Jointly Learning Band Selection and Filter Array Design for Hyperspectral Imaging

Ke Li[1] Dengxin Dai[2] Luc Van Gool[1,3]
[1]CVL, ETH Zurich, [2] MPI for Informatics, [3]PSI, KU Leuven
{ke.li,vangool}@vision.ee.ethz.ch, ddai@mpi-inf.mpg.de

Abstract

A single-shot multispectral camera equipped with an optimized color filter array (CFA) has the potential to deliver a fast and low-cost hyperspectral (HS) imaging system. Previous solutions are largely restricted to designing demosaicing algorithms for fixed CFAs – be it the Bayer color pattern or evenly-spaced spectral multiplexing patterns. Since sampling and reconstruction are tightly-coupled, the ability to search for an optimal solution is severely constrained by using predefined CFAs. In this work, we simultaneously address the problem of spectral band selection, CFA design, image demosaicing, and spectral image recovery in a joint learning framework for single-shot HS imaging. We propose a reinforcement learning (RL) based method for spectral band selection and a novel neural network for CFA generation, image demosaicing, and HS image recovery. The final spectral reconstruction accuracy is used to supervise the training of the main network to maximize the synergies between those tightly-related tasks. The RL method regards the main network as an agent to collect reward. Our final method delivers a simple setup – as simple as an RGB camera – for HS imaging. Experimental results show that our method outperforms competing methods by a large margin.

1. Introduction

Hyperspectral (HS) imaging acquires images across many small intervals of the electromagnetic spectrum. It offers great advantages over standard RGB imaging for studying the spectral signatures of a large range of interesting target objects [40, 19, 18, 49, 68], such as body tissues, crops, fruits, seeds, and drugs. Given this advantage, one can easily predict that a marriage of HS images with modern deep neural networks can fully unleash the potential of HS images for many applications [58]. However, this has not happened and probably will not happen soon. The main obstacle in the way is the difficulty of obtaining HS images – acquiring HS images is still much harder than obtaining RGB images. There is still no camera that can record HS images of high spatial resolution at a high frame rate. Cameras for a compromised setting – high spectral but low spatial resolution – are getting common. Still they are expensive.

There are many attempts in the literature to address this issue, including using optimized illumination [53, 12], developing hybrid camera systems [41, 61], exploring digital light processing (DLP) projectors [22], and using random printed masks [67]. These approaches, however, all require additional devices. The mostly relevant works to ours are 1) recovering HS images from RGB images [50, 16, 52, 4] and 2) estimating high-resolution (HR) HS images from the low-resolution (LR) HS images [27, 34, 33]. While most of the research attention for HS image estimation in the computer vision field is paid now to these two streams of research, we find that their settings are sub-optimal and can be improved significantly. For RGB image based methods, the color filters and color filter array (CFA) are both predefined, and more importantly they were not designed for the purpose of HS image recovery. The LR HS image based methods also assume a predefined CFA pattern – the filters of all the narrow spectral bands are evenly distributed over the sensor units on the 2D image plane. We argue that filter band selection, CFA design, demosaicing and HS image recovery should be jointly learned. Those tasks are tightly-coupled and thus treating them separately severely constrains the capability of reaching the optimal solution.

In this work, we propose a novel approach to jointly learn all these relevant tasks. Specifically, we look back to image sampling and band optimization and optimize them together with HS image recovery. First, instead of using three fixed wide band filters or a fixed set of narrow band filters, our method learns to select filters from a large set of wide band filters. This setting has two advantages: 1) using the right filters and using the right number of them can let us find the optimal balance between spatial and spectral resolution in order to maximize the performance of HS image recovery; 2) compared to narrow band filters, wide band filters have higher light transmittance efficiency, meaning less imaging

noise and higher frame rate (less exposure time). We design a reinforcement learning (RL) method for this task.

Second, we develop a novel network that learns to optimize the CFA pattern together with image demosaicing and HS image recovery. There are three sub-networks, one for each task, and they are trained together in an end-to-end fashion. The CFA pattern generation network generates a CFA by taking as inputs the positional encoding of the CFA grid and the results of band selection by our RL method. The CFA is then used to obtain image measurements. Since our CFA is the result of an optimization method, the measurements of different bands can have different densities and different distribution patterns. This raises great challenges for image demosaiscing. We propose a novel solution to this by leveraging the power of sparse convolutions and local implicit image functions. This novel image demosaicing network offers a high level of flexibility – it can handle many bands, different measurement densities across bands, and different measurement patterns (even and uneven) across bands. To our best knowledge, this is the first demosaicing method that offers this level of flexibility.

Finally, we use a spatial-spectral prior network to convert the demosaiced multi-spectral (MS) images to the final HS images and use the image recovery loss to guide the training of these three (sub-)networks.

Given a set of filters and the CFA dimensions, our approach automatically finds the useful filters, determines their appearance frequency in the CFA, optimizes the CFA pattern, and minimizes the HS-reconstruction error with the corresponding image measurements. Our method produces high-quality HS reconstructions, outperforming previous methods by a large margin.

2. Related Work

Band Selection. Multispectral filter arrays (MSFA) has aroused great interests in academia and industry in past years, due to the simpler design, lower cost, higher portability and higher accuracy. As a result, there have been quite some research about its design [25] [38] [31] [23] [55] [45] [43][44]. There have been works that select spectral bands to increase performance of the final task [30, 21, 54, 57, 3, 17]. The selection can be done by using techniques such as mutual information between bands [21] or by visually checking the results [26]. Considering that deep learning is now a powerful method and has shown potential for spectral band selection [48, 46, 17], we decide to employ it to tackle our band selection problem.

CFA Design. Following the work of Bayer, a variety of new CFA design strategies have been proposed over the years [25, 39, 9]. The closest work to ours is the method proposed by Chakrabarti [8] which uses a CNN architecture to design a CFA from four predefined colors while training a demosaicing method jointly. While the spirit is similar, our work differs significantly. First, we consider a large number of bands. This increases the difficulty of CFA optimization and demosaicing. Therefore, novel algorithms have to be designed. Second, we address the task of HS image recovery with sparse MS measurements, which is harder.

The advance in CFA design for RGB cameras sparks great interests in the CFA design for MS cameras [55, 45, 43, 44, 1, 24, 38, 47, 15, 6, 63, 56]. An early work by Ramanath et al. presented a CFA pattern which is composed of seven bands and they are arranged hexagonally [55]. The first generic method for CFA design was developed by Miao et al. in which a binary tree and a checkerboard pattern were employed to arrange band filters [45][44]. This work extensively discussed the requirements for MSFA designs and carefully addressed them in their approach. However, they have not considered band selection and joint training of MSFA design and image demosaicing.

Many previous works manually determine the number of bands and arrange the filters in a very straightforward manner. For example, [7] presented a MSFA with 6 bands in 400-700 nm range arranged in 3×2 moxels. [1] evaluated 4 possible patterns for a 4-band filter array. [38] proposed a MSFA with 16 bandpass filters arranged in 4×4 moxels of which 15 are for visible and 1 for near-infrared. We refer the readers to this excellent paper by Lapray et al. [31] for a more comprehensive survey.

Hyperspectral Image Super-resolution (SR). There are three settings for HSI SR: 1) HS image SR from only RGBs [50, 16, 52, 4]; 2) HS image SR from LR HS images [28, 34, 33]; 3) HS image SR from LR HS images and HR RGB images [64, 27]. While these groups of methods constantly use the most recent learning methods, they largely ignore the image sampling problem. In other words, band selection and the spatial resolution of those bands have not been studied or optimized, even though they can play a significant role in HS image SR. HS image SR from RGB images offers the simplest setup. However, since commercial RGB cameras are tuned to mimic human trichromatic perception, their spectral response functions are not optimal for HS image reconstruction. As such, Nie et al. has verified the advantage of deeply learned filters over RGB cameras for HS imaging [51]. Sun et al. [59] has learned a IR-Cut Filter to be placed in front of the lens of RGB cameras to better capture spectral signals. HS image SR from LR HS images or fusion-based methods using both HR RGB images and LR HS images have gained quite a lot of research attention with notable works such as 3D convolutional network [42][34], grouped convolutions with shared parameters [36][27][33] and fusion net [64].

Demosaicing. Demosaicing has been a well-established field with many great works proposed [23, 29, 35, 47, 15, 6, 5]. The general idea is to use the measured signals at sparse locations to fill up the missing values in neighboring

locations. This can also be done by leveraging the dependencies across spectral bands. The measured signals are usually assumed to be evenly distributed on a 2D grid. Our novel demosaicing method can handle a large number of bands, arbitrary distribution patterns, and different densities of measurements across different bands.

3. Method

Although there is considerable literature in the development of demosaicing and super-resolution algorithms, to the best of our knowledge, significantly less [43] has been done for spectral (color) band selection and for CFA design/optimization. Moreover, almost all of the discussions are for the sake of HS image reconstruction. In this work, we jointly learn the spatial pattern for multiple color filters – that requires making a hard decision to use one of a discrete set of color filters at each pixel – along with a neural network that performs demosaicing and spectral recovery. Together, these enable the recovery of high-quality HS images. The pipeline of our main network is shown in Fig. 1.

3.1. Problem Definition

We formulate this task as that of reconstructing an HS image Y, $\mathrm{Y}(n) \in \mathbb{R}^K$ from a measured sensor image X, $\mathrm{X}(n) \in \mathbb{R}$, where $n = (u, v) \in \mathbb{Z}^2$ indexes pixel location and K is the total number of spectral bands. Along with this HS reconstruction task, we also need to learn a pattern for spectral color filters which determines the spectral color channel that each $\mathrm{X}(n)$ corresponds to. The spectral color channel is implemented by putting a color filter over the pixel sensor. For each pixel, its filter is selected out of a fixed set of C filters. In this work, we choose to use popular wide band filters such as *red*, *yellow* and *cyan*, instead of narrow band filters (10 nm - 40 nm wide usually). This choice is made because 1) wide band filters can be physically created easily at low cost and they are already widely available; and 2) wide filters have better light transmittance efficiency, which means less imaging noise and higher frame rate.

We use I, $\mathrm{I}(n)$ in $\mathbb{R}^C$ to denote the intensity measurements corresponding to each of these color channels, and a binary selection map $\mathcal{M}$, $\mathcal{M}(n) \in \{0, 1\}^C$ with $|\mathcal{M}(n)| = 1$ to encode the color (spectral) filter array (CFA) patterns. The corresponding sensor measurements are then given by $\mathrm{X}(n) = \mathrm{M}(n)^T \mathrm{I}(n)$. In order to make the filter array design intuitive and feasible, we follow existing literature (e.g. the Bayer pattern) and assume that M repeats periodically every m pixels, and therefore $\bar{\mathrm{M}} \in \{0, 1\}^{m \times m \times C}$. As an example, Bayer pattern has $m = 2$ and $C = 3$ for RGB image recovery.

Given a training set consisting of C-channel input MS images I and the corresponding K-channel output HS images Y, the goal is to learn the CFA pattern M jointly with a reconstruction algorithm that maps the corresponding measurements X to the full HS image Y. First, we propose a reinforcement learning based method $\mathcal{G}_b$ for band selection, i.e. to learn the number of appearance $\mathbf{h}(c)$, $c \in \{1, ..., C\}$ for each of the C color filters in $\bar{\mathrm{M}}$, where $\mathbf{h}(c) \in \{0, 1, ..., m^2\}$ and $|\mathbf{h}| = m^2$. Furthermore, we develop a network $\mathcal{G}_s$ to generate $\bar{\mathrm{M}}$, which uses $\mathbf{h}$ as guidance. Once having $\bar{\mathrm{M}}$, we then map the input I to measurements X. The learnable parameters of $\mathcal{G}_s$ encode the learned CFA pattern $\bar{\mathrm{M}}$. Third, we design a novel demosaicing network $\mathcal{G}_d$ that outputs demosaiced images for all measured MS channels. Those demosaiced MS images are then feed into a spectral recovery network $\mathcal{G}_r$ to recover the full HS images Y. Please refer to Fig.1 for the visual representation of those elements. We train $\mathcal{G}_s$, $\mathcal{G}_d$, and $\mathcal{G}_r$ all together, with respective to an HS reconstruction loss and a loss imposed on M to respect the band selection result $\mathbf{h}$. The band selection network $\mathcal{G}_b$ is trained by treating the other three networks as its agent to compute reward (the HS image recovery accuracy) in order to take actions (modifications to $\mathbf{h}$).

3.2. CFA Pattern Generation

The CFA pattern generation network $\mathcal{G}_s$ is a small convolutional neural network (CNN) that takes as input the spatial information of each pixel in a window of size $m \times m$ pixels. The output of the network is the zero-one selection mask $\bar{\mathrm{M}}$. The spatial information is obtained by using the 2D positional embedding method that has been used in vision transformers [14]. Specifically, two sets of embeddings are learned, each for one of the axes. This leads to X-embedding and Y-embedding. We concatenate the two embeddings to get the final positional embedding for a pixel.

The key challenge lies in generating the optimal CFA pattern mask $\bar{\mathrm{M}}$ as it requires to learn a hard non-differentiable decision between C^{m^2} possibilities. To address this, we adopt the method proposed by Chakrabarti [8] that adds a temperature parameter τ to the soft-max function:

$$\bar{\mathrm{M}}(n) = \texttt{Soft-max}\ [\tau_t \mathbf{f}(n)], \tag{1}$$

where t is the training iteration and $\mathbf{f}$ is the feature input to the soft-max function. The temperature parameter τ increases with the training iteration. Therefore, the distribution of $\bar{\mathrm{M}}(n)$ can be effectively pushed to zeros and ones because of the increasing τ_t. This special design makes sure that the CFA pattern can be updated via SGD while also shifting towards making a hard choice. We use the quadratic schedule proposed in [8] to increase τ_t.

Periodic padding is used here as $\bar{\mathrm{M}}$ is used periodically to generate M. The network $\mathcal{G}_s$ is trained with two losses - one based on the spectral reconstruction quality through $\mathcal{G}_r$ (Sec. 3.3) and the other is derived from the total number of appearance for each band in $\bar{\mathrm{M}}$ as indicated by $\mathbf{h}(c)$ (Sec.

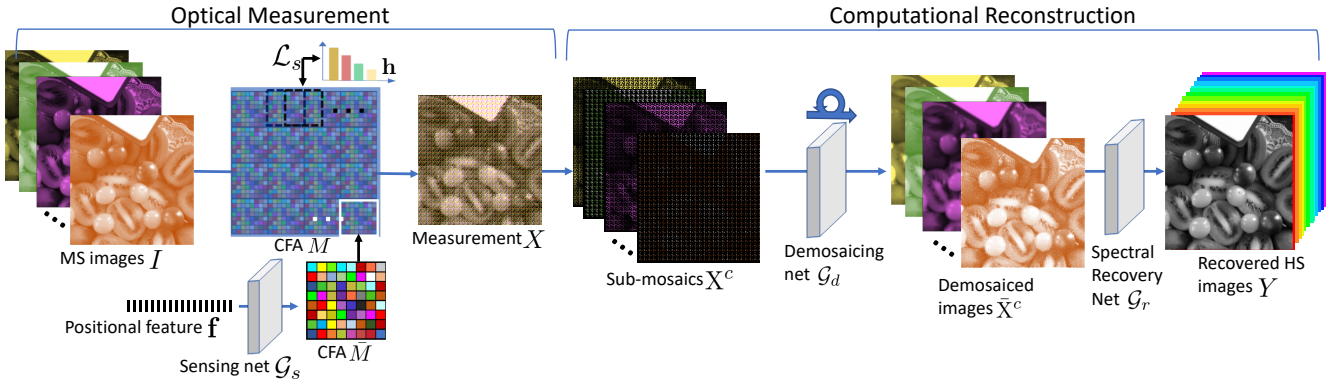


Figure 1: The overview of our method: left is optical measurement of the monochromatic image mosaic X by a learned CFA M, right shows reconstruction by a MS demosacing network $\mathcal{G}_d$ and a spectral recovery network $\mathcal{G}_r$.

3.4). For the band selection loss, we have two requirements: 1) the frequency of selected bands by mask $\bar{\mathrm{M}}$ should be consistent with the required band frequency $\mathbf{h}$; 2) repeating $\bar{\mathrm{M}}$ over the whole image should lead to a sampling strategy that all bands are sampled as evenly as possible over the entire image. The second constraint is known as *Spatial Uniformity* in the literature [21].

We therefore define the CFA pattern loss on the overall mask M instead of the pattern mask $\bar{\mathrm{M}}$. Specifically, we densely sample patches of size $m \times m$ (the same size as $\bar{\mathrm{M}}$) at a stride of $\lfloor m/2 \rfloor$. This will leads to a total number of P patches $\ddot{\mathrm{M}}$. The CFA pattern loss is then defined as

$$\mathcal{L}_s = \frac{1}{P} \sum_{p=1}^{P} \| \sum_{n=1}^{m^2} \ddot{\mathrm{M}}(n) - \mathbf{h} \|^2. \tag{2}$$

Since the patches $\ddot{\mathrm{M}}$ are densely sampled at a stride smaller than m, so some of patches reside over multiple neighboring CFA masks but we anyway still force them to be consistent with $\mathbf{h}$, such that this single loss fulfills the two requirements at the same time.

During training, we generate the corresponding $\mathrm{X}(n)$ vectors using Eq. 1 above, and the layer then outputs sensor measurements based on the C-channel input $\mathrm{I}(n)$ as $\mathrm{X}(n) = \mathrm{M}(n)^T \mathrm{I}(n)$. Once the training is complete, we replace $\mathrm{M}(n)$ with its zero-one version as $\mathrm{M}(n)^c = 1$ for $c = \texttt{argmax}_c \mathbf{f}^c(n)$, and 0 otherwise.

3.3. Demosaicing and Spectral Recovery

Given the sensor measurement image, i.e. the monochromatic image mosaic X, we need to performs two tasks: demosaicing to fill in the missing values for each of the C wide color channels and to convert the densified C-channel MS image to the desired K-channel HS image. Therefore, we decompose the task into a spatial reconstruction sub-task and a spectral reconstruction sub-task, and design corresponding networks for them.

3.3.1 Sparse Implicit Demosaicing

Recent deep learning based methods apply CNNs to the image mosaics [23] to recover the dense images. These existing methods have two problems: the image mosaics are sparse, which makes standard convolutions a sub-optimal choice as spatial dependencies will include spurious information from these uninformative areas, and computational power is wasted on uninformative areas. This issue is especially severe when there are many color bands in the monochromatic image mosaic – the submosaic for each band is sparse. Another problem is that the shared filters in standard convolutions have a fixed size. This is problematic when different submosaics have very different sparsity levels. This is actually the case when the number of input bands is large and when the mosaic pattern is generated by an optimization algorithm as it is in our case – some bands are less important and thus having sparser appearance.

To address all these issues, we propose a novel Sparse Implicit Demosaicing network built on top of sparse convolutions, implicit image function, and grouped networks. Specifically, we use the highly efficient Minkowski convolutions [13] and an extension of the implicit image function developed for image super-resolution task [11]. Below we first present our method for a single channel and then its extension for multiple input channels.

Sparse feature encoding. In order to process the monochromatic image mosaic X, we first need to lift it into C sub-mosaics X^c:

$$\mathrm{X}^c(n) = \begin{cases} \mathrm{X}(n) & \text{if } \mathrm{M}^c(n) = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

For sparse convolutions, a sparse tensor T is represented as a *coordinate* matrix N and a *feature* matrix Z:

$$\mathrm{N} = \begin{bmatrix} u_1 & v_1 \\ \vdots & \vdots \\ u_J & v_J \end{bmatrix}, \quad \mathrm{Z} = \begin{bmatrix} \mathbf{z}_1 \\ \vdots \\ \mathbf{z}_J \end{bmatrix}, \tag{4}$$

where $n_j = (u_j, v_j)$ are pixel coordinates, and $\mathbf{z}_j \in \mathbb{R}^Q$ is the corresponding feature vector. Note that we use n and (u, v) interchangeably for pixel coordinates. As input, the sub-mosaic X^c of size $U \times V \times 1$ is *sparsified* by gathering the positions of its valid pixels as coordinates and the intensity values as features. Once the input sub-mosaic image X^c is sparsified, its information is encoded through a series of Sparse Residual Blocks (SRB). Following the design by Guizilini et al. [20], each SRB is composed of three parallel branches, each with a different number of sparse convolutional blocks. However, we remove the max pooling layers to keep the same resolution at all feature levels. The outputs of the three branches are added together to form the input for the next SRB. We use four SRBs in total. The output features of the last SRB are mapped back to the 2D image plane:

$$\mathrm{Z}^c(u, v) = \begin{cases} \mathbf{z}(u, v) & \text{if } \mathrm{M}^c(u, v) = 1, \\ \mathbf{0} & \text{otherwise.} \end{cases} \tag{5}$$

Dense image decoding. We then use a local image implicit function to decode the feature map Z^c to obtain a dense image $\bar{\mathrm{X}}^c$. Following [11], we parameterize the decoding function f_θ as an MLP that takes the form:

$$\bar{\mathrm{X}}^c(n) = f_\theta(\mathbf{z}^c(n'), n - n'), \tag{6}$$

where $\mathbf{z}^c(n')$ is the nearest latent code from n in Z^c.

The idea is that a continuous image is represented as a 2D feature map $\mathrm{Z}^c \in \mathbb{R}^{U \times V \times D}$, where D is the feature dimension. This can be viewed as $\sum_{u,v} \mathrm{M^c(u, v)}$ latent codes 'sparsely' spread in the 2D domain, as indicated by the location of ones in M^c. This function f_θ is shared by all the images.

As pointed out by [11], a direct use of Eq.6 can lead to discontinuous predictions for the 'border' pixels where the selection of the nearest latent code $\mathbf{z}^c(n')$ switches. We follow the general idea of [11] and address this by using a local ensemble so that Eq.6 is extended to

$$\bar{\mathrm{X}}^c(n) = \frac{\sum_{t=1,2,3,4} \frac{1}{\|n - n'_t\|^2} \cdot f_\theta(\mathbf{z}^c(n'_t), n - n'_t)}{\sum_{t=1,2,3,4} \frac{1}{\|n - n'_t\|^2}}, \tag{7}$$

where $\mathbf{z}^c(n'_t)$ ($t \in \{1, 2, 3, 4\}$) are the 4 nearest latent codes for query location n.

Grouped demosaicing. By now, we have a demosaicing method for a single image channel. Another challenge facing us before delivering a flexible demosaicing algorithms is to handle a large number of color bands. To address this, we use the idea of grouped convolutions with group size 1, i.e. each band is a single group. This means that we share the same sparse feature encoding network and the same dense image decoding network across all bands. We generate the densified images $\bar{\mathrm{X}}^c, c = [0, 1, ..., C]$ for all bands, and then concatenate them to form the final C-channel demosaiced image $\bar{\mathrm{X}} \in \mathbb{R}^{U \times V \times C}$. We now have our complete demosaicing method for our MS image measurements.

Remarks. Thanks to the great properties of our sparse encoding and dense image decoding via implicit image function, our method is able to handle a large number of color channels even when the density of their measurements is different, the mosaic patterns of different bands are different, and the measurements are not evenly spaced on the 2D image domain. To our best knowledge, this is the first deep network approach that can offer this level of flexibility. Note that the local implicit function work for RGB image super-resolution [11] assumes that the measurements are evenly spaced. Consequently, the method can use standard CNNs directly for feature extraction. Furthermore, it only handle RGB images which have much fewer bands. Therefore, while our work is built on this excellent work, our contributions are significant.

3.3.2 Spectral Recovery

Given the demosaic results $\bar{\mathrm{X}} \in \mathbb{R}^{U \times V \times C}$, the goal in this section is to convert this C-band MS image to the desired K-band HS image Y. For this, we employ the spatial-spectral prior network [28]. We use this network to learn the spectral transformation as $\bar{\mathrm{X}}$ and Y have the same spatial resolution.

In order to capture both spatial and spectral correlation of the recovered HS images, we follow [28] and combine the L1 loss and the spatial-spectral total variation (SSTV) loss [2]. SSTV is used to encourage smooth results in both spatial domain and spectral domain and it is defined as:

$$\mathcal{L}_{\text{SSTV}} = \frac{1}{N} \sum_{n=1}^{N} (||\nabla_{\mathrm{h}} \hat{\mathrm{Y}}^n||_1 + ||\nabla_{\mathrm{w}} \hat{\mathrm{Y}}^n||_1 + ||\nabla_{\mathrm{c}} \hat{\mathrm{Y}}^n||_1), \tag{8}$$

where ∇_{h}, ∇_{w}, and ∇_{c} compute gradient along the horizontal, vertical and spectral directions, resp. The reconstruction loss is:

$$\mathcal{L}_r = \mathcal{L}_1 + \mathcal{L}_{\text{SSTV}}. \tag{9}$$

Note that all the described sub-networks including CFA pattern generation, grouped demosaicing and spectral recovery are parts of the same network and they can be trained together in an end-to-end manner. Therefore, after including the CFA pattern loss in Eq. 2, the overall loss is

$$\mathcal{L} = \mathcal{L}_r + \lambda \mathcal{L}_s. \tag{10}$$

We use $\lambda = 10$ as $\mathcal{L}_s$ is generally much smaller than $\mathcal{L}_r$.

3.4. RL-based Band Selection

As discussed in Sec. 3.1, we need to determine the number of pixels that each of the C pre-defined bands should

have for the CFA of size m^2. This appearance histogram is indicated by $\mathbf{h}$. In this section, we propose a reinforcement learning (RL) based method G_b for this task.

Solution Space. $\mathbf{h}$ is a valid proposal if it satisfies this constraint: $\mathbf{h}(c) \in \{0, 1, ..., m^2\}$ and $|\mathbf{h}| = m^2$.

Action. We define actions as modifications to $\mathbf{h}$: $\mathbf{h}^{t+1}(c) = \mathbf{h}^t(c) + \dot{\mathbf{h}}(c)$, where $\dot{\mathbf{h}}(c)$ is a number randomly sampled from $\{-1, 0, 1\}$ and t indicates the optimization step of the RL method. We truncate $\mathbf{h}^{t+1}(c)$ to $[0, m^2]$ after having the modification. If the total number of filters is not m^2, we need to remove or add filters until reaching m^2. If having less, we add one by one a randomly selected band. Otherwise, we remove one by one a randomly-selected filter that have non-zero filters.

Value Function. The final spectral reconstruction performance ρ by the network, i.e. the combination of $\mathcal{G}_s$, $\mathcal{G}_d$ and $\mathcal{G}_r$, is the reward that our RL method maximizes. We train a small neural network $\mathcal{G}_v$ to approximate this value function by training on all collected training pairs $(\mathbf{h}, \rho)$ collected over time.

Epsilon-Greedy Algorithm. We use a simple Epsilon-Greedy Algorithm for the search. Epsilon-Greedy is a simple method to balance exploration and exploitation by choosing between exploration and exploitation randomly. With a small probability ϵ, we propose a completely random $\mathbf{h}^{t+1}$. Otherwise, we use the $\mathbf{h}^{t+1}$ produced by performing the action that is considered the best by our value function $\mathcal{G}_v^t$ at each step t. Note that the total number of valid actions is very large, so we use the value function to select the best action out of 30 proposed actions in each iteration. Every time a neural network is trained with a new $\mathbf{h}^{t+1}$ and gets evaluated, we have one more training sample for $\mathcal{G}_v$, which is then retrained with the new training set for better value function approximation. We train our RL method for a sequence of length T.

The network architectures and spectral sensitivity curves of the filters are shown in the supplementary material.

4. Experiments

4.1. Experimental Setup

Datasets. Two public datasets, CAVE dataset[65] and Harvard dataset[10], are used to evaluate our method. The CAVE dataset includes 32 images of 512 $\times$ 512 pixels. Those images have 31 bands ranging from 400 to 700 nm at a step of 10 nm. They are splitted into two parts: 20 images for training and 12 for testing. As for the Harvard dataset, there are 50 images of 1392 $\times$ 1040 pixels in total. The images contain 31 bands but range from 420 nm to 720 nm. We use 40 images for training and 10 for testing. For both datasets, the training patch size is set to 128 $\times$ 128 pixels at a stride of 64 pixels. For CAVE, we divide the data values by 65536 to map them to $[0, 1]$. For Harvard, we multiple all data values by 20 to roughly map them to the same range.

Evaluation Metrics Six standard metrics are employed to evaluate the performance of all methods. They are cross correlation (CC)[37], spectral sample mapper (SAM)[66], root mean square error (RMSE), erreur relative globale adimensionnelle de synthese (ERGAS)[60], peak signal-to-noise ratio (PSNR), and structure similarity (SSIM)[62].

Parameters. The number of MS input bands C is set to 12. We used 12 commonly used wide band filters: *Near IR* (N), *Dark Red* (DR), *Light Red* (LR), *Orange* (O), *Green* (G), *Photopic* (P), *Light Green* (LG), *Cyan* (C), *Green-Blue* (GB), *Absorptive Visible* (AV), *Blue* (B), and *Indigo* (I). The response functions of the filters are downloaded from the midopt database[1]. We will provide the detailed data in suppl material. Note that these 12 filters are by no means the best. We use them as they are common, diverse, and their response functions are available. Since the two HS datasets both offer 31 HS bands, we set $K = 31$ for our method. The size of our CFA is 8×8, i.e. $m = 8$. The sequence length of training for RL is set as $T = 200$. For the RL method, $\epsilon = 0.05$ and PSNR is used for ρ in Sec.3.4. We train the network for 20 epoches. For all our networks, we use the Adam optimizer and the initial learning rate is set to 0.0001. The batch size is set to 8. The training is done on one GTX TITAN X GPU.

4.2. Comparison to other methods

We compare to two state-of-the-art (SOTA) HS SR methods based on LR HS images: SSPSR [28] and MCNet [34], one SOTA HS SR method based on RGB images: AWAN Network [32], one SOTA RGB image super-resolution method LIIF [11], for which we change their input and output dimensions from 3 to 31 to perform HS image SR, and one SOTA RGB CFA-Demosaicing method [23] for which we changed the output channel from 3 to 31.

Since some methods cannot be directly used for our task, we calculate the scaling rate for them so that they will use the same amount of input pixels as our method uses. Specifically, for our method, the number of input pixels is $1/31$ that of the output pixels. Since the outputs of all methods are the same, i.e. the K-channel HS images Y, we use the ratio $1/31$ to calculate the size of the input for other methods. Therefore, for SSPSR, MCNet and LIIF, their input LR HS images are 5.6 ($31^{0.5} = 5.6$) times smaller in each of the 2D spatial dimensions than the output HS images. Since SSPSR and MCNet cannot handle arbitrary scale, we report their performance for scaling factor $\times 4$. Note that this gives a clear advantage to these two methods.

We report the main results in Table 1. It can be found that our method significantly outperforms all other comparison methods, even when a large advantage is given to SSPSR and MCNet. They solve a simpler super-resolution task

[1] midopt: `https://midopt.com/filters/bandpass/`

	Method	RMSE ↓	CC↑	MPSNR↑	MSSIM↑	ERGAS↓	SAM ↓
CAVE	SSPSR [28]	0.01245	0.99317	42.13787	0.96457	3.55146	3.83398
	MCNet [34]	0.01245	0.99283	42.25978	0.96465	3.56246	3.84976
	AWAN [32]	0.02814	0.93421	37.5632.	0.92312	3.84321	3.98342
	LIIF [11]	0.01884	0.92957	38.62866	0.94675	3.63215	3.75236
	CFA-Demosaicing [23]	0.01433	0.99109	42.11352	0.96345	3.53781	3.76934
	Ours	**0.01146**	**0.99746**	**43.70456**	**1.02484**	**3.43840**	**3.45786**
Harvard	SSPSR [28]	0.01352	0.96059	40.81499	0.92806	3.05007	3.24930
	MCNet [34]	0.01405	0.96009	40.59229	0.92658	3.10529	2.59147
	AWAN [32]	0.02437	0.92285	36.45873	0.91432	4.74317	7.97364
	LIIF [11]	0.02139	0.93542	37.58447	0.93996	4.63137	8.24718
	CFA-Demosaicing [23]	0.01467	0.99092	42.02315	0.95849	3.55324	3.12792
	Ours	**0.01219**	**0.99233**	**43.01718**	**0.98299**	**2.90819**	**2.81355**

Table 1: Results on the CAVE and the Harvard dataset. Note that SSPSR and MCNet solve a simpler task, i.e. solving a $\times 4$ SR task instead of a $\times 5.6$ one.

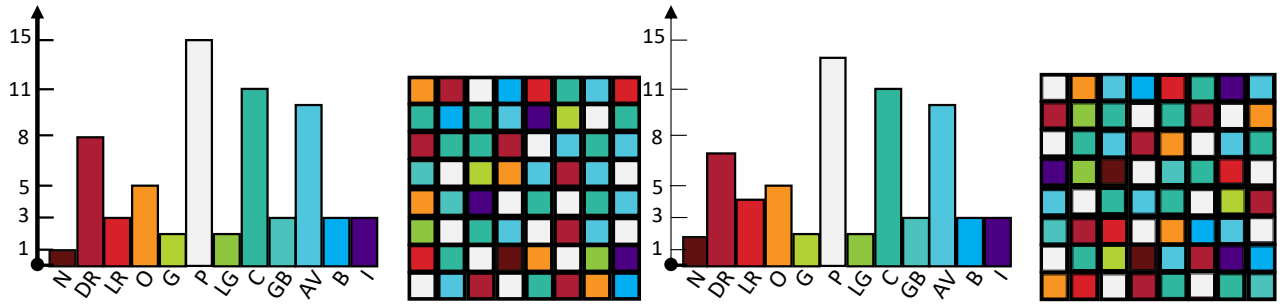
(a) Bilinear Interpolation (b) Demosaicing [23] (c) Ours (d) Ground Truth

Figure 2: Visual Results. Spectral band 5, 15, and 25 are used as the R, G, and B channel of a color image for this visualization. Better to see on screen.

instead of a $\times 5.6$ one. The latter one roughly has the same level of difficulty as our task. The superiority of our method is due to its flexibility to find a balance between spatial resolution and spectral resolution. The search of this balance is driven by the performance of the final spectral image recovery. Our method optimizes all these relevant sub-tasks jointly while others only focus on part of the game and mostly ignore the image sampling part. This limits their capability of finding the optimal solution. Compared to [23], our method is able to learn with a much larger set of filters (12 vs. 3) and is able to handle more irregular and sparse CFA patterns. These all contribute to its good performance. The visual results in Fig. 2 further show that our method can recover both spectral bands and spatial structures more accurately than other methods. More visual results are shown in the supplementary material.

4.3. Ablation Studies

We further study the contribution of these components of our method: the filter set, filter appearance frequency prediction, CFA learning, and the demosaicing method. The spectral recovery network is an existing, top-performing HS SR network, so we will not compare it with other alternative network in this work. All the results of our ablation study on the CAVE dataset are shown in Table 2. There are a few insights can be drawn from the table. First, one can see that a very basic baseline method using a large set of filters (row 3) can outperform methods using RGB filters and the Bayer pattern (row 1 and row 2). This highlights the fact that RGB images are not optimal for recovering HS images. Second, the table (row 3 vs. row 6, row 4 vs. row 7, and row 5 vs. row 8) shows that using the right number of filters for each spectral band (band frequency) is very important. Therefore, our dedicated RL-based band selection method is useful and crucial. Third, it can also be found from the table (row 3 vs. row 4, and row 6 vs. row 7) that our CFA method is effective, showing that better arrangement of color filters is important as well. Finally, the table shows that when all the components are combined (row 9), our method yields the best performance, showing that all the proposed components are important and the end-to-end learning synergizes them well.



(a) Bands: C (b) CFA: C (c) Bands: H (d) CFA: H

Figure 3: The bands and the CFAs determined: C for CAVE and H for Harvard.

Table 2: Component ablation of our method.

	Filter Set	Band Frequency	CFA	Demosaicing	HS Recovery	RMSE ↓
1	{R,G,B}	{0.25,0.5,0.25}	Bayer	Bilinear Interpolation	LIIF [11]	0.01884
2	{R,G,B}	Ours	Ours	Ours	Ours	0.01693
3	Ours	{1/12, 1/12, ..., 1/12}	Random	Bilinear Interpolation	Ours	0.01735
4	Ours	{1/12, 1/12, ..., 1/12}	Ours	Bilinear Interpolation	Ours	0.01677
5	Ours	{1/12, 1/12, ..., 1/12}	Ours	Ours	Ours	0.01623
6	Ours	Ours	Random	Bilinear Interpolation	Ours	0.01478
7	Ours	Ours	Ours	Bilinear Interpolation	Ours	0.01322
8	Ours	Ours	Random	Ours	Ours	0.01379
9	Ours	Ours	Ours	Ours	Ours	**0.01146**

4.4. Selected Bands and CFAs

In Fig. 3, we show the appearance times of the 12 filters identified by our method, and the corresponding CFAs determined by our method. The results are in line with intuition that some filters are indeed more important than others. Filters that are selected the most are: *Dark Red*, *Orange*, *Photopic*, *Cyan*, and *Absorptive Visible*. They span over the whole spectral range that our target HS images lie in; they are also evenly spaced so that for every spectral region there is a high-resolution image. It seems that the algorithm 'sacrifices' some spectral bands in exchange for high spatial resolution of some other bands. It strikes a balance between spatial resolution and spectral resolution. The least selected filter is *Near IR*. This also makes sense as most parts of its response function lie outside of the considered spectral range. The learned CFA shows that the filters of all bands are quite uniformly distributed. This is beneficial for the demosaicing algorithm. Note that the result in Fig. 3c (a) and (b) are not exactly consistent. This is normal as (b) is an optimization result guided by (a) so there can be slight inconsistency.

One can also find from the figure that the spectral bands and their frequencies identified by our method are highly consistent over the two datasets, though they are not identical. The CFAs look quite different. We further investigate whether the bands and the CFAs identified are transferable to a different dataset. That is, the two CFAs identified generate similar performance on both datasets. For this experiment, we take the bands and CFA identified on one dataset and apply them to the other dataset where we only re-train the demosaicing and spectral recovery networks. Our experiments show that we get on par performance (for CAVE to Harvard, the RMSE changes from 0.01219 to 0.01224, and for Harvard to CAVE, the RMSE changes from 0.01146 to 0.01157). That means the bands and CFA found on one dataset can be transferred to other datasets.

5. Conclusions

In this work, we have developed a method for a fast and low-cost hyperspectral (HS) imaging system. The method has achieved state-of-the-art performance by jointly learning multiple related tasks: spectral band selection, CFA optimization, image demosaicing for irregular measurements, and spectral recovery. We have developed specialised neural networks for all these tasks and they can trained jointly to avoid sub-optimal solutions. Experiments show that our method outperforms other methods significantly. Designing hardware prototypes for this method is our future work.

References

[1] Hemant Kumar Aggarwal and Angshul Majumdar. Multi-spectral demosaicing technique for single-sensor imaging. In *National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG)*, 2013.

[2] H. K. Aggarwal and A. Majumdar. Hyperspectral image denoising using spatio-spectral total variation. *IEEE Geoscience and Remote Sensing Letters*, 13(3):442–446, 2016.

[3] Boaz Arad and Ohad Ben-Shahar. Filter selection for hyperspectral estimation. In *ICCV*, 2017.

[4] Boaz Arad, Radu Timofte, Ohad Ben-Shahar, Yi-Tun Lin, and Graham D. Finlayson. Ntire 2020 challenge on spectral reconstruction from an rgb image. In *CVPRW*, 2020.

[5] Boaz Arad, Radu Timofte, Rony Yahel, Nimrod Morag, Amir Bernat, Yaqi Wu, Xun Wu, Zhihao Fan, Chenjie Xia, Feng Zhang, Shuai Liu, Yongqiang Li, Chaoyu Feng, Lei Lei, Mingwei Zhang, Kai Feng, Xun Zhang, Jiaxin Yao, Yongqiang Zhao, Suina Ma, Fan He, Yangyang Dong, Shufang Yu, Difa Qiu, Jinhui Liu, Mengzhao Bi, Beibei Song, Wen-Fang Sun, Jiesi Zheng, Bowen Zhao, Yanpeng Cao, Jiangxin Yang, Yanlong Cao, Xiangyu Kong, Jingbo Yu, Yuanyang Xue, and Zheng Xie. Ntire 2022 spectral demosaicing challenge and data set. In *CVPRW*, 2022.

[6] Liheng Bian, Yugang Wang, and Jun Zhang. Generalized msfa engineering with structural and adaptive nonlocal demosaicing. *IEEE Transactions on Image Processing*, 30:7867–7877, 2021.

[7] Johannes Brauers and Til Aach. A color filter array based multispectral camera. In *12. Workshop Farbbildverarbeitung*. Ilmenau, 2006.

[8] Ayan Chakrabarti. Learning sensor multiplexing design through back-propagation. In *Advances in Neural Information Processing Systems*, 2016.

[9] Ayan Chakrabarti, William T. Freeman, and Todd Zickler. Rethinking color cameras. In *IEEE International Conference on Computational Photography (ICCP)*, 2014.

[10] Ayan Chakrabarti and Todd Zickler. Statistics of real-world hyperspectral images. In *CVPR 2011*, pages 193–200. IEEE, 2011.

[11] Yinbo Chen, Sifei Liu, and Xiaolong Wang. Learning continuous image representation with local implicit image function. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021.

[12] Cui Chi, Hyunjin Yoo, and Moshe Ben-Ezra. Multi-spectral imaging by optimized wide band illumination. *International Journal of Computer Vision*, 86(2):140–151, 2010.

[13] Christopher Choy, JunYoung Gwak, and Silvio Savarese. 4d spatio-temporal convnets: Minkowski convolutional neural networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3075–3084, 2019.

[14] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021.

[15] Kai Feng, Yongqiang Zhao, Jonathan Cheung-Wai Chan, Seong G. Kong, Xun Zhang, and Binglu Wang. Mosaic convolution-attention network for demosaicing multispectral filter array images. *IEEE Transactions on Computational Imaging*, 7:864–878, 2021.

[16] Y. Fu, T. Zhang, Y. Zheng, D. Zhang, and H. Huang. Hyperspectral image super-resolution with optimized rgb guidance. In *CVPR*, 2019.

[17] Ying Fu, Tao Zhang, Yinqiang Zheng, Debing Zhang, and Hua Huang. Joint camera spectral response selection and hyperspectral image recovery. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(1):256–272, 2022.

[18] Alexander F.H. Goetz. Three decades of hyperspectral remote sensing of the earth: A personal view. *Remote Sensing of Environment*, 113:S5 – S16, 2009.

[19] A.A. Gowen, C.P. O'Donnell, P.J. Cullen, G. Downey, and J.M. Frias. Hyperspectral imaging – an emerging process analytical tool for food quality and safety control. *Trends in Food Science & Technology*, 18(12):590 – 598, 2007.

[20] Vitor Guizilini, Rares Ambrus, Wolfram Burgard, and Adrien Gaidon. Sparse auxiliary networks for unified monocular depth prediction and completion. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.

[21] B. Guo, S.R. Gunn, R.I. Damper, and J.D.B. Nelson. Band selection for hyperspectral image classification using mutual information. *IEEE Geoscience and Remote Sensing Letters*, 3(4):522–526, 2006.

[22] Shuai Han, Imari Sato, Takahiro Okabe, and Yoichi Sato. Fast spectral reflectance recovery using dlp projector. *International journal of computer vision*, 110(2):172–184, 2014.

[23] Bernardo Henz, Eduardo S. L. Gastal, and Manuel M. Oliveira. Deep joint design of color filter arrays and demosaicing. *Computer Graphics Forum*, 37(2):389–399, 2018.

[24] John Hershey and Zhengyou Zhang. Multispectral digital camera employing both visible light and non-visible light sensing on a single image sensor, Dec. 2 2008. US Patent 7,460,160.

[25] Keigo Hirakawa and Patrick J. Wolfe. Spatio-spectral color filter array design for optimal image recovery. *IEEE Transactions on Image Processing*, 17(10):1876–1890, 2008.

[26] A. Ifarraguerri and M.W. Prairie. Visual method for spectral band selection. *IEEE Geoscience and Remote Sensing Letters*, 1(2):101–106, 2004.

[27] Junjun Jiang, He Sun, Xianming Liu, and Jiayi Ma. Learning spatial-spectral prior for super-resolution of hyperspectral imagery. *IEEE Transactions on Computational Imaging*, 6:1082–1096, 2020.

[28] J. Jiang, H. Sun, X. Liu, and J. Ma. Learning spatial-spectral prior for super-resolution of hyperspectral imagery. *IEEE Transactions on Computational Imaging*, 6:1082–1096, 2020.

[29] S Kaur and Vijay Kumar Banga. A survey of demosaicing: Issues and challenges. *International Journal of Science, Engineering and Technologies*, 2(1):2, 2015.

[30] N. Keshava. Best bands selection for detection in hyperspectral processing. In *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.01CH37221)*, volume 5, pages 3149–3152 vol.5, 2001.

[31] Pierre-Jean Lapray, Xingbo Wang, Jean-Baptiste Thomas, and Pierre Gouton. Multispectral filter arrays: Recent advances and practical implementation. *Sensors*, 14(11):21626–21659, 2014.

[32] Jiaojiao Li, Chaoxiong Wu, Rui Song, Yunsong Li, and Fei Liu. 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 (CVPR) Workshops*, 2020.

[33] Ke Li, Dengxin Dai, and Luc Van Gool. Hyperspectral image super-resolution with spectral mixup and heterogeneous datasets. In *WACV*, 2021.

[34] Qiang Li, Qi Wang, and Xuelong Li. Mixed 2d/3d convolutional network for hyperspectral image super-resolution. *Remote Sensing*, 12(10), 2020.

[35] Xin Li, Bahadir Gunturk, and Lei Zhang. Image demosaicing: A systematic survey. In *Visual Communications and Image Processing 2008*, volume 6822, page 68221J. International Society for Optics and Photonics, 2008.

[36] Yong Li, Lei Zhang, Chen Dingl, Wei Wei, and Yanning Zhang. Single hyperspectral image super-resolution with grouped deep recursive residual network. In *2018 IEEE Fourth International Conference on Multimedia Big Data (BigMM)*, pages 1–4. IEEE, 2018.

[37] Laetitia Loncan, Luis B De Almeida, José M Bioucas-Dias, Xavier Briottet, Jocelyn Chanussot, Nicolas Dobigeon, Sophie Fabre, Wenzhi Liao, Giorgio A Licciardi, Miguel Simoes, et al. Hyperspectral pansharpening: A review. *IEEE Geoscience and remote sensing magazine*, 3(3):27–46, 2015.

[38] Yue M. Lu, Clément Fredembach, Martin Vetterli, and Sabine Süsstrunk. Designing color filter arrays for the joint capture of visible and near-infrared images. In *ICIP*, 2009.

[39] Yue M Lu and Martin Vetterli. Optimal color filter array design: Quantitative conditions and an efficient search procedure. In *Digital Photography V*, volume 7250, page 725009. International Society for Optics and Photonics, 2009.

[40] Guolan Lua and Baowei Fei. Medical hyperspectral imaging: a review. *Journal of Biomedical Optics*, 2014.

[41] Chenguang Ma, Xun Cao, Xin Tong, Qionghai Dai, and Stephen Lin. Acquisition of high spatial and spectral resolution video with a hybrid camera system. *International journal of computer vision*, 110(2):141–155, 2014.

[42] Shaohui Mei, Xin Yuan, Jingyu Ji, Yifan Zhang, Shuai Wan, and Qian Du. Hyperspectral image spatial super-resolution via 3d full convolutional neural network. *Remote Sensing*, 9(11):1139, 2017.

[43] Lidan Miao and Hairong Qi. The design and evaluation of a generic method for generating mosaicked multispectral filter arrays. *IEEE Transactions on Image Processing*, 15(9):2780–2791, 2006.

[44] Lidan Miao, Hairong Qi, Rajeev Ramanath, and Wesley E Snyder. Binary tree-based generic demosaicking algorithm for multispectral filter arrays. *IEEE Transactions on Image Processing*, 15(11):3550–3558, 2006.

[45] Lidan Miao, Hairong Qi, and Wesley E Snyder. A generic method for generating multispectral filter arrays. In *ICCP*, 2004.

[46] Andreas Michel, Wolfgang Gross, Fabian Schenkel, and Wolfgang Middelmann. Hyperspectral band selection within a deep reinforcement learning frame-

work. In *IEEE International Geoscience and Remote Sensing Symposium*, 2020.

[47] Yusukex Monno, Sunao Kikuchi, Masayuki Tanaka, and Masatoshi Okutomi. A practical one-shot multispectral imaging system using a single image sensor. *IEEE Transactions on Image Processing*, 24(10):3048–3059, 2015.

[48] Lichao Mou, Sudipan Saha, Yuansheng Hua, Francesca Bovolo, Lorenzo Bruzzone, and Xiao Xiang Zhu. Deep reinforcement learning for band selection in hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–14, 2022.

[49] N. M. Nasrabadi. Hyperspectral target detection : An overview of current and future challenges. *IEEE Signal Processing Magazine*, 31(1):34–44, 2014.

[50] Rang M. H. Nguyen, Dilip K. Prasad, and Michael S. Brown. Training-based spectral reconstruction from a single rgb image. In *ECCV*, 2014.

[51] Shijie Nie, Lin Gu, Yinqiang Zheng, Antony Lam, Nobutaka Ono, and Imari Sato. Deeply learned filter response functions for hyperspectral reconstruction. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018.

[52] Seoung Wug Oh, Michael S Brown, Marc Pollefeys, and Seon Joo Kim. Do it yourself hyperspectral imaging with everyday digital cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2461–2469, 2016.

[53] Jong-Il Park, Moon-Hyun Lee, Michael D Grossberg, and Shree K Nayar. Multispectral imaging using multiplexed illumination. In *International Conference on Computer Vision*, 2007.

[54] J.C. Price. Spectral band selection for visible-near infrared remote sensing: spectral-spatial resolution tradeoffs. *IEEE Transactions on Geoscience and Remote Sensing*, 35(5):1277–1285, 1997.

[55] Rajeev Ramanath, Wesley E Snyder, Griff L Bilbro, and William A Sander. Robust multispectral imaging sensors for autonomous robots. *Tech. Rep.*, 2001.

[56] Travis W. Sawyer, Michaela Taylor-Williams, Ran Tao, Ruqiao Xia, Calum Williams, and Sarah E. Bohndiek. Opti-msfa: a toolbox for generalized design and optimization of multispectral filter arrays. *Opt. Express*, 30(5):7591–7611, 2022.

[57] S.B. Serpico and L. Bruzzone. A new search algorithm for feature selection in hyperspectral remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 39(7):1360–1367, 2001.

[58] Alberto Signoroni, Mattia Savardi, Annalisa Baronio, and Sergio Benini. Deep learning meets hyperspectral image analysis: A multidisciplinary review. *Journal of Imaging*, 5(5), 2019.

[59] Bo Sun, Junchi Yan, Xiao Zhou, and Yinqiang Zheng. Tuning ir-cut filter for illumination-aware spectral reconstruction from rgb. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2021.

[60] Lucien Wald. *Data fusion: definitions and architectures: fusion of images of different spatial resolutions*. Presses des MINES, 2002.

[61] Lizhi Wang, Zhiwei Xiong, Dahua Gao, Guangming Shi, Wenjun Zeng, and Feng Wu. High-speed hyperspectral video acquisition with a dual-camera architecture. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2015.

[62] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[63] Renjie Wu, Yuqi Li, Xijiong Xie, and Zhijie Lin. Optimized multi-spectral filter arrays for spectral reconstruction. *Sensors*, 19(13):2905, 2019.

[64] Qi Xie, Minghao Zhou, Qian Zhao, Deyu Meng, Wangmeng Zuo, and Zongben Xu. Multispectral and hyperspectral image fusion by ms/hs fusion net. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1585–1594, 2019.

[65] Fumihito Yasuma, Tomoo Mitsunaga, Daisuke Iso, and Shree K Nayar. Generalized assorted pixel camera: postcapture control of resolution, dynamic range, and spectrum. *IEEE transactions on image processing*, 19(9):2241–2253, 2010.

[66] Roberta H Yuhas, Alexander FH Goetz, and Joe W Boardman. Discrimination among semi-arid landscape endmembers using the spectral angle mapper (sam) algorithm. In *Proc. Summaries 3rd Annu. JPL Airborne Geosci. Workshop*, volume 1, pages 147–149, 1992.

[67] Yuanyuan Zhao, Hui Guo, Zhan Ma, Xun Cao, Tao Yue, and Xuemei Hu. Hyperspectral imaging with random printed mask. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019.

[68] Tiancheng Zhi, Bernardo R Pires, Martial Hebert, and Srinivasa G Narasimhan. Multispectral imaging for fine-grained recognition of powders on complex backgrounds. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019.