

Illumination Spectrum Estimation for Multispectral Images via Surface Reflectance Modeling and Spatial-Spectral Feature Generation

Hyejin Oh¹ Woo-Shik Kim² Sangyoon Lee² YungKyung Park³ Je-Won Kang^{1*}
¹ Dept. of Electronic and Electrical Engineering, Ewha W. University, Seoul, South Korea
² Samsung Advanced Institute of Technology, South Korea
³ Dept. of Color Design, Ewha W. University, Seoul, South Korea

{hyejin5, yungkyung.park, jewonk}@ewha.ac.kr {wooshik.kim, sangy00n.lee}@samsung.com

Abstract

Multispectral (MS) images contain richer spectral information than RGB images due to their increased number of channels and are widely used for various applications. However, achieving accurate estimation in MS images remains challenging, as previous studies have struggled with spectral diversity and the inherent entanglement between the illuminant and surface reflectance spectra. To tackle these challenges, in this paper, we propose a novel Illumination spectrum estimation technique for MS images via Surface reflectance modeling and Spatial-spectral feature generation (ISS). The proposed technique employs a learnable spectral unmixing (SU) block to enhance surface reflectance modeling, which was unattempted in the illumination spectrum estimation, and a feature mixing block to fuse spectral and spatial features of MS images with cross-attention. The features are refined iteratively and processed through a decoder to produce an illumination spectrum estimator. Experimental results demonstrate that the proposed technique achieves state-of-the-art performance in illumination spectrum estimation in various MS image datasets. The code is available at <https://github.com/heyjinnii/ISS-MSI.git>.

1. Introduction

Multispectral (MS) imaging sensors capture a broader range of spectral bands and provide richer information than in RGB images [29, 35]. As these imaging sensors become compact and cost-effective [14], they are actively used in various image processing applications [3, 42, 44, 60, 70]. In computer vision, computational color constancy is used to replicate the perceptual stability of the human visual system, enabling reliable color rendering under varying lighting conditions [7, 34]. In MS imaging, the color constancy

*Je-Won Kang is a corresponding author.

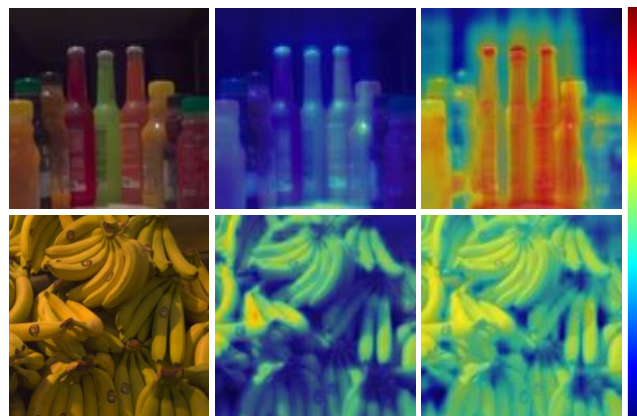


Figure 1. Our model synthesizes spectral features derived from surface reflectance modeling with CNN features, iteratively refining the fused representation in an IRSS framework. The features enhance the accuracy of illumination spectrum estimation by effectively accounting for the reflectance characteristics of objects within the image, as demonstrated in the last column, in contrast to the conventional deep features shown in the middle column.

becomes even more crucial than in the conventional RGB imaging [14, 26, 28, 40, 54, 72], because MS images with the increased number of channels are highly sensitive to illumination changes. The color constancy is achieved by solving an illumination and reflectance spectra separation (IRSS) problem [48], which is an underconstrained problem, requiring the simultaneous recovery of both the illumination spectrum and surface reflectance [72]. Accordingly, achieving accurate estimation in MS images remains a challenge [9, 21, 72].

Recent advances in sensor technology have enabled the acquisition of high-resolution, high-quality MS images [14]. The separation is more challenging in the real-world MS images, because both illumination and reflectance contribute more intricately to the observed spectrum and can vary more dynamically over their channels [30, 31, 71].

Previous studies have attempted to address the IRSS problem by imposing several constraints to MS images. In [2, 10, 72], scene illumination was separated from surface reflectance using low-rank matrix factorizations. However, relying on low-dimensional subspaces of surface reflectance would incur instability when the number of channels is insufficient relative to the basis functions [72]. Despite the limited research to date, deep learning techniques have been applied to MS images for illumination spectrum estimation [14, 40]. However, the previous deep models have largely overlooked the joint analysis of the illumination spectrum and surface reflectance.

To overcome the limitations of existing methods, we propose a comprehensive deep learning-based IRSS framework, dubbed **I**llumination spectrum estimation for MS images via **S**urface reflectance modeling and **S**patial-spectral feature generation (ISS). Spectral unmixing (SU) [25] is employed in this framework to decompose the measured spectrum into a set of endmembers, representing the spectral reflectance characteristics of specific surfaces, and abundances, which indicate the proportion of each endmember. SU has been widely used to analyze MS images in various fields, such as material science [68] and remote sensing [22, 64], yet few studies have been conducted on illumination spectrum estimation. In the proposed technique, rather than deriving the exact surface reflectance of objects, which is practically challenging, we focus on integrating refined reflectance surface modeling into a deep learning-based illumination spectrum estimation approach. Gaussian Mixture Model (GMM) was applied to generate robust endmembers and abundances [46, 53, 74]. We extend this approach by incorporating learnable parameters for endmembers and abundances, providing a robust and flexible decomposition method.

Existing learning-based methods have attempted to build an IRSS framework in their models. However, they may suffer from a limited amount of reflectance data [39] or rely on image prior such as segmentation masks [63] and clusters [15] that are not closely related to illumination spectrum estimation. Their methods were challenging to address diverse illumination spectrum and complicated scenes. In this paper, we propose a feature mixing block that combines the results of endmembers and abundance maps with deep visual features, fully exploiting both spatial information and spectral characteristics of the surface reflectance in the input MS image. Conventional deep learning-based feature estimation methods, commonly employed for illumination estimation, often focus on specific bright areas within a single object and misclassify irrelevant regions as distinct features. As shown in Fig. 1, even in shaded areas where previous methods would struggle, the proposed ISS technique appropriately captures the relevant regions for illumination spectrum estimation, treating both bright and shaded areas

as the parts of the same material. The features generated by the feature mixing block are fed into a decoder to reconstruct a reflectance map and into an illumination estimator to establish the IRSS framework, finally estimating illumination spectrum of an MS image in the output.

The contributions of this paper are as follows:

- We propose a novel deep learning-based IRSS framework to accurately estimate illumination spectrum, addressing the challenges of spectral diversity in MS images.
- We introduce a learnable SU module and a feature mixing block, which combines refined spectral features from surface reflectance modeling with deep visual features using cross-attention, effectively capturing both spatial and spectral characteristics of MS images.
- We evaluate the performance of the proposed technique on various image datasets, demonstrating its superiority to existing studies.

2. Related Works

2.1. Illumination Spectrum Estimation

Color constancy is achieved through two steps, including illumination estimation and chromatic adaptation. Early color constancy studies have focused on illumination estimation from RGB images. Statistics-based methods [8, 33, 62] and deep learning-based methods [1, 3, 4, 20, 36, 43, 56, 67] were developed to improve the accuracy of illumination estimation. These methods only used RGB images to separate illumination and reflectance.

In recent studies, MS images were used to address color constancy by exploiting rich spectral information, which allows for more accurate results. Previous methods extended well-established statistical models and relevant assumptions from RGB to MS spectral domains [13, 27, 61]. In [2, 10, 72], they separated illumination spectrum from surface reflectance using low-rank matrix factorizations. This approach enables the decomposition of the scene with few target objects.

Various learning-based illumination estimation techniques were studied for MS images. In [40], illumination spectrum estimation was modeled as a constrained matrix factorization problem. Spectral reflectance images required per-image calibration using a known reference or measurement device [39]. They used a synthesized training set with simulated lighting conditions. However, the limited number of true reflectance data relative to diverse lighting conditions would lead to performance degradation in complex scenes. In [15], an unsupervised method for segmenting illuminant regions was developed. They assumed the illumination spectrum to be piecewise constant as a probabilistic clustering problem. However, this assumption was relatively simple to address various illumination spectrum. In [63], an image segmentation method was used to lo-

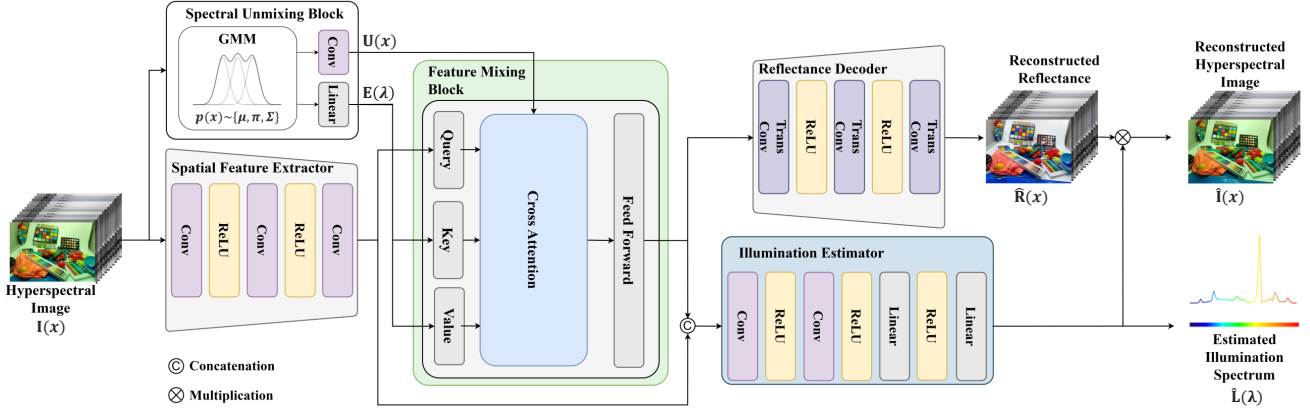


Figure 2. The proposed network uses several sub-networks including a spectral feature extractor, spectral unmixing block, and feature mixing block to generate spatial and spectral features, which are iteratively refined. The decoder uses reflectance decoder and illumination estimator, consisting of an IRSS framework, and estimates illumination spectrum as a result.

calize regions dominantly affected by illumination within the image, which were then used for illumination estimation. However, the gaps between the segmentation results and key regions needed for accurate illumination estimation were problematic. In [14], an input MS image was divided into multiple triplets, processed using a pyramid-based convolutional color constancy block. However, the spectral information of illumination and surface reflectance among different channels was not fully exploited.

2.2. Spectral Unmixing

The SU [25] has been used to extract unique spectrum from various materials in MS images. Previous SU methods used linear [5] and non-linear mixing models [19]. These methods are based on the assumption that each pixel can be expressed as a combination of endmember spectra and corresponding abundances. GMM [23, 46, 74] and deep learning-based SU methods [6, 16, 50, 52, 57, 59, 69] have been studied, extensively. GMM-based models [53] address spectral variability by modeling each endmember as a probabilistic distribution rather than a fixed value, tailored to real-world applications. Deep learning approaches, on the other hand, utilize neural networks to automatically learn the underlying structure of MS data. CNN-based [69] and autoencoder-based models [16, 50, 52, 59] have demonstrated effectiveness in enhancing unmixing accuracy at the cost of computational complexity.

Our method introduces a more refined, learnable SU model based on the GMM, which can model reflectance surface with high spectral variability. The spectral features is combined with a deep model and is further processed, which can adaptively address the spatial and spectral characteristics of input MS images.

2.3. Attention Mechanism in Spectral Imaging

Attention mechanisms of transformers have been widely used to capture spectral characteristics in spectral imaging [11, 18, 37, 38, 55, 73]. Early studies employed spectral and spatial features independently. In [11, 55, 73], the spectral and spatial features are addressed with sequential processing. In [18, 37, 38], the features are processed in parallel using separated branches and combined with late fusion. While these methods were computationally efficient, they failed to effectively fuse the features for illumination spectrum estimation. Recently, cross-attention [51, 65] and 3D convolution [17, 24] were adopted to fuse the spectral and spatial features of MS images. More recently, transformer architectures have been more actively employed for this purpose [32, 41, 45]. Cross-attention mechanisms facilitate the modeling of global dependencies between spectral and spatial features, enabling the effective learning of complex feature relation.

3. Proposed Method

3.1. MS Image Formulation Model

In an IRSS framework [48], an MS image $\mathbf{I} \in \mathbb{R}^{H \times W \times C}$ with height H , width W , and channel C is modeled as

$$\mathbf{I}(x) = \int_{\Lambda} \mathbf{R}(x, \lambda) \cdot \mathbf{L}(\lambda) \cdot \mathbf{S}(\lambda) d\lambda, \quad (1)$$

where $\mathbf{R}(x, \lambda)$ represents surface reflectance of a spatial coordinate x and wavelength λ ranged within Λ . $\mathbf{L}(\lambda)$ and $\mathbf{S}(\lambda) \in \mathbb{R}^C$ represent illumination spectrum and camera sensitivity, respectively.

In the joint IRSS problem, given camera sensitivity, it is necessary to accurately model reflectance surface to produce $\mathbf{L}(\lambda)$. In our approach, the reflectance is modeled as

the product of abundances and endmembers through the SU, representing a detailed spectral characteristics, given as,

$$\mathbf{R}(x, \lambda) = \sum_{k=1}^K \mathbf{U}_k(x) \cdot \mathbf{E}_k(\lambda), \quad (2)$$

where $\mathbf{E}_k(\lambda)$ and $\mathbf{U}_k(x)$ denote the k -th endmember and its abundance map, respectively. $\mathbf{I}(x)$ is modeled by incorporating Eq. (2) as follows:

$$\mathbf{I}(x) = \int_{\Lambda} \left(\sum_{k=1}^K \mathbf{U}_k(x) \cdot \mathbf{E}_k(\lambda) \right) \cdot \mathbf{L}(\lambda) \cdot \mathbf{S}(\lambda) d\lambda, \quad (3)$$

where the SU-based reflectance surface model is used to provide more robust and flexible approach to accurately decompose the components. Fig. 2 presents this process with further details in the following subsections.

3.2. Surface Reflectance Representation with Learnable SU

In the proposed technique, the SU is used to provide detailed spectrum profile of reflectance surface, and the GMM is employed to decompose observed spectrum into a set of endmember spectra and their corresponding abundances. The components are transformed to learnable tensors to accomplish a deep IRSS framework in the following.

Based on the GMM, a probability density of an observed spectrum $\mathbf{I}_x \in \mathbb{R}^C$ at each pixel x is modeled as:

$$p(\mathbf{I}_x) = \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{I}_x | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \quad (4)$$

where K is the number of normal distributions, and $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ are the the mean vector and covariance matrix. π_k is a weighting coefficient, in which $\sum_{k=1}^K \pi_k = 1$.

In Eq. (4), $p(\mathbf{I}_x)$ represents a likelihood that all K spectrum can explain the observed spectrum, quantified by the linear combination of $\mathcal{N}(\mathbf{I}_x | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ and its proportion π_k , and, thus $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ represent the k -th endmember spectrum and the spectral variability within the endmember. To derive the optimal parameter set $\{\pi_k^*, \boldsymbol{\mu}_k^*, \boldsymbol{\Sigma}_k^*\}$, given a set of observed spectra $\mathbf{I}_x \in \mathbf{I}$ of training samples, we apply an Expectation-Maximization (EM) algorithm [12] to iteratively refine the parameters.

Based on the results, we directly obtain the end-member spectra from the mean vectors, *i.e.*, $\boldsymbol{\mu}^* = \{\boldsymbol{\mu}_1^*, \boldsymbol{\mu}_2^*, \dots, \boldsymbol{\mu}_K^*\}$. Then, we calculate a set of abundance $\boldsymbol{\gamma}^* = \{\boldsymbol{\gamma}_1^*, \boldsymbol{\gamma}_2^*, \dots, \boldsymbol{\gamma}_K^*\}$. $\boldsymbol{\gamma}_k^*(x)$ is computed at each pixel coordinate x as follows:

$$\boldsymbol{\gamma}_k^*(x) = \frac{\pi_k^* \mathcal{N}(\mathbf{I}_x | \boldsymbol{\mu}_k^*, \boldsymbol{\Sigma}_k^*)}{\sum_{j=1}^K \pi_j^* \mathcal{N}(\mathbf{I}_x | \boldsymbol{\mu}_j^*, \boldsymbol{\Sigma}_j^*)}, \quad (5)$$

where $\boldsymbol{\gamma}_k^*(x)$ consists of an abundance map for the k -th end-member spectrum.

In previous studies [47, 72], surface reflectance has been approximated using a low-dimensional model based on a subspace assumption, where a small number of spectral bases (e.g., 3) is sufficient for most natural images. These findings could lead to a tight subspace expression of the surface reflectance for practical use cases. In our experiments, we empirically selected K and observed best results with $K = 5$, aligning with the subspace assumption.

In the proposed model, the parameters pass through learnable layers, allowing for a more flexible representation that adapts to real-world images, mathematically expressed,

$$\mathbf{E} = W_e \boldsymbol{\mu}^* \in \mathbb{R}^{K \times C}, \quad \mathbf{U} = W_u \boldsymbol{\gamma}^* \in \mathbb{R}^{K \times H \times W}, \quad (6)$$

where W_e and W_u are convolution layers. \mathbf{E} and \mathbf{U} are the tensors of endmembers and abundance maps, respectively, as the final output of the SU block.

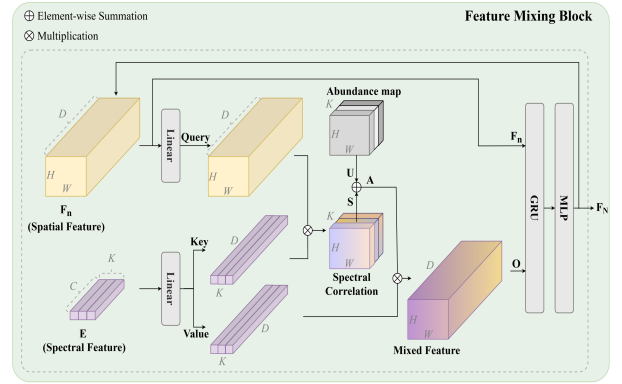


Figure 3. Cross-attention operations in a feature mixing block. It uses input spectral and visual features from a SU block and CNN, respectively, and produces the spatial-spectral features as output, iteratively refined with an N -step GRU.

3.3. Feature Mixing Block with Cross-Attention

Feature mixing block (FMB) uses both the SU features and CNN features as input features. The CNN features are from a spatial feature extractor (SFE) which consists of 3 layers of convolutions, as illustrated in Fig. 2. Previous model-based methods neglected spatial information [2, 10, 72], degrading performance in complex scenes. In contrast, deep learning-based methods captured visual features but struggled to integrate physical modeling [14, 40]. In the proposed method, the SU features and CNN features are used complementarily to generate spatial-spectral features tailored for MS images. Fig. 3 presents the detailed operations in the FMB. The spectral and spatial features are fused using cross-attention and iteratively refined through a gated recurrent unit (GRU).

The FMB employs a cross-attention mechanism of a transformer to fuse spectral and spatial features as shown in Fig. 3. In the process, given $\mathbf{I} \in \mathbb{R}^{H \times W \times C}$, the spatial features derived from CNN are used as queries, and the endmembers from a spectral unmixing block are used as keys and values as follows:

$$\mathbf{Q} = W_q \mathbf{F}_n, \quad \mathbf{K} = W_k \mathbf{E}, \quad \mathbf{V} = W_v \mathbf{E}, \quad (7)$$

where $\mathbf{Q} \in \mathbb{R}^{D \times H \times W}$ is the query, and $\mathbf{K}, \mathbf{V} \in \mathbb{R}^{K \times D}$ are the key and value. \mathbf{F}_n is a visual feature, which is iteratively refined using a n -step GRU. \mathbf{F}_0 is initialized to \mathbf{F}_{cnn} as visual features of a SFE network. D is its spectral dimension. $W_q \in \mathbb{R}^{D \times D}$, and $W_k, W_v \in \mathbb{R}^{D \times C}$ are learnable linear transformations to embed the features to a cross-attention block.

Spectral-wise correlations $\mathbf{S} \in \mathbb{R}^{K \times H \times W}$ between a query of spatial locations in \mathbf{F}_n and a key of spectral characteristics of \mathbf{E} are computed as follows:

$$\mathbf{S} = \mathbf{Q} \cdot \mathbf{K}^T, \quad (8)$$

where the attention scores calculate the relevance between the visual features and spectral features.

These spectral correlations are modified by adding abundance map $\mathbf{U} \in \mathbb{R}^{K \times H \times W}$ obtained from the SU block, which generating an attention weight $\mathbf{A} \in \mathbb{R}^{K \times H \times W}$ as follows:

$$\mathbf{A} = \text{softmax}(\mathbf{S} + \mathbf{U}), \quad (9)$$

where softmax is a softmax function. Then, we compute an attention output through a weighted sum between attention weights and values, given as:

$$\mathbf{O} = \mathbf{A} \cdot \mathbf{V}, \quad (10)$$

where $\mathbf{O} \in \mathbb{R}^{D \times H \times W}$ is the attention output.

In the FMB, the output undergoes an update process using a GRU followed by a two-layer fully connected network (MLP), which iteratively refines the output feature with the query features. Mathematically, this can be expressed as:

$$\mathbf{F}_{n+1} = \text{MLP}(\text{GRU}(\mathbf{F}_n, \mathbf{O})), \quad (11)$$

where \mathbf{F}_n is updated for N iterations to fuse the spectral and spatial features, until the final output \mathbf{F}_N is generated. \mathbf{F}_N represents a refined reflectance feature, maintaining the same dimensions as the input CNN features while incorporating spectral information from the SU method. N is set to 6 in our experiments. The Fig. 4 shows the gradually refined features during the iterative attention process of FMB.

We present a pseudo-code implementation in Algorithm 1.

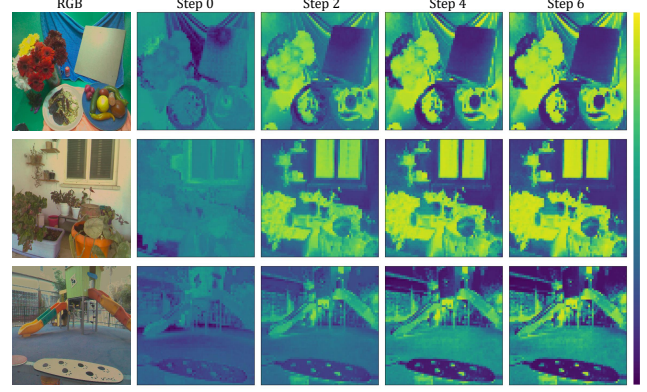


Figure 4. Progressive feature refinement visualization through iterative stages in FMB. The leftmost image shows a rendered RGB of input MS image. From the initial query feature of FMB, the features are gradually refined from step 0 to step 6.

Algorithm 1: SU-CNN Feature Mixing Block

Input: $\mathbf{F}_0 \leftarrow \mathbf{F}_{cnn} \in \mathbb{R}^{D \times H \times W}$: CNN feature
 $\mathbf{E} \in \mathbb{R}^{K \times C}$: Endmember from SU block
 $\mathbf{U} \in \mathbb{R}^{K \times H \times W}$: Abundance map from SU block
Output: $\mathbf{F}_{\text{ref}} \leftarrow \mathbf{F}_N \in \mathbb{R}^{D \times H \times W}$: Updated feature map for N iterations

$\mathbf{K} \leftarrow W_k \mathbf{E}$ // Key projection
 $\mathbf{V} \leftarrow W_v \mathbf{E}$ // Value projection

Function FeatureMixing($\mathbf{F}_0, \mathbf{K}, \mathbf{V}, \mathbf{U}$):

```

for  $n = 0$  to  $N - 1$  do
     $\mathbf{Q} \leftarrow W_q \mathbf{F}_n$  // Query projection
     $\mathbf{S} \leftarrow \mathbf{Q} \cdot \mathbf{K}^T$  // Spectral correlation
     $\mathbf{A} \leftarrow \text{softmax}(\mathbf{S} + \mathbf{U})$  // Attention weights
     $\mathbf{O} \leftarrow \mathbf{A} \cdot \mathbf{V}$  // Attention output
     $\mathbf{F}_{n+1} \leftarrow \text{MLP}(\text{GRU}(\mathbf{F}_n, \mathbf{O}))$  // Feature update with GRU and MLP
end
return  $\mathbf{F}_N$ 
 $\mathbf{F}_N \leftarrow \text{FeatureMixing}(\mathbf{F}_0, \mathbf{K}, \mathbf{V}, \mathbf{U})$ 

```

3.4. Illumination Spectrum Estimation in Decoder Blocks

In decoder blocks, the proposed network generates an estimation of the illumination spectrum, which is denoted by $\hat{\mathbf{L}}$, at the end. The network uses two branches to calculate not only the illumination spectrum estimation but also the surface reflectance approximation, which is denoted by $\hat{\mathbf{R}}$. The approximation of the surface reflectance in the decoder enables the network to use a joint IRSS framework, rather than producing an exact surface reflectance in the absence of GT. As shown in Fig. 2, \mathbf{F}_N is fed into a reflectance decoder, consisting of three transposed convolution layers, to

calculate $\hat{\mathbf{R}}$ as intermediate results in the IRSS. Specifically, we use 3×3 convolution kernels in the implementation. The reflectance decoder uses the only \mathbf{F}_N as input to facilitate the optimization and fully focus on the reflectance prediction.

The illumination spectrum block is built with a series of convolutional and linear layers, in which an input is constructed by concatenating \mathbf{F}_{cnn} and \mathbf{F}_N . The subsequent convolutional layers are employed to extract spatially localized spectral features, effectively capturing illumination in the target scene. We employ two layers of convolution kernel with stride 2. Rectified linear unit (ReLU) activation function is applied after all convolution and linear layers in the block, except for the final output linear layer.

3.5. Loss function

For training the proposed architecture, we employ a reconstruction loss \mathcal{L}_{RC} and an illumination estimation loss \mathcal{L}_{AE} . \mathcal{L}_{RC} is computed using an \mathcal{L}_1 distance between \mathbf{I} and $\hat{\mathbf{I}}$:

$$\mathcal{L}_{RC} = \sum_{x \in (h,w)} \sum_{\lambda \in \Lambda} \|\hat{\mathbf{I}}(x, \lambda) - \mathbf{I}(x, \lambda)\|_1, \quad (12)$$

where the loss is computed over all spatial spectral locations, i.e., $h = 1 \cdots H$ and $w = 1 \cdots W$, and all channels Λ . $\hat{\mathbf{I}}$ is reconstructed using an integral of $\hat{\mathbf{L}}$ and $\hat{\mathbf{R}}$,

We also utilize an angular error (AE), \mathcal{L}_{AE} , defined as:

$$\mathcal{L}_{AE} = \arccos \left(\frac{\hat{\mathbf{L}} \cdot \mathbf{L}}{\|\hat{\mathbf{L}}\| \|\mathbf{L}\|} \right), \quad (13)$$

where \mathbf{L} and $\hat{\mathbf{L}}$ are the ground-truth (GT) and predicted illumination spectrum vector, respectively. For illumination spectrum estimation, capturing relative differences of illuminant among multi-channels is crucial. Spectral orientation defines its color characteristics, which would affect color constancy, while the intensity of illuminant reflects brightness. Thus, we use an angular error to focus on the orientation by measuring the angle between vectors.

The total loss function \mathcal{L} is defined as:

$$\mathcal{L} = \alpha_1 \mathcal{L}_{RC} + \alpha_2 \mathcal{L}_{AE}, \quad (14)$$

where α_1 and α_2 are weighting factors, which are set to 0.5 and 1, respectively, in our experiments.

4. Experimental Results

4.1. Spectral Reflectance Datasets

We evaluate the proposed technique using two spectral reflectance datasets: KAUST [40] and CAVE [66]. In the datasets, we use 31 channels ranging from $400nm$ to $700nm$ at $10nm$ intervals. We relight the scenes by multiplying the surface reflectance of each pixel with an illuminant spectra matrix with 59 standard illuminants from

[49] resulting in a 31-channel spectral image. We randomly select 80% of the KAUST reflectance and illuminant data for training and the remaining 20% used for testing as in [40]. CAVE dataset is utilized only for testing the models. We evaluate the performance of the proposed technique with existing methods including both statistic-based and learning-based methods, GrayEdge [27], LRMF [72], ISNL [58], PWIR [54], DUN [40] and BeyondRGB [14]. We used mean- ΔA_{MS} and std- ΔA_{MS} for evaluation metrics, which are the average and standard deviation of an angular error measured between the GT and predicted illumination spectrum.

Tab. 1 exhibits the performance comparisons, showing that the proposed technique achieves the best results on both datasets. The proposed technique produces 5.55 and 3.29 in terms of mean- ΔA_{MS} and std- ΔA_{MS} , respectively, in KAUST dataset. For CAVE dataset, which is not used for model training, it shows reliable performance achieving 5.93 and 3.90 in mean- ΔA_{MS} and std- ΔA_{MS} , respectively. These results demonstrate that the proposed technique holds stable performance on unseen reflectance data.

Dataset	Method	mean- $\Delta A_{MS} \downarrow$	std- $\Delta A_{MS} \downarrow$
KAUST	GrayEdge [27]	11.46	8.02
	LRMF [72]	24.06	23.49
	ISNL [58]	20.05	11.46
	PWIR [54]	28.07	21.20
	DUN [40]	9.17	6.30
	BeyondRGB [14]	7.30	4.45
	Ours	5.55	3.29
CAVE	GrayEdge [27]	18.33	14.89
	LRMF [72]	20.05	15.47
	ISNL [58]	22.92	16.62
	PWIR [54]	34.95	21.20
	DUN [40]	16.04	13.75
	BeyondRGB [14]	6.64	3.62
	Ours	5.93	3.90

Table 1. Performance comparisons of the proposed technique and existing methods using KAUST and CAVE datasets.

4.2. Beyond RGB Dataset

The BeyondRGB [14] dataset is a comprehensive, real-world dataset, designed for MS image-related studies, focused on mobile devices. The dataset is organized into ‘‘Lab’’ and ‘‘Field,’’ which include a variety of objects with diverse colors and textures in indoor scenes and real-world data in outdoor scenes, both encompassing complex scenes with rich object content. For experimental results, we use the same data-split protocol of [14]. For comparisons, we used PWIR and BeyondRGB, because the other methods required reflectance data for training [40] or the input dimension needs to match the number of spectral bands in GT spectra [27, 58, 72]

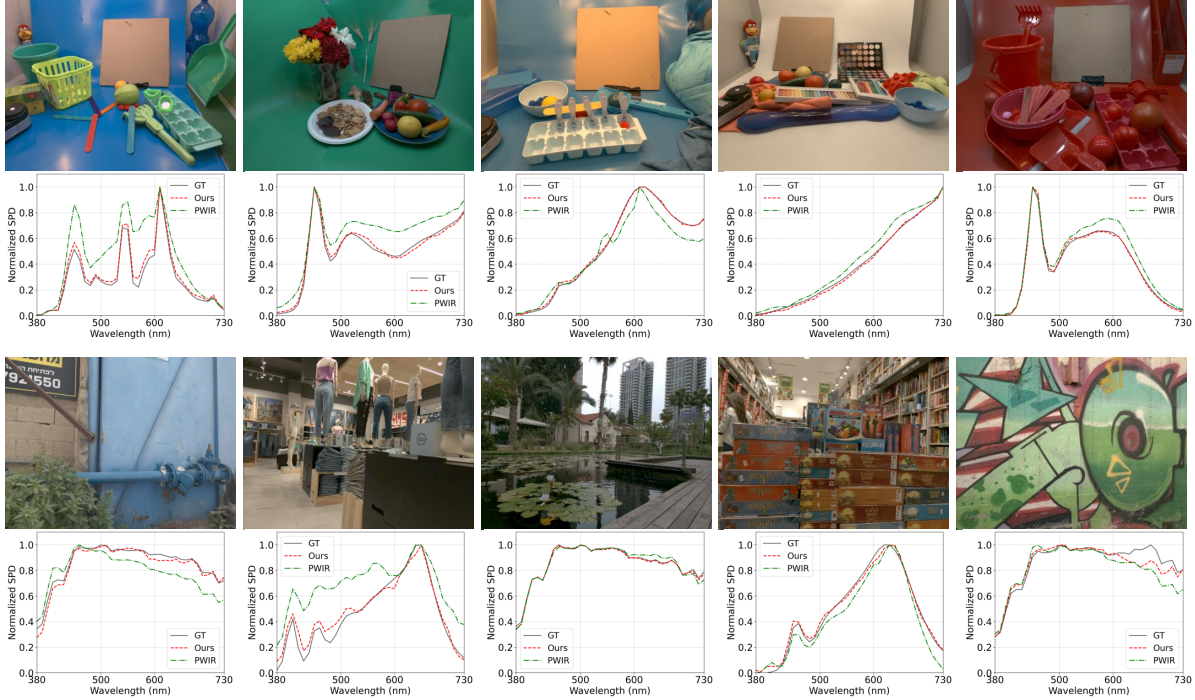


Figure 5. Each of two rows shows MS images of BeyondRGB dataset in the first row and their estimated illuminant spectrum in the second row. The results of our model and PWIR are plotted with red and green dashed lines, respectively. The GT is shown with black solid lines.

Tab. 2 shows the accuracy of the tested methods. It is clearly demonstrated with the results that the proposed technique provides superior accuracy compared to all the tested methods. Our technique significantly reduces the angular errors, achieving values of 3.65 and 5.97 in the Lab and Field datasets, respectively, in mean- ΔA_{MS} . In the BeyondRGB dataset, the proposed technique presents a greater performance gap compared to the second-best algorithm [14] than in KAUST and CAVE datasets, which can present the effectiveness of our approach in complex scenes.

Dataset	Method	mean- $\Delta A_{MS} \downarrow$	std- $\Delta A_{MS} \downarrow$
Lab	PWIR [54]	27.10	10.86
	BeyondRGB [14]	5.92	2.92
	Ours	3.65	2.72
Field	PWIR [54]	16.31	6.07
	BeyondRGB [14]	7.22	5.54
	Ours	5.97	3.90

Table 2. Performance comparisons of the proposed technique and existing methods using real-world MS dataset, BeyondRGB.

Fig. 5 shows the results of illumination spectrum using the proposed technique in comparison with PWIR [54]. The top row of the image consists of Lab images, captured with controlled and artificial illuminations. From left to right, the illuminations are the spectrum of fluorescent, mixed LED, hallogen lamp, incandescent light bulb, and LED illumi-

nant, which are found in daily environments. The bottom row contains real-world Field images, where outdoor images are illuminated by natural sunlight. The estimated illumination spectrum of our model, represented by the red dashed line, demonstrate a significantly more detailed alignment with the shape and peaks of the GT compared to PWIR. In the first image of the Lab set, the proposed technique shows a superior performance in estimating complex illumination conditions, such as those with multiple peaks characteristic of fluorescent lighting. Even for the bottom row comprising a wider variety of scenes, our model demonstrates robust and accurate illumination estimation performance.

4.3. Surface Reflectance Modeling

We investigate the reconstruction results of surface reflectance on the KAUST and CAVE datasets, in which GT reflectance data is available. As shown in Fig. 6, we compare the estimated reflectance from our reflectance decoder and corresponding GT reflectance. From the results, we observe that even though the network has not directly seen GT reflectance during the training process, the surface reflectance is sufficiently considered while estimating illumination from images. Even in complex optical conditions like refractive surfaces with shadows shown in the window and glasses scenes, and reflection and diffraction effects demonstrated by the CD, our method effectively reconstruct the surface reflectance preserving these challeng-

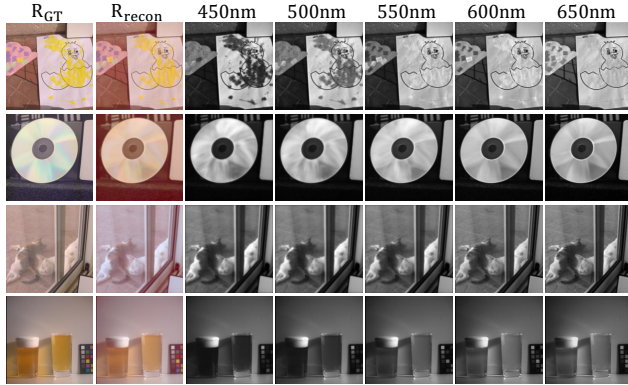


Figure 6. Visualization of reconstructed surface reflectance for KAUST and CAVE datasets. The first and second columns display GT and reconstructed surface reflectance rendered in RGB. The following columns display the estimated surface reflectance at each wavelength, respectively.

ing visual properties. The results align with our motivation for effective disentanglement of illumination and reflectance through physics-based approach and validate the effectiveness of our proposed method.

4.4. Ablation Study

Network architecture: Tab. 3 presents the effectiveness of the proposed modules consisting of the network architecture. When the modules are turned off one-by-one, we observe a considerable reduction in accuracy. First, without SU block, the network relies solely on visual features without using spectral features, resulting in an approximate increase of 1.12 in angular error. Second, when disabling FMB, the spatial and spectral features are not fused appropriately but used independently to estimate the illumination spectrum. It increases approximately 3.65 of an angular error. Lastly, when the SFE module is turned off, we observe a dramatic degradation in accuracy. This result is expected, as the network is unable to exploit any visual features from the input image without the SFE. Fig. 7 shows the output features of the FMB, when disabling SFE or SU. Compared to the rendered RGB image on the left, the synthesized feature can be effectively used for illumination estimation by localizing regions relevant to the task.

SU	FMB	SFE	mean- $\Delta A_{MS} \downarrow$	\mathcal{L}_{AE}	\mathcal{L}_{RC}	mean- $\Delta A_{MS} \downarrow$
✗	✓	✓	5.52			
✓	✗	✓	8.05	✓	✗	4.72
✓	✓	✗	24.64	✗	✓	73.59
✓	✓	✓	4.40	✓	✓	4.40

Table 3. Ablation studies of the sub-modules of the proposed network (left) and training loss functions (right).

Loss function: We conducted ablation tests in the loss terms. In the proposed method, using both \mathcal{L}_{AE} and \mathcal{L}_{RC}

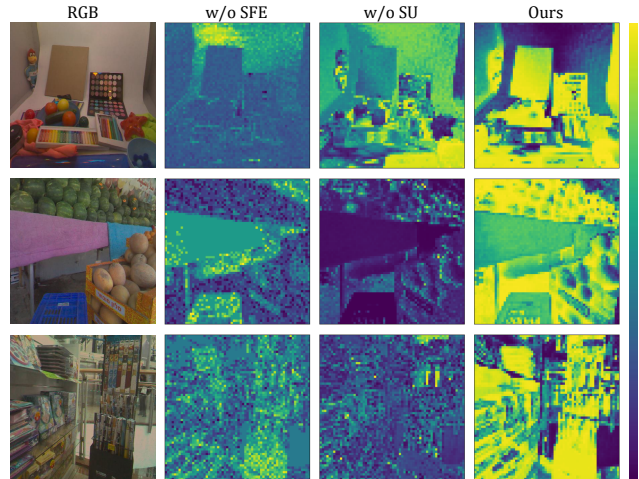


Figure 7. Output features from an FMB block in the ablation test. The second and the third columns display the features, when disabling SFE and SU, respectively. Our features in the last column highlight more relevant regions including high spectral diversity in local surfaces to achieve more enhanced accuracy.

together achieves the best performance. When \mathcal{L}_{RC} is not used, the performance decreases by approximately 0.32 compared to using both the losses. By incorporating \mathcal{L}_{RC} , the network can further learn the impact of a global illumination spectrum on the reconstructed image in terms of spatial representation. When excluding \mathcal{L}_{AE} , we observed substantial performance drops (e.g. 73 in mean- ΔA_{MS}). These results are expected, since relying solely on the reflectance for the image reconstruction may provide insufficient constraints for illumination estimation in the absence of GT. \mathcal{L}_{AE} plays a crucial role in the estimation.

5. Conclusion

In this paper, we proposed illumination spectrum estimation technique for MS images. The proposed ISS technique used a learnable SU block which modeled surface reflectance, and a feature mixing block which fused spatial and spectral information using cross-attention operations. The spatial-spectral features were efficiently used to focus on important regions in complex scenes to enhance the accuracy of illumination estimation. Experimental results demonstrated that the proposed ISS technique provided a superiority to existing studies in various MS image datasets.

In our future study, we will investigate illumination estimation under locally varying and multi-illuminant scenarios. While the current model is designed to estimate a global illumination spectrum, the proposed ISS technique can be extended to improve accuracy in estimating multiple illuminant by efficiently disentangling reflectance properties under locally varying lighting conditions.

Acknowledgment

This work was partly supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (RS-2021-II212068, Artificial Intelligence Innovation Hub, 40%) and was partly supported by the NRF grant funded by MSIT (No.NRF-2022R1A2C4002052, 60%).

References

- [1] Mahmoud Afifi and Michael S Brown. Sensor-independent illumination estimation for dnn models. *arXiv preprint arXiv:1912.06888*, 2019. 2
- [2] Dongsheng An, Jinli Suo, Haoqian Wang, and Qionghai Dai. Illumination estimation from specular highlight in a multi-spectral image. *Optics Express*, 23(13):17008–17023, 2015. 2, 4
- [3] Simone Bianco, Claudio Cusano, and Raimondo Schettini. Color constancy using cnns. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 81–89, 2015. 1, 2
- [4] Simone Bianco, Claudio Cusano, and Raimondo Schettini. Single and multiple illuminant estimation using convolutional neural networks. *IEEE Transactions on Image Processing*, 26(9):4347–4362, 2017. 2
- [5] José M Bioucas-Dias, Antonio Plaza, Nicolas Dobigeon, Mario Parente, Qian Du, Paul Gader, and Jocelyn Chanussot. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE journal of selected topics in applied earth observations and remote sensing*, 5(2):354–379, 2012. 3
- [6] Ricardo Augusto Borsoi, Tales Imbiriba, and José Carlos Moreira Bermudez. Deep generative endmember modeling: An application to unsupervised spectral unmixing. *IEEE Transactions on Computational Imaging*, 6:374–384, 2019. 3
- [7] David H Brainard and Brian A Wandell. Analysis of the retina theory of color vision. *JOSA A*, 3(10):1651–1661, 1986. 1
- [8] Gershon Buchsbaum. A spatial processor model for object colour perception. *Journal of the Franklin institute*, 310(1): 1–26, 1980. 2
- [9] Ayan Chakrabarti, Keigo Hirakawa, and Todd Zickler. Color constancy with spatio-spectral statistics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(8):1509–1519, 2011. 1
- [10] Xiaochuan Chen, Mark S Drew, and Ze-Nian Li. Illumination and reflectance spectra separation of hyperspectral image data under multiple illumination conditions. *Electronic Imaging*, 29:194–199, 2017. 2, 4
- [11] Yushi Chen, Hanlu Jiang, Chunyang Li, Xiuping Jia, and Pedram Ghamisi. Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. *IEEE transactions on geoscience and remote sensing*, 54(10):6232–6251, 2016. 3
- [12] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society: series B (methodological)*, 39(1):1–22, 1977. 4
- [13] Arjan Gijsenij, Theo Gevers, and Joost Van De Weijer. Improving color constancy by photometric edge weighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(5):918–929, 2011. 2
- [14] Ortal Glatt, Yotam Ater, Woo-Shik Kim, Shira Werman, Oded Berby, Yael Zini, Shay Zelinger, Sangyoon Lee, Heejin Choi, and Evgeny Soloveichik. Beyond rgb: A real world dataset for multispectral imaging in mobile devices. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 4344–4354, 2024. 1, 2, 3, 4, 6, 7
- [15] Lin Gu, Cong Phuoc Huynh, and Antonio Robles-Kelly. Segmentation and estimation of spatially varying illumination. *IEEE transactions on image processing*, 23(8):3478–3489, 2014. 2
- [16] Rui Guo, Wei Wang, and Hairong Qi. Hyperspectral image unmixing using autoencoder cascade. In *2015 7th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pages 1–4. IEEE, 2015. 3
- [17] Amina Ben Hamida, Alexandre Benoit, Patrick Lambert, and Chokri Ben Amar. 3-d deep learning approach for remote sensing image classification. *IEEE Transactions on geoscience and remote sensing*, 56(8):4420–4434, 2018. 3
- [18] Siyuan Hao, Wei Wang, Yuanxin Ye, Tingyuan Nie, and Lorenzo Bruzzone. Two-stream deep architecture for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 56(4):2349–2361, 2017. 3
- [19] Rob Heylen, Mario Parente, and Paul Gader. A review of nonlinear hyperspectral unmixing methods. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(6):1844–1868, 2014. 3
- [20] Yuanming Hu, Baoyuan Wang, and Stephen Lin. Fc4: Fully convolutional color constancy with confidence-weighted pooling. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4085–4094, 2017. 2
- [21] Anya Hurlbert. Challenges to color constancy in a contemporary light. *Current Opinion in Behavioral Sciences*, 30: 186–193, 2019. 1
- [22] Marian-Daniel Iordache, José M Bioucas-Dias, and Antonio Plaza. Sparse unmixing of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 49(6):2014–2039, 2011. 2
- [23] Qiwen Jin, Yong Ma, Erting Pan, Fan Fan, Jun Huang, Hao Li, Chenhong Sui, and Xiaoguang Mei. Hyperspectral unmixing with gaussian mixture model and spatial group sparsity. *Remote Sensing*, 11(20):2434, 2019. 3
- [24] Murali Kanthi, T Hitendra Sarma, and C Shobha Bindu. A 3d-deep cnn based feature extraction and hyperspectral image classification. In *2020 IEEE India Geoscience and Remote Sensing Symposium (InGARSS)*, pages 229–232. IEEE, 2020. 3
- [25] Nirmal Keshava and John F Mustard. Spectral unmixing. *IEEE signal processing magazine*, 19(1):44–57, 2002. 2, 3

- [26] Haris Ahmad Khan. *Multispectral constancy for illuminant invariant representation of multispectral images*. PhD thesis, Université Bourgogne Franche-Comté; Norwegian University of Science and . . . , 2018. 1
- [27] Haris Ahmad Khan, Jean-Baptiste Thomas, Jon Yngve Hardeberg, and Olivier Laligant. Illuminant estimation in multispectral imaging. *JOSA A*, 34(7):1085–1098, 2017. 2, 6
- [28] Haris Ahmad Khan, Jean-Baptiste Thomas, and Jon Yngve Hardeberg. Towards highlight based illuminant estimation in multispectral images. In *Image and Signal Processing: 8th International Conference, ICISP 2018, Cherbourg, France, July 2-4, 2018, Proceedings 8*, pages 517–525. Springer, 2018. 1
- [29] Muhammad Jaleed Khan, Hamid Saeed Khan, Adeel Yousaf, Khurram Khurshid, and Asad Abbas. Modern trends in hyperspectral image analysis: A review. *Ieee Access*, 6:14118–14129, 2018. 1
- [30] Dongyoung Kim, Jinwoo Kim, Junsang Yu, and Seon Joo Kim. Attentive illumination decomposition model for multi-illuminant white balancing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25512–25521, 2024. 1
- [31] Furkan Kınlı, Doğa Yılmaz, Barış Özcan, and Furkan Kıraç. Deterministic neural illumination mapping for efficient auto-white balance correction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1139–1147, 2023. 1
- [32] Fanqiang Kong, Tongbo Cao, Yunsong Li, Dan Li, and Kedi Hu. Multi-scale spatial-spectral attention network for multispectral image compression based on variational autoencoder. *Signal Processing*, 198:108589, 2022. 3
- [33] Edwin H Land. The retinex theory of color vision. *Scientific american*, 237(6):108–129, 1977. 2
- [34] Edwin H Land. Recent advances in retinex theory. In *Central and Peripheral Mechanisms of Colour Vision: Proceedings of an International Symposium Held at The Wenner-Gren Center Stockholm, June 14–15, 1984*, pages 5–17. Springer, 1985. 1
- [35] David Landgrebe. Hyperspectral image data analysis. *IEEE Signal processing magazine*, 19(1):17–28, 2002. 1
- [36] Bing Li, Haina Qin, Weihua Xiong, Yangxi Li, Songhe Feng, Weiming Hu, and Stephen Maybank. Ranking-based color constancy with limited training samples. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(10):12304–12320, 2023. 2
- [37] Rui Li, Shunyi Zheng, Chenxi Duan, Yang Yang, and Xiqi Wang. Classification of hyperspectral image based on double-branch dual-attention mechanism network. *Remote Sensing*, 12(3):582, 2020. 3
- [38] Xian Li, Mingli Ding, and Aleksandra Pižurica. Deep feature fusion via two-stream convolutional neural network for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 58(4):2615–2629, 2019. 3
- [39] Yuqi Li, Chong Wang, Jieyu Zhao, and Qingshu Yuan. Efficient spectral reconstruction using a trichromatic camera via sample optimization. *The Visual Computer*, 34:1773–1783, 2018. 2
- [40] Yuqi Li, Qiang Fu, and Wolfgang Heidrich. Multispectral illumination estimation using deep unrolling network. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2672–2681, 2021. 1, 2, 4, 6
- [41] Shuaiqi Liu, Siyuan Liu, Shichong Zhang, Bing Li, Weiming Hu, and Yu-Dong Zhang. Ssau-net: A spectral–spatial attention-based u-net for hyperspectral image fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–16, 2022. 3
- [42] Zhenqi Liu, Xinyu Wang, Yanfei Zhong, Meng Shu, and Chen Sun. Siamhyper: Learning a hyperspectral object tracker from an rgb-based tracker. *IEEE Transactions on Image Processing*, 31:7116–7129, 2022. 1
- [43] Yi-Chen Lo, Chia-Che Chang, Hsuan-Chao Chiu, Yu-Hao Huang, Chia-Ping Chen, Yu-Lin Chang, and Kevin Jou. Clcc: Contrastive learning for color constancy. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8053–8063, 2021. 2
- [44] Zhongyu Lou, Theo Gevers, Ninghang Hu, Marcel P Lucassen, et al. Color constancy by deep learning. In *BMVC*, pages 76–1, 2015. 1
- [45] Qing Ma, Junjun Jiang, Xianming Liu, and Jiayi Ma. Reciprocal transformer for hyperspectral and multispectral image fusion. *Information Fusion*, 104:102148, 2024. 3
- [46] Yong Ma, Qiwen Jin, Xiaoguang Mei, Xiaobing Dai, Fan Fan, Hao Li, and Jun Huang. Hyperspectral unmixing with gaussian mixture model and low-rank representation. *Remote Sensing*, 11(8):911, 2019. 2, 3
- [47] Laurence Maloney. Evaluation of linear models of surface spectral reflectance with small numbers of parameters. In *Physics-based vision: Principles and practice*. Jones and Bartlett, 1992. 4
- [48] Laurence T Maloney and Brian A Wandell. Color constancy: a method for recovering surface spectral reflectance. In *Readings in Computer Vision*, pages 293–297. Elsevier, 1987. 1, 3
- [49] Thomas Mansencal, Michael Mauderer, Michael Parsons, Nick Shaw, Kevin Wheatley, Sean Cooper, Jean D. Vandenberg, Luke Canavan, Katherine Crowson, Ofek Lev, Katrin Leinweber, Shriramana Sharma, Troy James Sobotka, Dominik Moritz, Matt Pppp, Chinmay Rane, Pavithra Eswaramoorthy, John Mertic, Ben Pearlstine, Manuel Leonhardt, Olli Niemitalo, Marek Szymanski, Maximilian Schambach, Sianyi Huang, Mike Wei, Nishant Joywardhan, Omar Wagih, Pawel Redman, Joseph Goldstone, Stephen Hill, Jedediah Smith, Frederic Savoie, Geetansh Saxena, Saransh Chopra, Ilia Sibiriyakov, Tim Gates, Gajendra Pal, Nicolas Tessore, Aurélien Pierre, François-Xavier Thomas, Sabarish Srinivasan, and Tucker Downs. Colour 0.4.2, 2022. 6
- [50] Savas Ozkan, Berk Kaya, and Gozde Bozdagi Akar. Endnet: Sparse autoencoder network for endmember extraction and hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 57(1):482–496, 2018. 3
- [51] Yishu Peng, Yuwen Zhang, Bing Tu, Qianming Li, and Wujing Li. Spatial–spectral transformer with cross-attention for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–15, 2022. 3

- [52] Ying Qu and Hairong Qi. udas: An untied denoising autoencoder with sparsity for spectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 57(3):1698–1712, 2018. 3
- [53] Douglas A Reynolds et al. Gaussian mixture models. *Encyclopedia of biometrics*, 741(659-663), 2009. 2, 3
- [54] Antonio Robles-Kelly and Ran Wei. A convolutional neural network for pixelwise illuminant recovery in colour and spectral images. In *2018 24th International Conference on Pattern Recognition (ICPR)*, pages 109–114. IEEE, 2018. 1, 6, 7
- [55] Swalpa Kumar Roy, Gopal Krishna, Shiv Ram Dubey, and Bidyut B Chaudhuri. Hybridsn: Exploring 3-d–2-d cnn feature hierarchy for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 17(2):277–281, 2019. 3
- [56] Wu Shi, Chen Change Loy, and Xiaoou Tang. Deep specialized network for illuminant estimation. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14*, pages 371–387. Springer, 2016. 2
- [57] Alberto Signoroni, Mattia Savardi, Annalisa Baronio, and Sergio Benini. Deep learning meets hyperspectral image analysis: A multidisciplinary review. *Journal of imaging*, 5(5):52, 2019. 3
- [58] Tong Su, Yu Zhou, Yao Yu, Xun Cao, and Sidan Du. Illumination separation of non-lambertian scenes from a single hyperspectral image. *Optics express*, 26(20):26167–26178, 2018. 6
- [59] Yuanchao Su, Jun Li, Antonio Plaza, Andrea Marinoni, Paolo Gamba, and Somdatta Chakravorty. Daen: Deep autoencoder networks for hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 57(7):4309–4321, 2019. 3
- [60] Hao Sun, Xiangtao Zheng, and Xiaoqiang Lu. A supervised segmentation network for hyperspectral image classification. *IEEE Transactions on Image Processing*, 30:2810–2825, 2021. 1
- [61] Jean-Baptiste Thomas. Illuminant estimation from uncalibrated multispectral images. In *2015 Colour and Visual Computing Symposium (CVCS)*, pages 1–6. IEEE, 2015. 2
- [62] Joost Van De Weijer, Theo Gevers, and Arjan Gijsenij. Edge-based color constancy. *IEEE Transactions on image processing*, 16(9):2207–2214, 2007. 2
- [63] Donik Vršnak, Ilija Domislović, Marko Subašić, and Sven Lončarić. Framework for illumination estimation and segmentation in multi-illuminant scenes. *IEEE access*, 11:2128–2137, 2023. 2
- [64] Xia Xu and Zhenwei Shi. Multi-objective based spectral unmixing for hyperspectral images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 124:54–69, 2017. 2
- [65] Kai Yang, Hao Sun, Chunbo Zou, and Xiaoqiang Lu. Cross-attention spectral–spatial network for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–14, 2021. 3
- [66] Fumihito Yasuma, Tomoo Mitsunaga, Daisuke Iso, and Shree K Nayar. Generalized assorted pixel camera: post-capture control of resolution, dynamic range, and spectrum. *IEEE transactions on image processing*, 19(9):2241–2253, 2010. 6
- [67] Huanglin Yu, Ke Chen, Kaiqi Wang, Yanlin Qian, Zhaoxiang Zhang, and Kui Jia. Cascading convolutional color constancy. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 12725–12732, 2020. 2
- [68] Jinkai Zhang, Benoit Rivard, and Arturo Sanchez-Azofeifa. Derivative spectral unmixing of hyperspectral data applied to mixtures of lichen and rock. *IEEE Transactions on Geoscience and Remote Sensing*, 42(9):1934–1940, 2004. 2
- [69] Xiangrong Zhang, Yujia Sun, Jingyan Zhang, Peng Wu, and Licheng Jiao. Hyperspectral unmixing via deep convolutional neural networks. *IEEE Geoscience and Remote Sensing Letters*, 15(11):1755–1759, 2018. 3
- [70] Chunhui Zhao, Boao Qin, Shou Feng, Wenxiang Zhu, Weiwei Sun, Wei Li, and Xiuping Jia. Hyperspectral image classification with multi-attention transformer and adaptive superpixel segmentation-based active learning. *IEEE Transactions on Image Processing*, 32:3606–3621, 2023. 1
- [71] Runxing Zhao, Zhiwen Wang, Wuyuan Guo, and Canlong Zhang. Multi-scene image enhancement based on multi-channel illumination estimation. *Expert Systems with Applications*, 226:120271, 2023. 1
- [72] Yinqiang Zheng, Imari Sato, and Yoichi Sato. Illumination and reflectance spectra separation of a hyperspectral image meets low-rank matrix factorization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1779–1787, 2015. 1, 2, 4, 6
- [73] Zilong Zhong, Jonathan Li, Zhiming Luo, and Michael Chapman. Spectral–spatial residual network for hyperspectral image classification: A 3-d deep learning framework. *IEEE Transactions on Geoscience and Remote Sensing*, 56(2):847–858, 2017. 3
- [74] Yuan Zhou, Anand Rangarajan, and Paul D Gader. A gaussian mixture model representation of endmember variability in hyperspectral unmixing. *IEEE Transactions on Image Processing*, 27(5):2242–2256, 2018. 2, 3