), the algorithm does not account for the noise and artifacts introduced during image acquisition. In addition, over-exposure and under-exposure can further exacerbate image quality degradation. Therefore, we adopt the uptake model proposed by Cai et al. [33] and incorporate the exposure components $\widetilde{E} \in \mathrm{R}^{\mathrm{H}\times\mathrm{W}}$ and reflection components $\widetilde{R} \in \mathrm{R}^{\mathrm{H}\times\mathrm{W}\times 3}$ into the original equation. By incorporating these components into Equation (4), we obtain Equation (5). Then, by multiplying both sides by the exposure-adj map $\hat{E}$ to ensure that $\hat{E} \odot E = 1$, we derive Equation (6).

$$I = \left(R + \widetilde{R}\right) \odot \left(E + \widetilde{E}\right) \tag{5}$$

$$I \odot \hat{E} = R + R \odot (\hat{E} \odot \widetilde{E}) + \left(\widetilde{R} \odot (E + \hat{E})\right) \odot \hat{E} \tag{6}$$

where $\left(\widetilde{R} \odot (E + \hat{E})\right) \odot \hat{E}$ denotes noise and artifacts after being affected by the exposure-adj map, and $R \odot (\hat{E} \odot \widetilde{E})$ denotes over-exposure or under-exposure as well as color distortion caused by the $\hat{E}$ adjustment process. We can simplify Equation (6) as

$$Iea = I \odot \hat{E} = R + C \tag{7}$$

where $Iea \in \mathrm{R}^{\mathrm{H}\times\mathrm{W}\times 3}$ denotes the image adjusted by the exposure estimator and $C \in \mathrm{R}^{\mathrm{H}\times\mathrm{W}\times 3}$ denotes the overall corruption term. Thus, our network can be represented as

$$(Iea, Fea) = EE(I, Ep), \tag{8}$$

$$Rrec = EFMST(Iea, Fea), \tag{9}$$

where EE denotes the exposure estimator and $EFMST$ denotes the spectral reconstructor. The image I and illumination priori E_p (E_p = $\mathrm{mean_c}(I)$) are input into EE to obtain the exposure-adjusted image I_{ea} and the exposure-adjusted feature F_{ea}, which are then input into the *EFMST*++ module for spectral reconstruction to obtain R_{rec}.

2.1. *Exposure Estimator (EE)*

The architecture of EE is shown in Figure 2a. First, we fuse the inputs I and E_p and apply a 1 × 1 convolution to expand the channel dimension. Instead of using traditional convolutions that focus on high-frequency information [34], we employ the third-order WTConv [35], which is more focused on low-frequency information (as shown in Figure 3a). This also expands the receptive field to further extract features. Finally, another 1 × 1 convolution is applied to recover the three-channel exposure adjustment map $\hat{E}$. The exposure-adjusted image I_{ea} is then obtained using element-wise multiplication of EE and I.

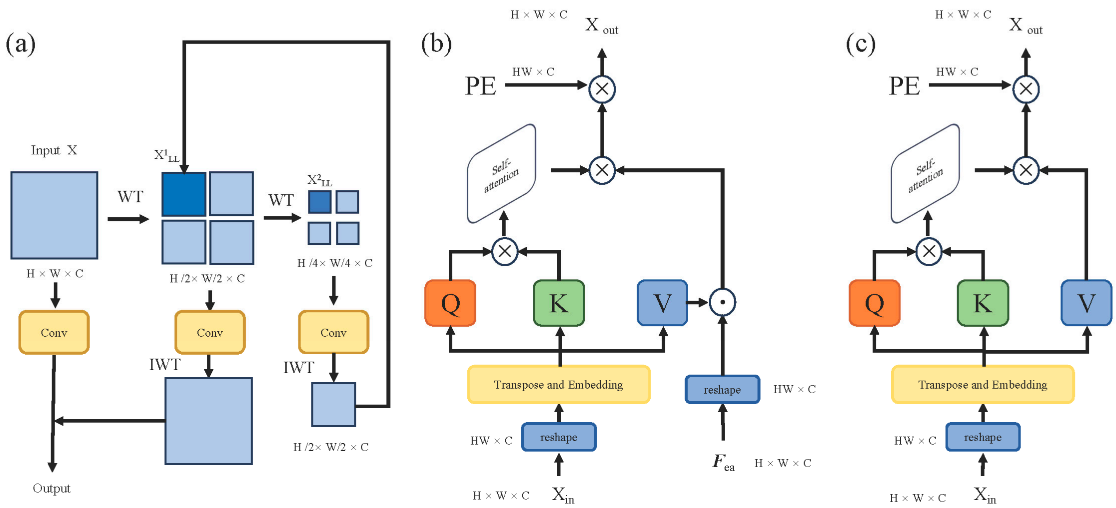

Figure 3. Architecture of (**a**) WTConv; (**b**) EG-MSA; (**c**) S-MSA.

2.2. *Exposure-Fused Multi-Stage Spectral-Wise Transformer (EFMST++)*

The structure of EFMST++ is shown in Figure 2b. It consists of two 3*3 convolutions, one EGT module, and two SST modules connected in series. The first 3*3 convolution performs the dimensional transformation, then the features are further extracted by the EGT and SST modules, and the last 3*3 convolution obtains the final spectral image. The U-shaped structure of SST and EGT is shown in Figure 2c, which consists of a symmetric encoder and decoder, with embedding and mapping blocks as a single 3*3 convolutional layer. The encoder reduces computational complexity and extracts deep features through two layers of down-sampling. Each layer consists of an attention block (EGT or SST) and a 4×4 strided convolution. After each layer, the width and height of the features are halved, while the feature dimension is doubled. The decoder is symmetric to the encoder and performs two layers of up-sampling on the features. Each layer contains a deconv 2×2 operation and an attention block. After each convolutional layer, the width and height of the features are doubled, while the feature dimension is halved. Additionally, skip connections are used between the encoder and decoder to minimize information loss during the sampling process.

2.3. EGT and SST Modules

The difference between EGT and SST lies in the EGAB and SAB, while the difference between EGAB and SAB lies in the EG-MSA and S-MSA, whose structures are shown in Figures 2d and 3. Both EGAB and SAB consist of two normalization layers, a self-attention block (EG-MSA or S-MSA), and a feed-forward network. The intrinsic principle of the S-MSA [8] is to transpose the input features before the self-attention calculation to compute the self-attention along the spectral dimension. However, considering that there may be different exposure conditions in the same image, regions with better exposure conditions can provide semantic contextual representations to help regions with weak exposure conditions. Therefore, we used EG-MSA with the exposure feature Fea added to S-MSA to allow regions with different exposure conditions to interact with each other to guide the computation of self-attention.

3. Experiment

The dataset ARAD_1K provided by NTIRE 2022 was used [36], which contains 1000 data pairs, and was divided into training, validation, and test sets at a ratio of 18:1:1. Each HSI at a size of 482×512 has 31 wavelengths from 400 nm to 700 nm. Due to the nature of the competition, the test set data are not public, and we also used the validation set results to judge the model effect, just as with other participating models.

We first chose the models that won the Spectral Reconstruction Challenge as test targets. For each model, the hyperparameters, learning rate, and other settings were kept consistent with their original version. It is worth noting that, unlike earlier models, MST++ uses max/min normalization to process the data and achieve excellent reconstruction accuracy. We were inspired by this, as the operation means that incorrectly exposed images undergo an initial exposure correction before being fed into the model, opening up more possibilities for the model to combat exposure variations. We then used the same max/min normalization to process the data when testing other models. The model performance is evaluated using three different metrics as usual. The first metric is the mean relative absolute error (MRAE) the second is the root-mean-square error (RMSE), as shown in Equations (10) and (11), and the third metric is the peak signal-to-noise ratio (PSNR).

$$MRAE = \frac{1}{N}\sum_{p=1}^{N}\left(\left|\mathbf{I}_{HSI}^{p} - \mathbf{I}_{SR}^{p}\right| / \mathbf{I}_{HSI}^{p}\right), \tag{10}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{p=1}^{N}\left(\mathbf{I}_{HSI}^{p} - \mathbf{I}_{SR}^{p}\right)^{2}}, \tag{11}$$

where $\mathbf{I}_{HSI}$ and $\mathbf{I}_{SR}$ denote the ground truth and reconstructed spectral reflectance, respectively, and N is the number of reflectance samples; for spectral wavelengths ranging from 400 nm to 700 nm with sampling intervals of 10 nm, N is equal to 31. The smaller the MRAE or RMSE, the better the performance of the model.

The adaptability of these models to variations in input image exposure was initially evaluated. Using the published version of each model, we adjusted the exposure level of all tested images with different values of K (K = 0.2, 0.75, 1, 1.5, 2.5) and fed them into the released reconstruction model. After that, we calculated the average reconstruction error of all the tested images; a summary of the results is presented in Table 1. We can see that the reconstruction error of the four advanced models increased with the value change of K, and the higher the exposure deviation from K = 1, the larger the reconstruction error. The results in Table 1 show that the current deep learning models are all exposure-dependent.

Table 1. Results of different published models tested with different exposure coefficients.

Method	K = 0.2			K = 0.75			K = 1			K = 1.5			K = 2.5			*Average*		
	MRAE	**RMSE**	**PSNR**	**MRAE**	**RMSE**	**PSNR**	**MRAE**	**RMSE**	**PSNR**	**MRAE**	**RMSE**	**PSNR**	**MRAE**	**RMSE**	**PSNR**	**MRAE**	**RMSE**	**PSNR**
HSCNN++	0.546	0.086	23.74	0.404	0.061	26.12	0.381	0.059	26.36	0.477	0.069	24.64	0.931	0.116	19.81	0.548	0.078	24.12
AWAN	0.415	0.054	27.67	0.275	0.040	30.29	0.250	0.037	31.22	0.477	0.049	28.06	0.994	0.101	22.32	0.482	0.056	27.91
HRNet	0.432	0.064	25.82	0.361	0.055	26.74	0.348	0.055	26.89	0.539	0.069	24.98	1.013	0.120	20.12	0.539	0.073	24.91
Liang	0.466	0.062	25.71	0.415	0.056	26.51	0.408	0.054	26.62	0.545	0.071	24.16	1.211	0.140	18.52	0.609	0.077	24.30
MST++	0.352	0.046	28.75	0.205	0.029	32.38	0.165	0.025	34.32	0.566	0.062	26.23	1.563	0.156	18.04	0.570	0.064	27.94

To solve the defect in current models, as shown in Table 1, we treat the multispectral reconstruction as the image enhancement task and, at the same time, try to make these models exposure-invariant, to improve the reconstruction accuracy. This calls for a new image dataset containing various exposure levels for each image, along with the corresponding ground truth multispectral images. However, no such dataset is currently available.

Therefore, we created an augmented dataset based on the ARAD 1K dataset by multiplying the original image with four random values of K (random Seed = 0). To keep the data balance of over-exposed and under-exposed images in the augmented dataset, two random values of K between 0 and 1 and two values of K between 1 and 3 were set in creating the database. Therefore, together with the original image, we obtained five random exposures of each original image, and a total of 4500 training images and 250 validation digital images, corresponding to 900 training and 50 validation multispectral images, were included in the augmented data. Some of the digital images are shown in Figure 4.

Figure 4. Selected digital images from the augmented dataset.

We retrained these models using the augmented dataset, adjusting the batch size to 10 for memory reasons and keeping all other settings unchanged. The results are summarized in Table 2. We can see that the models trained with the new dataset show a significant increase in reconstruction error, compared with the results of $K = 1$ in Table 1. Among them, the SOTA model MST++ for spectral reconstruction from RGB in NTIRE 2022 still showed the best results among these advanced models. However, the reconstruction errors of MRAE and RMSE still significantly increased from 0.165 and 0.025 to 0.236 and 0.040, while the PSNR decreased from 34.32 to 30.96. For the other tested models, even though they were retrained using the augmented dataset and may have a certain ability to combat exposure changes (compared with the overall results of other K values in Table 1), except for the AWAN model, they showed relatively poor results. All the models apart from the MST++ model are based on traditional CNN architectures, which have limitations in capturing long-range dependencies between spatial regions. This is the reason for their relatively poor performance in terms of accuracy. In contrast, the MST++ model introduces a global spatial-wise MSA, which leverages long-range dependencies in the spectral dimension to enhance reconstruction accuracy. However, it overlooks the relationships between local regions in the spatial domain, which is particularly critical for images requiring exposure correction.

Table 2. Results of training and testing of different models using the augmented dataset.

Method	MRAE	RMSE	PSNR
HSCNN++	0.459	0.078	24.45
AWAN	0.306	0.051	28.88
HRNet	0.429	0.071	25.39
Liang	0.516	0.089	23.37
MST++	0.236	0.040	30.96

The proposed model in Figure 2 was also trained with the augmented dataset. The parameter optimization algorithm was the Adam modification. The learning rate was initialized as 0.0004 and the cosine annealing scheme was adopted for 300 epochs. The results are shown in Table 3 and compared with those for the MST++ model.

Table 3. Comparison of reconstruction results for two models.

Method	MRAE	RMSE	PSNR
Ours	0.229	0.037	31.43
MST++	0.236	0.040	30.96

The results in Table 3 show that our model was better than the MST++ for all three evaluation metrics. In addition, for a more intuitive comparison of the multispectral reconstruction results between these two models, we also show some examples of the results. Figure 5 illustrates the reconstructed results of the test image ARAD_1K_0933 in different and random exposure levels. The column furthest to the left in Figure 5 is a digital image of the ARAD_1K_0933, and the comparison of reconstructed spectral reflectance curves of ROI is indicated by the yellow square.

The GT column in Figure 5 shows the ground truth multispectral image at wavelengths of 430 nm, 500 nm, 570 nm, and 640 nm from top to bottom, for the ROI indicated by the red rectangle. The other columns show a detailed comparison between our model and the MST++ model based on different exposure levels and different wavelengths. The lower-left corner shows the reconstructed spectral curves for the two methods at different exposures within the yellow matrix region of the image, including the GT values; by zooming in, we can see that although our model has improved the multispectral reconstruction accuracy to some extent, both our model and the MST++ model showed some noise and artifacts of the reconstructed multispectral image in specific wavelengths, especially in the long wavelength of the visible range. These noises and artifacts are undoubtedly caused by errors in multispectral reconstruction.

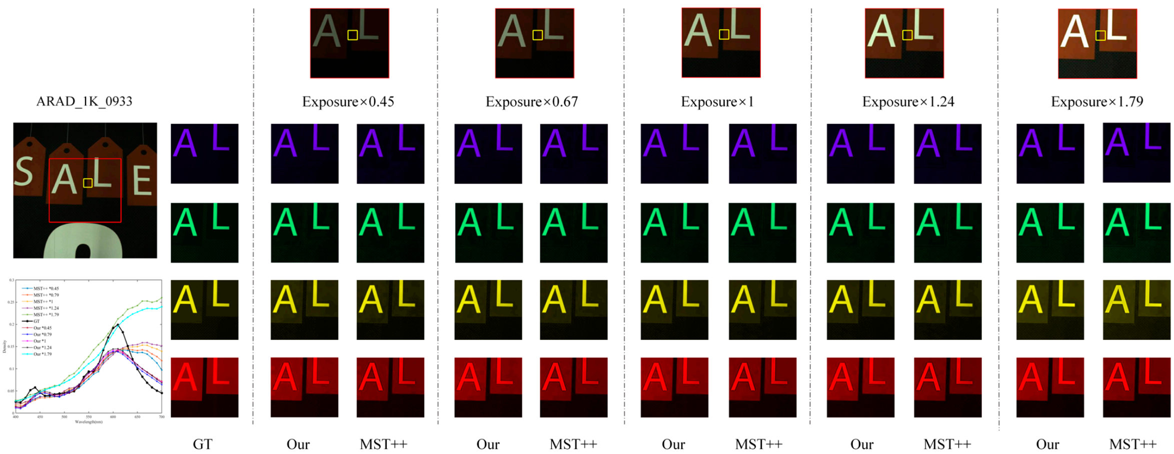

Figure 5. Example of multispectral reconstruction result comparison between our model and MST++ model according to different exposure levels.

The reconstructed spectral reflectance in Figure 5 shows that the reconstructed spectral reflectance curve of ROI is significantly shifted from the ground truth after 600 nm, while our model performs better than the MST++ model for most exposure levels. However, for the exposure value of 1.79, the reconstructed spectral curves of both our model and the MST++ model show serious deviation from the ground truth. This may be caused by the over-exposure of the digital image. We also selected three validation datasets (ARAD_1K_0916, ARAD_1K_0923, ARAD_1K_0945) containing a larger range of exposure values to observe this phenomenon in Figure 6. The presentation is the same as for ARAD_1K_0933, and the exposure magnification is included in the legend. At lower exposure values, both methods still achieve better reconstruction results, but the curve shifts significantly with increasing exposure magnification. To further analyze and compare the performance of our model and the MST++ model, we counted their error metrics in under-exposure ($K < 1$), normal exposure ($K = 1$), and over-exposure ($K > 1$). These are summarized in Table 4. The corresponding intuitive error comparison is plotted in Figure 7.

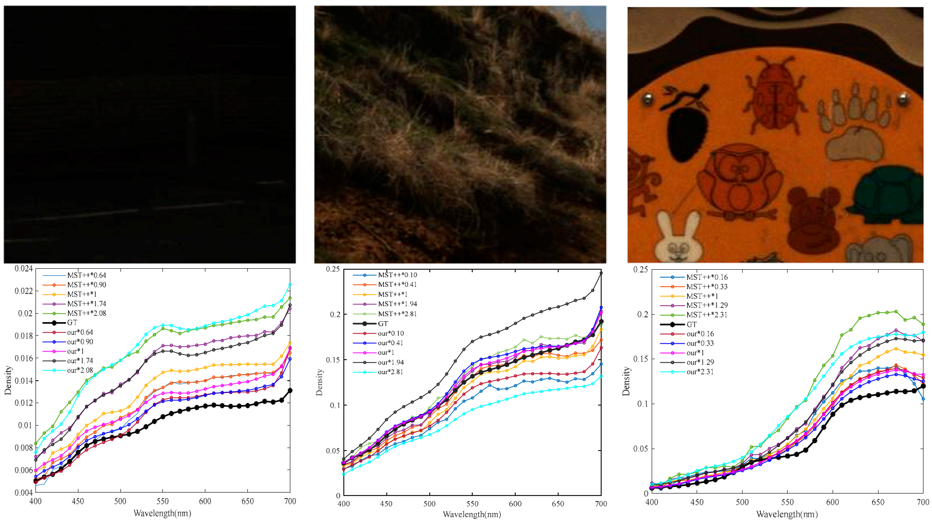

Figure 6. Three more examples of multispectral reconstruction result comparisons between our model and MST++ according to different exposure levels. From left to right: ARAD_1K_0916, ARAD_1K_0923, and ARAD_1K_0945.

Table 4. Comparison of reconstruction results of two models in terms of three exposure conditions: under-exposure (K < 1), normal exposure (K = 1), and over-exposure (K > 1).

Method	Our Model			MST++		
	MRAE	**RMSE**	**PSNR**	**MRAE**	**RMSE**	**PSNR**
$K < 1$	0.233	0.035	31.55	0.239	0.039	31.37
$K = 1$	0.216	0.034	32.06	0.220	0.037	31.53
$K > 1$	0.240	0.042	30.68	0.248	0.048	29.89

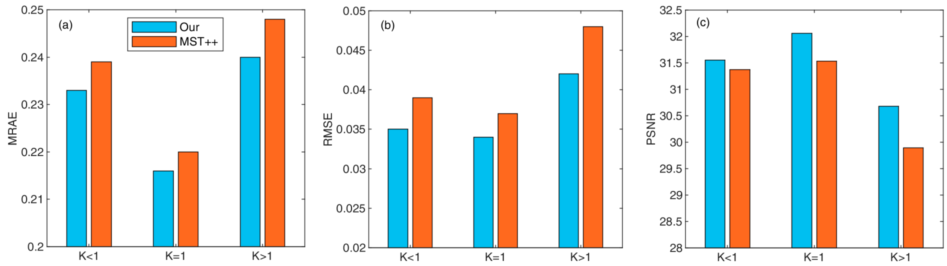

Figure 7. Intuitive error comparison of reconstruction results of two models in terms of three exposure conditions: (**a**) under-exposure (K < 1), (**b**) normal exposure (K = 1), and (**c**) over-exposure (K > 1).

It can be seen from Table 4 and Figure 7 that our model is slightly superior to the MST++ model in all three conditions. In addition, both models achieved the best results when $K = 1$, the second best results when $K < 1$, and the largest error when $K > 1$, which indicates that they are still slightly sensitive to exposure changes.

The statistical results in Table 4 and Figure 7 meet our expectations as the exposure estimation module in our model based on Retinex theory is commonly used for low-light image enhancement, and it is hard to recover the details of serious over-exposure in a digital image when the color information is completely covered by the illumination. Therefore, this issue leads to a somewhat large multispectral reconstruction error for the test situation of $K > 1$. This can also be verified with the reconstructed spectral reflectance curves in Figures 4 and 6, where there are very obvious shifts from the ground truth of the reconstructed spectral curves when $K = 1.79$.

To verify the validity of the model, we performed ablation experiments using MST++ as the baseline model for the validation dataset. The EE module was divided into two versions: EE(DWConv), which uses depth-wise separable convolution; and EE(WTConv), which uses WTConv. The experimental results are shown in Table 5. When the EE(DWConv) module was applied to the baseline model, the MRAE decreased by 0.003, and the PSNR increased by 0.05. Replacing EE(DWConv) with EE(WTConv) further reduced the MRAE by 0.002 and increased the PSNR by 0.11, confirming the effectiveness of the EE module and WTConv.

Table 5. Ablation testing results of the proposed method, with MST++ as baseline.

Ablation	Params (M)	MRAE	RMSE	PSNR
Baseline	1.62	0.236	0.040	30.96
EE(DWConv)	1.77	0.233	0.039	31.01
EE(WTConv)	1.79	0.231	0.041	31.12
EE+EGT	1.79	0.229	0.037	31.43

Afterwards, when we replaced the SST module with the EGT module with added exposure features, the MRAE and RMSE again decreased by 0.002 and 0.04, respectively, and the PSNR increased by 0.31. This further illustrates the effectiveness of using exposure features to guide the computation of self-attention. Additionally, the table shows the changes in the number of parameters for various combinations, indicating that our model only increased by 0.17 M parameters compared to the baseline model, demonstrating that the performance improvement comes at the cost of only a small increase in parameters, which does not impose significant computational pressure in practical applications.

4. Conclusions

In conclusion, we investigated the current deep learning-based advanced multispectral reconstruction models and identified the inherent defects of their sensitivity to exposure changes in input images, which has limited their practical use. To deal with this issue, we decided to use an augmented exposure dataset to enhance the current models' ability to adapt to exposure changes. The results showed that after retraining with the augmented dataset, most of the current advanced multispectral reconstruction models showed improvement in dealing with exposure changes, but they still demonstrated relatively inferior results compared with their original models tested in unchanged exposure conditions. Among them, MST++ showed the best performance. Although an optimized model was proposed by referencing the MST++ model, the multispectral reconstruction accuracy still deviated from the ground truth, especially for the tested images at high exposure levels. In the future, for both deep learning- and machine learning-based multispectral reconstruction studies, more efforts should be devoted to enhancing the model's robustness by utilizing the relationship between local exposure regions or combining deep learning with physical exposure estimation to facilitate its application in practice.

Author Contributions: Conceptualization, J.L. and X.H.; methodology, J.L. and K.X.; investigation, X.H. and W.Z.; validation, X.H. and W.Z.; writing—original draft preparation, X.H. and W.Z.; writing—review and editing, J.L. and Z.W.; visualization, W.Z.; supervision, J.L., K.X. and Z.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (62305255) and Natural Science Foundation of Hubei Province (2022CFB537).

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Li, S.; Xiao, K.; Li, P. Spectra Reconstruction for Human Facial Color from RGB Images via Clusters in 3D Uniform CIELab* and Its Subordinate Color Space. *Sensors* **2023**, *23*, 810. [CrossRef]
2. Mirandilla, J.R.F.; Yamashita, M.; Yoshimura, M.; Paringit, E. Leaf Spectral Analysis for Detection and Differentiation of Three Major Rice Diseases in the Philippines. *Remote Sens.* **2023**, *15*, 3058. [CrossRef]
3. Qamar, F.; Dobler, G. Atmospheric correction of vegetation reflectance with simulation-trained deep learning for ground-based hyperspectral remote sensing. *Plant Methods* **2023**, *19*, 74. [CrossRef] [PubMed]
4. Liang, J.; Wan, X. Optimized method for spectral reflectance reconstruction from camera responses. *Opt. Express* **2017**, *25*, 28273–28287. [CrossRef]
5. Shen, H.L.; Wan, H.J.; Zhang, Z.C. Estimating reflectance from multispectral camera responses based on partial least-squares regression. *J. Electron. Imaging* **2010**, *19*, 020501. [CrossRef]
6. Xu, P.; Xu, H.; Diao, C.; Ye, Z. Self-training-based spectral image reconstruction for art paintings with multispectral imaging. *Appl. Opt.* **2017**, *56*, 8461–8470. [CrossRef] [PubMed]
7. Shi, Z.; Chen, C.; Xiong, Z.; Lin, D.; Wu, F. Hscnn+: Advanced cnn-based hyperspectral recovery from rgb images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 939–947.
8. Li, J.; Wu, C.; Song, R.; Li, Y.; Liu, F. Adaptive weighted attention network with camera spectral sensitivity prior for spectral reconstruction from RGB images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 462–463.

9. Zhao, Y.; Po, L.M.; Yan, Q.; Liu, W.; Liu, T. Hierarchical regression network for spectral reconstruction from RGB images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 422–423.
10. Vaswani, A. *Attention is all you need. Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2017.
11. Cai, Y.; Lin, J.; Lin, Z.; Wang, H.; Zhang, Y.; Pfister, H. Mst++: Multi-stage spectral-wise transformer for efficient spectral reconstruction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–20 June 2022; pp. 745–755.
12. Wang, Y.; Gu, Y.; Nanding, A. SSTU: Swin-Spectral Transformer U-Net for hyperspectral whole slide image reconstruction. *Comput. Med. Imaging Graph.* **2024**, *114*, 102367. [CrossRef]
13. Lin, Y.T. Finlayson G D. Exposure invariance in spectral reconstruction from rgb images. Color and Imaging Conference. *Soc. Imaging Sci. Technol.* **2019**, *27*, 284–289.
14. Lin, Y.T.; Finlayson, G.D. Physically plausible spectral reconstruction from RGB images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 532–533.
15. Liang, J.; Xin, L.; Zuo, Z.; Zhou, J.; Liu, A.; Luo, H.; Hu, X. Research on the deep learning-based exposure invariant spectral reconstruction method. *Front. Neurosci.* **2022**, *16*, 1031546. [CrossRef]
16. Zhang, J.; Sun, Y.; Chen, J.; Yang, D.; Liang, R. Deep-learning-based hyperspectral recovery from a single RGB image. *Opt. Lett.* **2020**, *45*, 5676–5679. [CrossRef]
17. Zhang, J.; Zhao, D.; Chen, J.; Sun, Y.; Yang, D.; Liang, R. Unsupervised learning for hyperspectral recovery based on a single RGB image. *Opt. Lett.* **2021**, *46*, 3977–3980. [CrossRef]
18. Cao, X.; Lian, Y.; Liu, Z.; Li, J.; Wang, K. Unsupervised spectral reconstruction from RGB images under two lighting conditions. *Opt. Lett.* **2024**, *49*, 1993–1996. [CrossRef] [PubMed]
19. Sidhu, R.K.; Sachdeva, J.; Katoch, D. Segmentation of retinal blood vessels by a novel hybrid technique-Principal Component Analysis (PCA) and Contrast Limited Adaptive Histogram Equalization (CLAHE). *Microvasc. Res.* **2023**, *148*, 104477. [CrossRef] [PubMed]
20. Reza, A.M. Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement. *J. VLSI Signal Process. Syst. Signal Image Video Technol.* **2004**, *38*, 35–44. [CrossRef]
21. Yan, Q.; Wang, H.; Ma, Y.; Liu, Y.; Dong, W.; Woźniak, M.; Zhang, Y. Uncertainty estimation in HDR imaging with Bayesian neural networks. *Pattern Recognit.* **2024**, *156*, 110802. [CrossRef]
22. Yan, Q.; Hu, T.; Sun, Y.; Tang, H.; Zhu, Y.; Dong, W.; Van Gool, L.; Zhang, Y. Towards high-quality hdr deghosting with conditional diffusion models. *IEEE Trans. Circuits Syst. Video Technol.* **2023**, *34*, 4011–4026. [CrossRef]
23. Bhalla, K.; Koundal, D.; Bhatia, S.; Imam Rahmani, M.K.; Tahir, M. Fusion of Infrared and Visible Images Using Fuzzy Based Siamese Convolutional Network. *Comput. Mater. Contin.* **2022**, *70*, 5503–5518. [CrossRef]
24. Baek, J.H.; Kim, D.; Choi, S.M.; Lee, H.J.; Kim, H.; Koh, Y.J. Luminance-aware color transform for multiple exposure correction. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 1–6 October 2023; pp. 6156–6165.
25. Yi, X.; Xu, H.; Zhang, H.; Tang, L.; Ma, J. Diff-retinex: Rethinking low-light image enhancement with a generative diffusion model. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 1–6 October2023; pp. 12302–12311.
26. Bai, J.; Yin, Y.; He, Q.; Li, Y.; Zhang, X. Retinexmamba: Retinex-based mamba for low-light image enhancement. *arXiv* **2024**, arXiv:2405.03349 2024.
27. Nguyen, H.; Tran, D.; Nguyen, K.; Nguyen, R. Psenet: Progressive self-enhancement network for unsupervised extreme-light image enhancement. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 2–7 January 2023; pp. 1756–1765.
28. Xu, S.; Chen, X.; Song, B.; Huang, C.; Zhou, J. CNN Injected transformer for image exposure correction. *Neurocomputing* **2024**, *587*, 127688. [CrossRef]
29. He, Q.; Zhang, J.; Chen, W.; Zhang, H.; Wang, Z.; Xu, T. OENet: An overexposure correction network fused with residual block and transformer. *Expert Syst. Appl.* **2024**, *250*, 123709. [CrossRef]
30. Afifi, M.; Derpanis, K.G.; Ommer, B.; Brown, M.S. Learning multi-scale photo exposure correction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 9157–9167.
31. Zhang, Q.; Nie, Y.; Zheng, W.S. Dual illumination estimation for robust exposure correction. *Comput. Graph. Forum* **2019**, *38*, 243–252. [CrossRef]
32. Zhang, Q.; Yuan, G.; Xiao, C.; Zhu, L.; Zheng, W.S. High-quality exposure correction of underexposed photos. In Proceedings of the 26th ACM international conference on Multimedia, Seoul, Republic of Korea, 22–26 October 2018; pp. 582–590.
33. Cai, Y.; Bian, H.; Lin, J.; Wang, H.; Timofte, R.; Zhang, Y. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 1–6 October 2023; pp. 12504–12513.

34. Gavrikov, P.; Keuper, J. Can Biases in ImageNet Models Explain Generalization? In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 16–22 June 2024; pp. 22184–22194.
35. Finder, S.E.; Amoyal, R.; Treister, E.; Freifeld, O. *Wavelet convolutions for large receptive fields. European Conference on Computer Vision*; Springer: Cham, Switzerland, 2025; pp. 363–380.
36. Arad, B.; Timofte, R.; Yahel, R.; Morag, N.; Bernat, A.; Cai, Y.; Lin, J.; Lin, Z.; Wang, H.; Zhang, Y.; et al. Ntire 2022 spectral recovery challenge and data set. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–20 June 2022; pp. 863–881.