DRCR Net: Dense Residual Channel Re-calibration Network with Non-local Purification for Spectral Super Resolution

Jiaojiao Li, Songcheng Du*, Chaoxiong Wu, Yihong Leng, Rui Song, Yunsong Li
The State Key Laboratory of Integrated Service Networks, Xidian University, Xian 710071, China
{jjli, rsong, ysli}@xidian.edu.cn, {dusongcheng, cxwu, yhleng}@stu.xidian.edu.cn

Abstract

Spectral super resolution (SSR) aims to reconstruct the 3D hyperspectral signal from a 2D RGB image, which is prosperous with the proliferation of Convolutional Neural Networks (CNNs) and increased access to RGB/hyperspectral datasets. Nevertheless, most CNN-based spectral reconstruction (SR) algorithms can only perform high reconstruction accuracy when the input RGB image is relatively 'clean' with foregone spectral response functions. Unfortunately, in the real world, images are contaminated by mixed noise, bad illumination conditions, compression, artifacts etc. and the existing state-of-the-art (SOTA) methods are no longer working well. To conquer these drawbacks, we propose a novel dense residual channel re-calibration network (DRCR Net) with non-local purification for achieving robust SSR results, which first performs the interference removal through a non-local purification module (NPM) to refine the RGB inputs. To be specific, as the main component of backbone, the dense residual channel re-calibration (DRCR) block is cascaded with an encoder-decoder paradigm through several cross-layer dense residual connections, to capture the deep spatial-spectral interactions, which further improve the generalization ability of the network effectively. Furthermore, we customize dual channel re-calibration modules (CRMs) which are embedded in each DRCR block to adaptively re-calibrate channel-wise feature response for pursuing high-fidelity spectral recovery. In the NTIRE 2022 Spectral Reconstruction Challenge, our entry obtained the 3rd ranking. Code will be made available online at https://github.com/jojolee6513/DRCR-net.

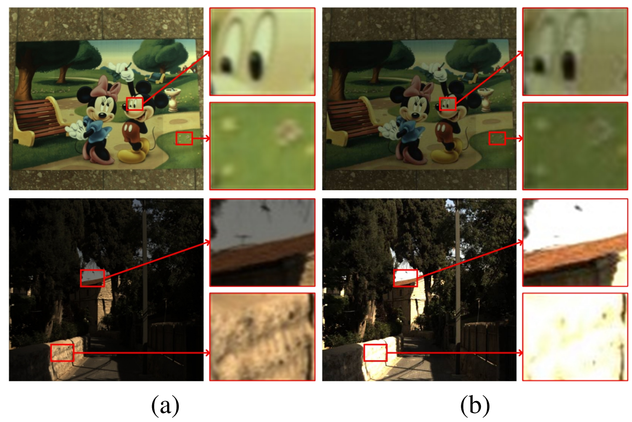

Figure 1. **Visual comparison.** (a) clean RGB images with more crisp edges and less artifacts which facilitate high-precision spectral reconstruction, (b) degraded RGB images that conform to the real world more but interfere with spectral super resolution. It is best to view the examples in color on a high-resolution display by zooming in close.

1. Introduction

The hyperspectral imaging equipment can acquire hundreds of thousands of fine, continuous and narrow-band spectra of an actual scene in a wide electromagnetic spectral range, and simultaneously collect the geometry, radiation or reflection information of the target, form a data cube. These spectral information with high spectral resolution have been proven to have promoting effect in various domains, including remote sensing, agriculture, as well as computer vision applications such as image classification [29,30] , target detection [21], face recognition [33], etc.

However, recent developments in HS imaging have been hampered by a bottleneck associated with its dependence on hardware conditions. There exists a heavy trade off between the spectral resolution and the spatial or temporal resolution, therefore, to obtain high spectral resolution, spatial or temporal resolution has to be reduced, which severely restricts the applications of hyperspectral images (HSIs) [24]. Additionally, the expensive equipment also remains an insurmountable barrier for large-scale acquisition of HSIs, despite researchers continuing to optimize and improve on the above issues. To achieve real-time and low-cost HSI data

*Corresponding Author

acquisition, several scholars tend to acquire HSIs data only from a corresponding RGB image and such a process is defined as spectral super resolution (SSR) or spectral reconstruction [13, 15, 18, 22].

Generally speaking, one of the main inherent challenges in such methods is their severe ill-posedness since an infinite amount of HSI data can be projected to the same input RGB image under the similar constraints. Attempting to alleviate this issue, a variety of methods have been presented in the literature. Early approaches mainly work by exploiting the inherent statistical priors of HSI such as sparsity, low rankness [4, 11, 32],etc. However, limited by the fewer datasets and the robustness and generalization of the methods themselves, the results of reconstruction in real scene were not satisfactory. Later, as the neural networks based methods gradually became prevalent and in view of impressive success in many computer vision tasks [9,20,27], CNN-based models are also exploited in the SSR task [17, 19, 36]. Despite the very accurate reconstruction performance has been implemented in SSR when dealing with the clean RGB data as shown in Fig. 1 (a), such as the NTIRE2020 and NTIRE2018 datasets [6,7] where the RGB images almost directly available through given HSIs and the corresponding spectral response curve (SRC), these methods are not robust enough to inferior RGB images as depicted in Fig. 1 (b) contaminating under some realistic conditions e.g. bad illumination (overexposure or underexposure), loss created by compression and information, mixed noise, and some inevitable artifacts. In addition, some existing models blindly pursue the algorithm complexity and ignore exploiting the information interaction between intermediate features to further improve the expressiveness of CNNs.

To solve these drawbacks, in this paper, we present a novel dense residual channel re-calibration network (DRCR Net) with non-local purification for SSR which takes noisy, overexposed and compressed RGB images as inputs. Since the aforementioned degradations all bring challenges to the SSR task, therefore, a simple but effective Non-local Purification module (NPM) which can adaptively adjust its own values by exploiting the adjacent pixels is first employed to refine the RGB input. Besides, due to that the artifacts have a negative impact on image quality at various scales, such as where noise affects at a local point level, compression at a local region level and poor illumination at a large area level, thus, a pyramid-like hierarchical structure is designed to solve this multi-scale problem. In addition, our backbone is designed through cascading numerous dense residual channel re-calibration (DRCR) blocks, which is composed of encoder and decoder with several robust cross-layer connections for deep spatial-spectral feature extraction and interactions between intermediate features. Moreover, enhancing the network's discriminative learning capabilities by integrating channel-wise feature interdependencies is particularly important. As such, the dual symmetrical channel re-calibration modules (CRMs) which are embedded in each DRCR block are developed to accurately re-calibrate channel-wise feature responses by explicitly predicting interdependencies between channels and to improve CNN learning capabilities.

The main contributions of our work can be summarized as the following.

- Generally, we proposed the DRCR Net with non-local purification for SSR, which takes an RGB image with severe artifacts as input. Experimental results on NTIRE 2022 Hyperspectral dataset demonstrate the effectiveness of DRCR Net, and our entry obtained the 3rd place on Spectral Reconstruction from RGB.

- As the main component of DRCR Net, DRCR block which is cascaded with an encoder-decoder paradigm through several cross-layer dense residual connections is employed to capture the deep spatial-spectral interactions and improve the generalization ability of the network.

- To eliminate the degeneration of various artifacts, such as noise, compression and poor illumination polluting the input RGB images at different scales, we design a simple but effective NPM to intrinsically reduce the impact of different scale artifacts on the subsequent reconstruction process via a hierarchical structure.

- Dual CRMs embedded in each DRCR block are developed to adaptively re-calibrate channel-wise feature responses through explicitly modeling interdependencies between channels for pursuing high-fidelity spectral recovery.

2. Related Work

Hyperspectral SSR from RGB images offers an excellent alternative to acquire hyperspectral data without alteration of the hyperspectral imaging devices, which was put into practice earlier when Agahian $et.al$ [2] formulated such problem as combination of a few basis functions generated from spectral reflectance data sets. Nguyen $et.al$ [23] put forward a radial basis function shallow network to normalize the scene illumination and recovered the scene reflectance by using RGB white-balancing. Later, Arad $et.al$ [5] leveraged the sparsity prior of HSI to tackle this issue which first created a sparse over-complete spectral dictionary by K-SVD algorithm [3] from their collected dataset and used the orthogonal matching pursuit (OMP) algorithm [10] to recover the HSI from RGB images on the basis of the spectral response curve (SRC) prior. Based on the large improvements over the above sparse coding

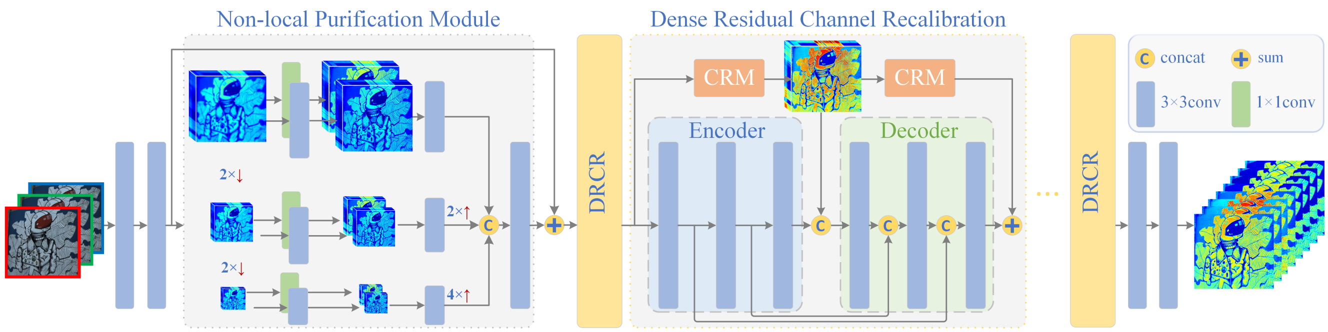

Figure 2. **Architecture of dense residual channel re-calibration network for Spectral Super Resolution from RGB Images.** ↓ and ↑ indicate downsampling and upsampling, respectively.

method, Aeschbacher *et.al* [1] presented a shallow learned SSR method which called A+ and pushed the performance.

These methods, however, are restricted to domain-specific images due to their poor expressiveness and limited generalizability. In recent years, as CNN has demonstrated superior performance for solving nonlinear problems and a number of studies have attempted to perform this three-to-many mapping by conducting an end-to-end CNN network. Xiong *et.al* [31] presented a unified deep learning based framework for SSR which input the spectrally up-sampled image and output the enhanced hyperspetral one. With the holding of the two spectral reconstruction competitions, NTIRE2018 [6] and NTIRE2020 [7], a large number of SSR algorithms with high restoration accuracy have been proposed. Shi *et.al* [25] proposed two advanced networks named as HSCNN-D and HSCNN-R, which both equipped with several dense blocks, respectively, and won the 1st and 2nd place on both the 'Clean' and 'Real World' tracks in the NTIRE2018 spectral reconstruction challenge. Li *et.al* [16] proposed an adaptive weighted attention network (AWAN) consisting of multiple dual residual attention blocks and won the championship in NTIRE2020 spectral reconstruction challenge on the 'Clean' track. Moreover, Zhao *et.al* [35] presented a four-level hierarchical regression network (HRNet) which replaced the downsampling and upsampling operation with PixelUnShuffle as well as PixelShuffle and was the winning method of 'real world' track. Fu *et.al* [12] took SRC in to account and explored implementing a SRC optimization layer based on an unsupervised CNN-based algorithm either for selecting a SRC from a dataset or designing a SRC that meets all the restrictions imposed by physical space. Very recently, Wu *et.al* [28] investigated multi-source and priors that include spatial contexts, semantic information from RGB images, deep feature-prior and band-wise correlations of HSIs for enhancing the accuracy of SSR. Zhang *et.al* [34] proposed a single Meta-Learning-Based model for recovering high spectral resolution HSI, which could be trained for a wide variety of input-output band settings. However, none of them take the degradation of the input RGB images in the real conditions into consideration and fail to achieve the similar superior reconstruction accuracy as in the clean dataset.

3. Our Proposed Method

3.1. Network Architecture

In this section, we describe our proposed dense residual channel re-calibration network (DRCR Net) in detail. Given $\mathbf{I}_{RGB}$ as the input of DRCR Net. As illustrated in Fig. 2, we first employ two convolutional layers to extract the shallow feature $\mathbf{F}_0$ as well as boost the number of bands from input RGB images.

$$\mathbf{F}_{SF} = H_{SFE\_convs}(\mathbf{I}_{RGB}), \tag{1}$$

where $H_{SFE\_convs}(\cdot)$ stands for front convolution operations. Then we use the extracted shallow feature $\mathbf{F}_{SF}$ as the input of the non-local purification module (NPM). Thus we can further have

$$\mathbf{F}^0 = H_{DF}(\mathbf{F}_{SF}), \tag{2}$$

where $H_{DF}(\cdot)$ represents our designed very simple but efficient NPM whose output $\mathbf{F}^0$ is then taken as the input of our multiple dense residual channel re-calibration (DRCR) blocks.

$$\begin{aligned}\mathbf{F}^m &= H^m_{DRCR}(\mathbf{F}^{m-1}) \\ &= H^m_{DRCR}(H^{m-1}_{DRCR}(\cdots H^1_{DRCR}(\mathbf{F}^0)\cdots)),\end{aligned} \tag{3}$$

where $\mathbf{F}^m$ and $\mathbf{F}^{m-1}$ denote the output and the input of the mth DRCR block, separately. $H^m_{DRCR}(\cdot)$ represents the mth DRCR block. To be specific, the DRCR blocks have a U-shaped structure in which the encoding part and decoding part consist of three 3×3 plain convolution layers, separately. Additionally, three concatenation operations between the encoding part and decoding part are utilized to

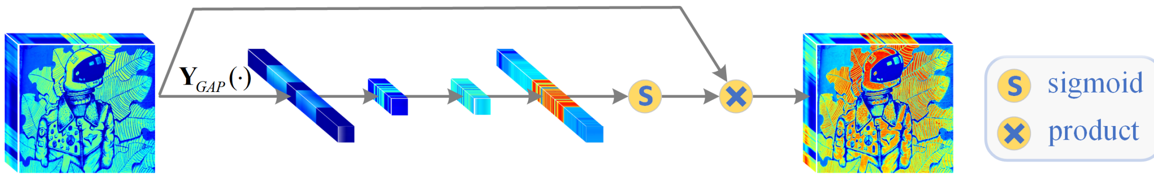

Figure 3. **Architecture of channel re-calibration module**. In the figure, $\mathbf{Y}_{GAP}(\cdot)$ represents the global average pooling operation.

explore the information interaction among the intermediate layers and such skip cross-layer connections help to alleviate the vanishing gradient problem. Besides, we employ the dual channel re-calibration module (CRM) to re-calibrate the features associated with the DRCR block along the channel dimension, where the first CRM $H^{(m,1)}_{DCRM}(\cdot)$ draws the calibration feature $\mathbf{F}^{m,1}_{RF}$ from the input of the mth DRCR block. The above process can be expressed as

$$\mathbf{F}^{(m,1)}_{RF} = H^{(m,1)}_{DCRM}(\mathbf{F}^{m-1}) \tag{4}$$

We then fuse the $\mathbf{F}^{(m,1)}_{RF}$ with the aggregated features in the middle layer of mth DRCR block, additionally, we also input the $\mathbf{F}^{(m,1)}_{RF}$ into the second DCRM $H^{(m,2)}_{DCRM}(\cdot)$ for further calibration of the channel-dimensional features and the above process can be formulated as

$$\mathbf{F}^{(m,2)}_{RF} = H^{(m,2)}_{DCRM}(\mathbf{F}^{(m,1)}_{RF}), \tag{5}$$

Therefor, the output of the ith convolution layer in mth DRCR block can be expressed as:

$$\mathbf{F}^{(m,i)} = \begin{cases} H^{(m,i)}_{DRCR\_conv}(\mathbf{F}^{m-1}) & i = 1 \\ H^{(m,i)}_{DRCR\_conv}(\mathbf{F}^{(m,i-1)}) & i = 2,3 \\ H^{(m,i)}_{DRCR\_conv}([\mathbf{F}^{(m,1)}_{RF}, \mathbf{F}^{(m,i-1)}]) & i = 4 \\ H^{(m,i)}_{DRCR\_conv}([\mathbf{F}^{m,i-1}, \mathbf{F}_{(m,7-i)}]) & i = 5,6, \end{cases} \tag{6}$$

where $H^{(m,i)}_{DRCR\_conv}(\cdot)$ denotes the ith convolution operation of mth DRCR block, and $[\cdot,\cdot]$ represents the concatenation operation of two features. Moreover, we add $\mathbf{F}^{(m,2)}_{RF}$ to the output of the last convolution layer $\mathbf{F}^(m,6)$, thus we can further have

$$\mathbf{F}^{m} = \mathbf{F}^{m,i-1} + \mathbf{F}^{(m,2)}_{RF}. \tag{7}$$

Finally, similar to the front structure of the network, we use two plain convolutions to aggregate features and map the number of bands to 31 to obtain the spectral reconstructed HSI $\mathbf{I}_{SR}$.

$$\mathbf{I}_{SR} = H_{AF\_convs}(\mathbf{F}^{m}), \tag{8}$$

where $H_{AF\_convs}(\cdot)$ stands for tail convolution operations.

3.2. Non-local Purification Module (NPM)

In the process of image generation and transmission, the image quality is often degraded due to the interference of various external conditions or human factors, which bring a very negative impact on the subsequent image processing such as SSR. Therefore, prior to processing the input image, artifact removal must be performed. And ideally, such practical algorithms should be flexible, efficient, and capable of handling various kinds of artifacts. Unfortunately, the current algorithms fall far short of all of these goals. In this paper, as shown in Fig. 3, we present a simple but efficient plug-and-play NPM to refine the degraded RGB images.

In general, a hierarchical pyramid structure is designed to exploit multi-scale information and remove the artifacts with different scales. For each level, the process can be decomposed into the acquisition of information from different receptive fields, non-local information guided adaptive purification and refinement.

Specifically, considering the computational cost and reconstruction accuracy, we choose a three-layer structure and first perform two downsampling operations on $\mathbf{F}_{SF}$,

$$\mathbf{F}_{SF\_2} = \mathbf{F}_{SF} \downarrow_2, \tag{9}$$

$$\mathbf{F}_{SF\_4} = \mathbf{F}_{SF} \downarrow_4 . \tag{10}$$

Then, for each level, the parallel 3×3 convolutional and 1×1 convolutional operation involve gathering information from both non-local as well as local sources. After that, concatenation operations followed by a 3×3 convolution are used to integrate the information captured by the different receptive fields. Thus, the local pixel value can be adjusted and refined based on the perception of non-local area information for artifact removal and self-purification. After that, we perform $2\times$ and $4\times$ bilinear upsampling operations on the output results of the two bottom layers, respectively, to achieve uniform spatial sizes. Finally, a concatenation operation and a 3×3 convolution are adopted again to integrate artifacts removal features from different scales. Due to the potential loss of boundary information during self-purification, a residual connection with the original feature $\mathbf{F}_{SF}$ is adopted.

3.3. Dual Channel Re-calibration Module (CRM)

Enhancing the discriminative learning capabilities of the network by integrating channel-wise feature interdependencies is especially important. As such, the dual CRMs based on [14] that are embedded in each DRCR block are

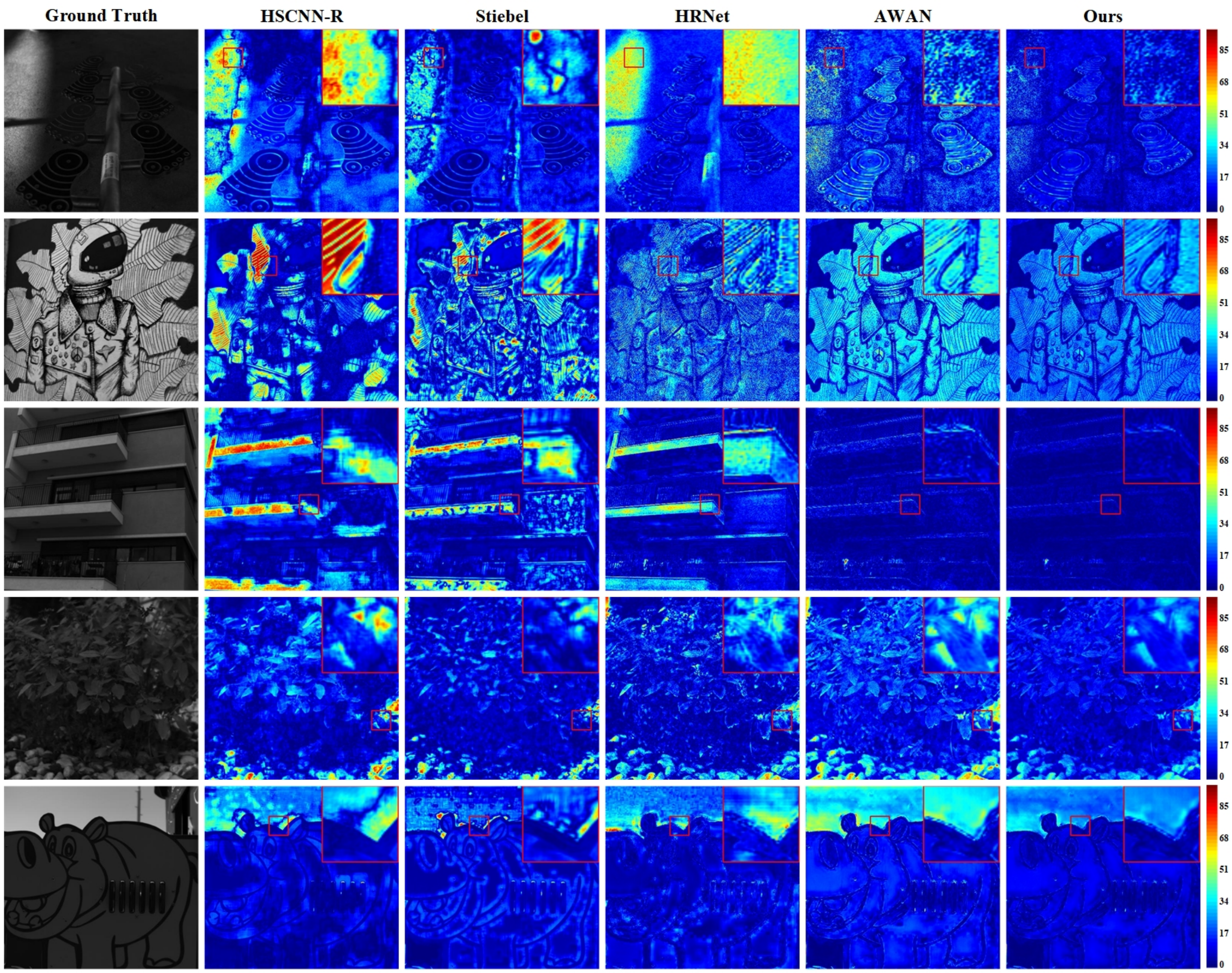

Figure 4. **Visual quality comparison of the** 18**-th band on five verification set images of NTIRE2022.** The ground truth,MRAE heat maps for HSCNN-R/Stibel/HRNet/AWAN/Ours methods. Note that the MRAE heat maps have been scaled for optimal display.

designed to adaptively calibrate channel-wise feature responses by explicitly modeling interdependencies between channels as well as strengthen the discriminant learning capability of CNNs.

Specifically, we first perform a global average pooling on the input of a DRCR block, for example $\mathbf{F}^{m-1}$ and turn it into a channel descriptor $\mathbf{F}_c \in R^{C\times 1\times 1}$. Then, to obtain the feature interdependencies in terms of channels, a three-layer fully connected structure which can perform nonlinear transformations is adopted. Thus, the output $\mathbf{F}_{CFI}$ can be formulated as

$$\mathbf{F}_{CFI} = \delta(\mathbf{W}_V(\mathbf{W}_U(\mathbf{W}_D(\mathbf{Y}_{GAP}(\mathbf{F}_{n-1}))))), \tag{11}$$

where $\mathbf{Y}_{GAP}(\cdot)$ denotes the global average pooling operation, and $\mathbf{W}_D(\cdot) \in R^{C\times C/r}$, $\mathbf{W}_U(\cdot) \in R^{C/r\times C/r}$, $\mathbf{W}_V(\cdot) \in R^{C/r\times C}$ are the weight sets of three-layer fully connected structure, r denotes the reduction ratio and $\delta(\cdot)$ represents the sigmoid function. Then we broadcast the $\mathbf{F}_{CFI}$ along the spatial dimension to channel re-calibration map $\mathbf{F}_{CRM} \in R^{C\times H\times W}$ and rescale the input through a dot product

$$\mathbf{F}_{RF}^{(m,1)} = \mathbf{F}_{CRM} \odot \mathbf{F}^{m-1}. \tag{12}$$

4. Experiments

4.1. Settings

Hyperspectral datasets. In this paper we use all three datasets published in the Spectral Reconstruction Challenge: NTIRE2018 [6], NTIRE2020 [7] and NTIRE2022 [8]. Considering that the non-local purification module (NPM) is specially designed in our module, therefore, in the NTIRE2018 and NTIRE2020 datasets we adopt 'Real

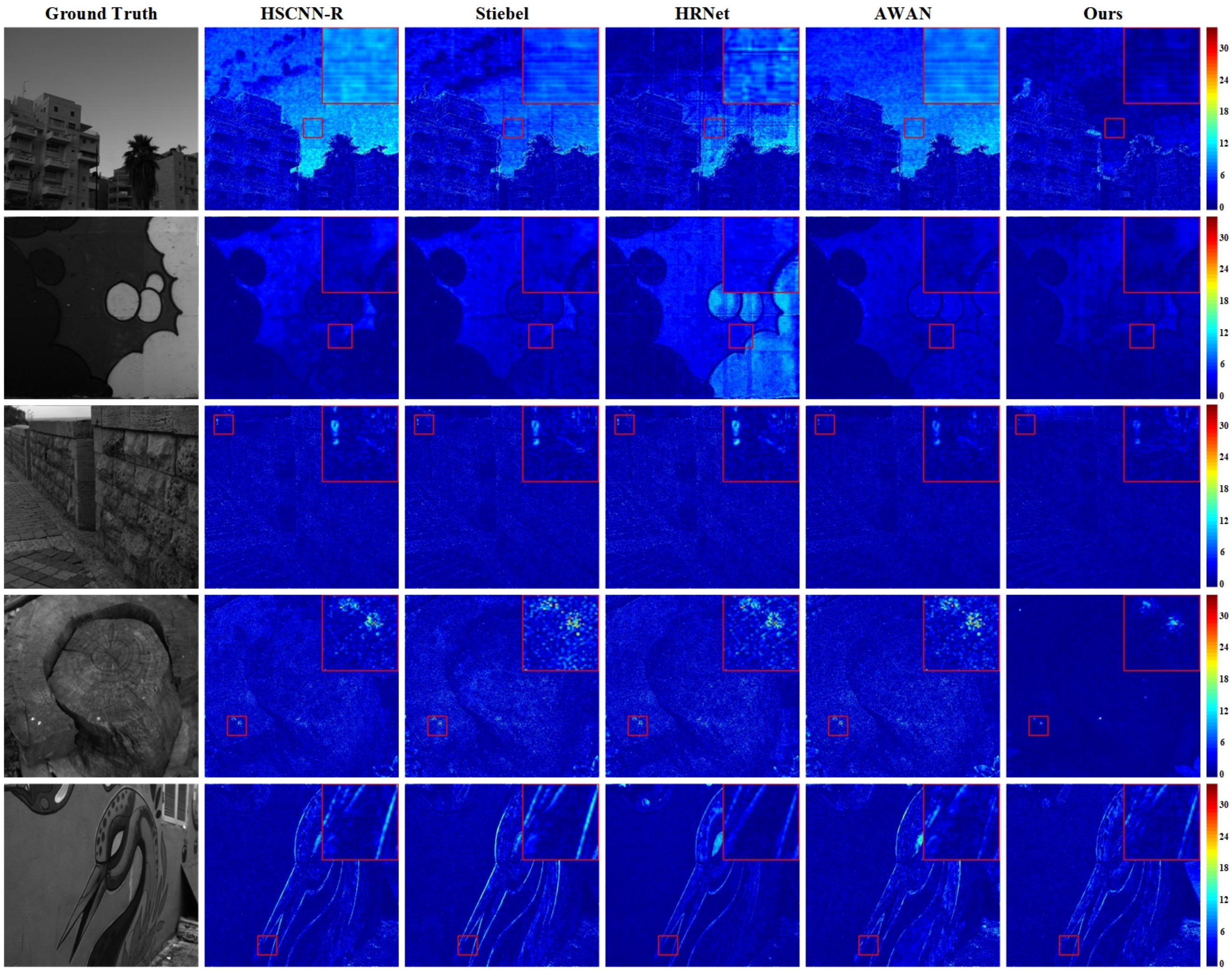

Figure 5. **Visual quality comparison of the** 18**-th band on five verification set images of NTIRE2020.** The ground truth,MRAE heat maps for HSCNN-R/Stibel/HRNet/AWAN/Ours methods. Note that the MRAE heat maps have been scaled for optimal display.

World' track in which the input RGB images quality suffer from noise and compression and the spectral response curve (SRC) prior is not allowed. With respect to NTIRE2022, we evaluate our DRCR Net on 'Spectral Recovery' track where the RGB image is first multiplied by the corresponding hyperspectral image (HSI) and the SRC, then normalized by every single image's maximum value, unified mean value to 0.18×255, added noise and compressed. Accordingly, the corresponding relationship between the pixels of different images is seriously damaged, which causes the problem of one object may have several different spectrums and different objects may also correspond to the same spectrum in the reconstruction results. Besides, the NTIRE2018 dataset consists of 256 RGB-HSI pairs for training, 5 and 10 pairs for validation and testing, separately. The spatial resolution of all images is 1392×1300. The NTIRE2020 dataset contains 450, 10 and 20 nature images for training, validation as well as testing which are 482×512 in spatial size. In the NTIRE2022 dataset, 900 pairs of training data, 50 pairs of validation data and testing data are provided respectively in which the spatial size of the HSI is consistent with the NTIRE2020 dataset. Due to the fact that the validation set of NTIRE2022 includes a HSI with some 0 value pixels which are not available for calculating the MRAE metric, another 49 HSIs are adopted. Additionally, in all three datasets, the HSI has 31 successive spectral bands ranging from 400 to 700 nanometer with a 10 nanometer spacing.

Evaluation metrics. Based on the evaluation metrics provided by the challenge, we evaluate the performance of our proposed method on the above three datasets using mean relative absolute error (MRAE) as well as root mean square error (RMSE) that squares the residuals, takes and

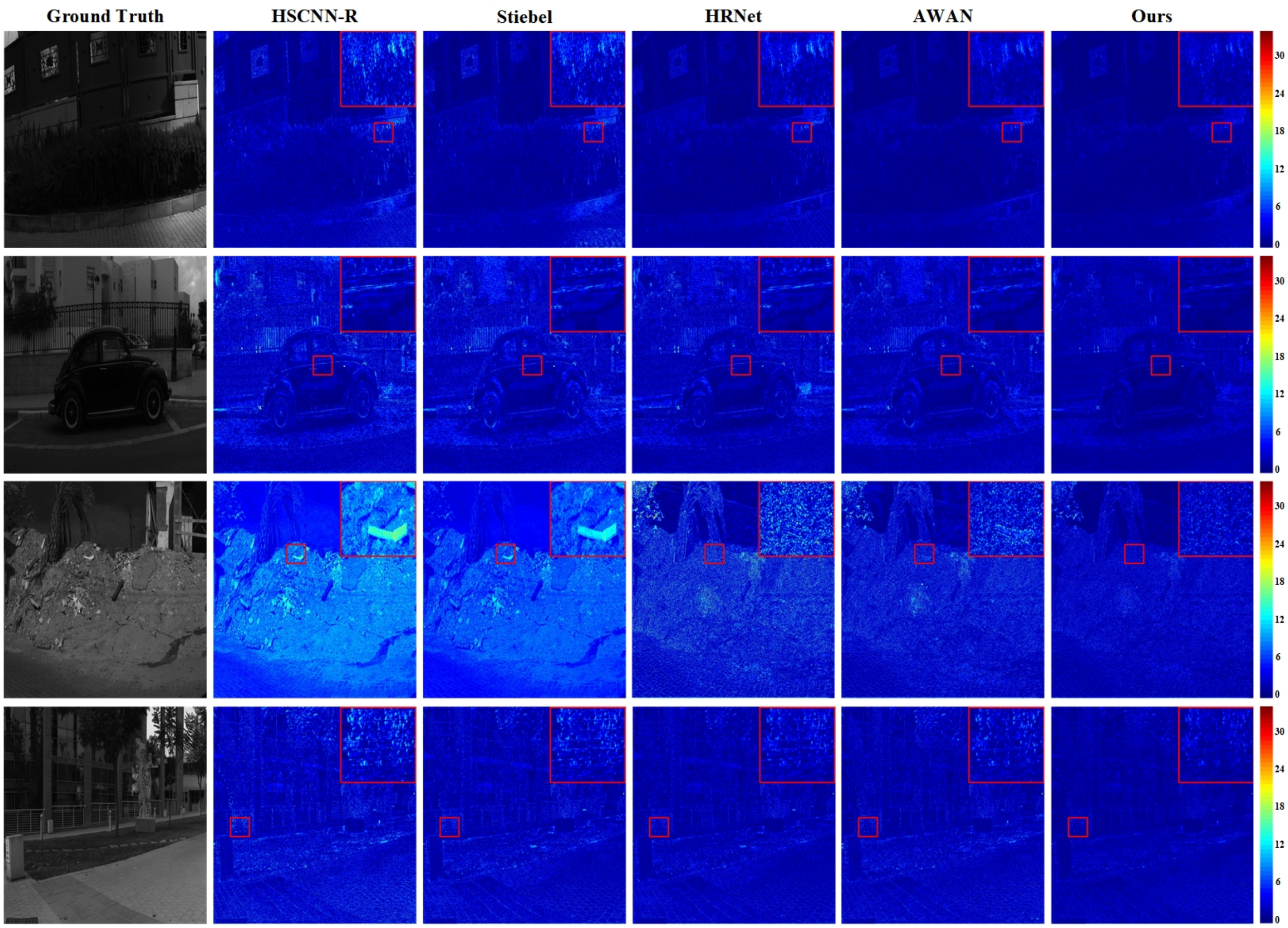

Figure 6. **Visual quality comparison of the 18-th band on five verification set images of NTIRE2018.** The ground truth,MRAE heat maps for HSCNN-R/Stibel/HRNet/AWAN/Ours methods. Note that the MRAE heat maps have been scaled for optimal display.

Table 1. Ablation study on the validation set of NTIRE 2022. We record the best MRAE and RMSE values.

	Baseline	w/o NPM	w/o CRM	ours
MRAE(↓)	0.3524	0.2157	0.3286	**0.1662**
RMSE(↓)	0.0564	0.0414	0.0519	**0.0332**

averages and finally a root of the result. These two metrics are defined as follows

$$\mathrm{MRAE} = \frac{1}{N}\sum_{n=1}^{N}\left(\left|\mathbf{I}_{GT}^{(n)} - \mathbf{I}_{SR}^{(n)}\right| / \mathbf{I}_{GT}^{(n)}\right), \quad (13)$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(\mathbf{I}_{GT}^{(n)} - \mathbf{I}_{SR}^{(n)}\right)^2}, \quad (14)$$

where $\mathbf{I}_{GT}^{(n)}$ and $\mathbf{I}_{SR}^{(n)}$ represent the value of the nth pixel in the ground truth and the reconstructed HSI, respectively.

Table 2. The final testing results of NTIRE 2022 Spectral Reconstruction from RGB Images Challenge. Our results are **highlighted**.

Team	MRAE(↓)	RMSE(↓)
THU-SIGS-MEAI	0.1131211099	0.02308144229
mialgo_ls	0.1247392967	0.02569337961
deeppf	**0.1766834871**	**0.03217000226**
Ptdoge	0.2106939205	0.03654118167
anjing_guo	0.2802988331	0.04161410992

Implementation details. For training details, we set the number of DRCR blocks to 10, and the channels of intermediate layer features to 100. The image pairs are cropped to 128×128 region before normalized to $[0, 1]$. The reduction

Table 3. The final testing results of NTIRE 2022 Spectral Reconstruction from RGB Images Challenge. The best and second best results are **highlighted** and underlined.

Method	NTIRE2018		NTIRE2020		NTIRE2022	
	MRAE(↓)	RMSE(↓)	MRAE(↓)	RMSE(↓)	MRAE(↓)	RMSE(↓)
DRCR Net	**0.0284**	**22.11**	**0.0664**	**0.0171**	**0.1662**	**0.0332**
AWAN [16]	0.0289	22.18	0.0668	0.0175	0.3135	0.0652
HR-net [35]	0.0292	22.45	0.0682	0.0179	0.3335	0.0656
HSCNN-R [25]	0.0297	22.88	0.0684	0.0182	0.3856	0.0661
Stiebel [26]	0.0312	23.88	0.0698	0.0187	0.4074	0.0691

ratio r value of the channel re-calibration module (CRM) is 8. For optimization, we choose Adam with $\beta_1 = 0.9$, $\beta_2 = 0.99$ and $\epsilon = 10^{-8}$. The learning rate is set as 0.0001 initially and a decay policy with a power of 1.5. Our DRCR Net has been implemented on the Pytorch framework and approximately 24 hours are required for training the NTIRE2022 dataset on 1 NVIDIA 3090Ti GPU.

4.2. Ablation Analysis

To verify the effects of NPM and CRM in our designed network, we carry out ablation studies on the NTIRE2022 dataset. The detailed experimental results are listed in the Table 1. Apparently, compared with all ablation settings, DRCR Net has the best performance on both MRAE and RMSE evaluation metrics. When we delete the NPM, both metrics drop significantly which verifies the necessity of our data purification process. Then, we carry out another ablation experiment that replaces the dual CRMs with direct connections. Compared with our DRCR Net, the accuracy decreases which proves the usefulness of performing re-calibration of spectral dimension features. Afterwards, we experimented with removing both modules to demonstrate their effectiveness. Obviously, the MRAE and RMSE decrease the most, which reveal the combination of them is significant for further improving the accuracy of SSR.

4.3. Results

Testing result on NTIRE 2022 challenge. In the official list of the competition, our proposed DRCR Net achieved 3rd place on the track of NTIRE 2022 Spectral Reconstruction from RGB challenge. The results on the testing dataset compared with other teams' are listed in Table 2.

Comparison with other architectures. To demonstrate the superior performance of our network, we utilize four existing SSR algorithms that have achieved SOAT results in previous NTIRE challenges, including AWAN [16], HR-net [35], HSCNN-R [25] and Stiebel [26]. Each of the above methods is based on the official dataset setup and are evaluated by the official unified evaluation indicators MRAE as well as RMSE, the numerical results of the validation set on NTIRE2018, NTIRE2020 and NTIRE2022 are summarized in Table 3. We can obviously observe that our DRDC net has a significantly leading reconstruction accuracy, especially on the 2022 validation dataset in both measurements. More precisely, our DRCR Net arquires a 46.99% decline in MRAE and a 49.08% decrease in RMSE compared with the second best AWAN method on the NTIRE2022 dataset. The reason may be the fact that our NPM can perform artifact removal and refinement of the input RGB images.

Visual results. To evaluate the perceptual quality of the reconstructed HSI, we depict the corresponding error maps of the 18th band of selected four official validation dataset images from NTIRE2022, NTIRE2020 and four images from NTIRE2018 in Fig. 4, Fig. 5 and Fig. 6, respectively. The error images are the heat maps of MRAE between the HSI recovered and the ground truth on each pixel which are shown to measure reconstruction quality. And apparently, our method achieves more accurate and robust reconstructed HSIs, and has less overall errors as can be seen from all above heat maps.

5. Conclusion

In this paper, we present a novel dense residual channel re-calibration network (DRCR Net) with non-local purification for SSR. Specifically, a non-local purification module (NPM) is proposed to perform artifact removal and self-purification through a hierarchical pyramid structure prior which can generate robust reconstruction results. Besides, a trainable dense residual channel re-calibration (DRCR) block which is cascaded with encoder-decoder paradigm through several cross-layer dense residual connections is designed for the deep feature extraction and to improve the generalization ability of the network. To further improve the recovery accuracy of DRCR Net, we develop a dual channel re-calibration Module (CRM) which can adaptively re-calibrate channel-wise feature response. Experimental results on three datasets demonstrate the superiority and effectiveness of our DRCR Net.

References

[1] Jonas Aeschbacher, Jiqing Wu, and Radu Timofte. In defense of shallow learned spectral reconstruction from rgb images. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 471–479, 2017. 3

[2] Farnaz Agahian, Seyed Ali Amirshahi, and Seyed Hossein Amirshahi. Reconstruction of reflectance spectra using weighted principal component analysis. *Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur*, 33(5):360–371, 2008. 2

[3] Michal Aharon, Michael Elad, and Alfred Bruckstein. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on signal processing*, 54(11):4311–4322, 2006. 2

[4] Naveed Akhtar, Faisal Shafait, and Ajmal Mian. Bayesian sparse representation for hyperspectral image super resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3631–3640, 2015. 2

[5] Boaz Arad and Ohad Ben-Shahar. Sparse recovery of hyperspectral signal from natural rgb images. In *European Conference on Computer Vision*, pages 19–34. Springer, 2016. 2

[6] B Arad, O Ben-Shahar, R NTIRE Timofte, L Van Gool, L Zhang, and M NTIRE Yang. challenge on spectral reconstruction from rgb images. In *Proceedings of the Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA*, pages 18–22, 2018. 2, 3, 5

[7] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 446–447, 2020. 2, 3, 5

[8] Boaz Arad, Radu Timofte, Rony Yahel, Nimrod Morag, Amir Bernat, et al. NTIRE 2022 spectral recovery challenge and dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 5

[9] Yan Bai, Jile Jiao, Wang Ce, Jun Liu, Yihang Lou, Xuetao Feng, and Ling-Yu Duan. Person30k: A dual-meta generalization network for person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2123–2132, June 2021. 2

[10] T Tony Cai and Lie Wang. Orthogonal matching pursuit for sparse signal recovery with noise. *IEEE Transactions on Information theory*, 57(7):4680–4688, 2011. 2

[11] Weisheng Dong, Fazuo Fu, Guangming Shi, Xun Cao, Jinjian Wu, Guangyu Li, and Xin Li. Hyperspectral image super-resolution via non-negative structured sparse representation. *IEEE Transactions on Image Processing*, 25(5):2337–2352, 2016. 2

[12] Ying Fu, Tao Zhang, Yinqiang Zheng, Debing Zhang, and Hua Huang. Joint camera spectral sensitivity selection and hyperspectral image recovery. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 788–804, 2018. 3

[13] Biebele Joslyn Fubara, Mohamed Sedky, and David Dyke. Rgb to spectral reconstruction via learned basis functions and weights. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 480–481, 2020. 2

[14] Jie Hu, Li Shen, Samuel Albanie, Gang Sun, and Enhua Wu. Squeeze-and-excitation networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(8):2011–2023, 2020. 4

[15] Jiaojiao Li, Songcheng Du, Rui Song, Chaoxiong Wu, Yunsong Li, and Qian Du. Hasic-net: Hybrid attentional convolutional neural network with structure information consistency for spectral super-resolution of rgb images. *IEEE Transactions on Geoscience and Remote Sensing*, 2022. 2

[16] 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 Workshops*, pages 462–463, 2020. 3, 8

[17] Jiaojiao Li, Chaoxiong Wu, Rui Song, Yunsong Li, and Weiying Xie. Residual augmented attentional u-shaped network for spectral reconstruction from rgb images. *Remote Sensing*, 13(1):115, 2020. 2

[18] Jiaojiao Li, Chaoxiong Wu, Rui Song, Yunsong Li, Weiying Xie, Lihuo He, and Xinbo Gao. Deep hybrid 2-d-3-d cnn based on dual second-order attention with camera spectral sensitivity prior for spectral super-resolution. *IEEE Transactions on Neural Networks and Learning Systems*, 2021. 2

[19] Jiaojiao Li, Chaoxiong Wu, Rui Song, Weiying Xie, Chiru Ge, Bo Li, and Yunsong Li. Hybrid 2-d–3-d deep residual attentional network with structure tensor constraints for spectral super-resolution of rgb images. *IEEE Transactions on Geoscience and Remote Sensing*, 59(3):2321–2335, 2020. 2

[20] Ke Li, Shijie Wang, Xiang Zhang, Yifan Xu, Weijian Xu, and Zhuowen Tu. Pose recognition with cascade transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1944–1953, June 2021. 2

[21] Yunsong Li, Yanzi Shi, Keyan Wang, Bobo Xi, Jiaojiao Li, and Paolo Gamba. Target detection with unconstrained linear mixture model and hierarchical denoising autoencoder in hyperspectral imagery. *IEEE Transactions on Image Processing*, 2022. 1

[22] Yi-Tun Lin and Graham D Finlayson. Physically plausible spectral reconstruction from rgb images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 532–533, 2020. 2

[23] Rang MH Nguyen, Dilip K Prasad, and Michael S Brown. Training-based spectral reconstruction from a single rgb image. In *European Conference on Computer Vision*, pages 186–201. Springer, 2014. 2

[24] Daniela Poli and Thierry Toutin. Review of developments in geometric modelling for high resolution satellite pushb-

room sensors. *The Photogrammetric Record*, 27(137):58–73, 2012. 1

[25] Zhan Shi, Chang Chen, Zhiwei Xiong, Dong Liu, and Feng Wu. Hscnn+: Advanced cnn-based hyperspectral recovery from rgb images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 939–947, 2018. 3, 8

[26] Tarek Stiebel, Simon Koppers, Philipp Seltsam, and Dorit Merhof. Reconstructing spectral images from rgb-images using a convolutional neural network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 948–953, 2018. 8

[27] Bohan Wu, Suraj Nair, Roberto Martin-Martin, Li Fei-Fei, and Chelsea Finn. Greedy hierarchical variational autoencoders for large-scale video prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2318–2328, June 2021. 2

[28] Chaoxiong Wu, Jiaojiao Li, Rui Song, Yunsong Li, and Qian Du. Hprn: Holistic prior-embedded relation network for spectral super-resolution. *arXiv preprint arXiv:2112.14608*, 2021. 3

[29] Bobo Xi, Jiaojiao Li, Yunsong Li, Rui Song, Weiwei Sun, and Qian Du. Corrections to "multiscale context-aware ensemble deep kelm for efficient hyperspectral image classification". *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–1, 2022. 1

[30] Bobo Xi, Jiaojiao Li, Yunsong Li, Rui Song, Yuchao Xiao, Qian Du, and Jocelyn Chanussot. Semisupervised cross-scale graph prototypical network for hyperspectral image classification. *IEEE Transactions on Neural Networks and Learning Systems*, 2022. 1

[31] Zhiwei Xiong, Zhan Shi, Huiqun Li, Lizhi Wang, Dong Liu, and Feng Wu. Hscnn: Cnn-based hyperspectral image recovery from spectrally undersampled projections. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 518–525, 2017. 3

[32] Jize Xue, Yong-Qiang Zhao, Yuanyang Bu, Wenzhi Liao, Jonathan Cheung-Wai Chan, and Wilfried Philips. Spatial-spectral structured sparse low-rank representation for hyperspectral image super-resolution. *IEEE Transactions on Image Processing*, 30:3084–3097, 2021. 2

[33] Xianyi Zhang and Haitao Zhao. Hyperspectral-cube-based mobile face recognition: A comprehensive review. *Information Fusion*, 74:132–150, 2021. 1

[34] Zhongyang Zhang, Zhiyang Xu, Zia Ahmed, Asif Salekin, and Tauhidur Rahman. Hyperspectral image super-resolution in arbitrary input-output band settings. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 749–759, 2022. 3

[35] Yuzhi Zhao, Lai-Man Po, Qiong Yan, Wei Liu, and Tingyu Lin. Hierarchical regression network for spectral reconstruction from rgb images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 422–423, 2020. 3, 8

[36] Zhiyu Zhu, Hui Liu, Junhui Hou, Sen Jia, and Qingfu Zhang. Deep amended gradient descent for efficient spectral reconstruction from single rgb images. *IEEE Transactions on Computational Imaging*, 7:1176–1188, 2021. 2