NTIRE 2022 Spectral Demosaicing Challenge and Data Set

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 WenFang Sun Jiesi Zheng Bowen Zhao
Yanpeng Cao Jiangxin Yang Yanlong Cao Xiangyu Kong Jingbo Yu
Yuanyang Xue Zheng Xie

Abstract

This paper presents the first challenge on demosaicing of natural spectral images for snapshot hyperspectral imaging systems (HIS) which utilize a multi-spectral filer array (MSFA), i.e.*, the recovery of whole-scene hyperspectral information from spatially sub-sampled hyperspectral information. This challenge expands the "ARAD_1K" data set to a first-of-its-kind large-scale data set for multi-spectral filter array demosaicing of natural scenes containing 1,000 images. Challenge participants were required to recover hyperspectral information from synthetically generated MSFA images simulating capture by a known calibrated snapshot mosaic hyperspectral camera. The challenge was attended by 157 teams, with 29 teams competing in the final testing phase, 7 of which provided detailed descriptions of their methodology which are included in this report. The performance of these submissions is reviewed and provided here as a gauge for the current state-of-the-art in multi-spectral filter array demosaicing of natural images.*

1. Introduction

Hyperspectral imaging systems (HIS) are able to record the distribution of light in a scene across a large number of narrow spectral bands [7]. The additional information which HISs can provide over conventional RGB cameras could offset the disadvantages of their larger size, significantly higher cost, and limited resolution. However, a strong gating factor for the use of traditional HISs in many computer vision applications is their long acquisition times due to the use of spatial or temporal scanning. Snapshot HISs overcome this limitation by rapidly acquiring both spectral and spatial information [17].

Among recent approaches to snapshot hyperspectral imaging, such as computed tomography [9,22] or light-field imaging [5,8], snapshot mosaic HISs or "multi-spectral filter array (MSFA) cameras" are emerging as a leading contender. MSFA cameras are snapshot HISs which employ a MSFA to rapidly acquire spectral information in a single exposure of a 2D image sensor [25]. MSFAs cameras spatially sub-sample the imaged scene in a similar manner to Bayer-filter-based RGB cameras. While Bayer filter RGB cameras employ a repeating 2×2 color filter array (CFA) or "mosaic", MSFA cameras will often utilize much larger 3×3, 4×4, or even 5×5 [18] CFAs. Figure 1 depicts a conventional RGB CFA compared to a 4×4 MSFA.

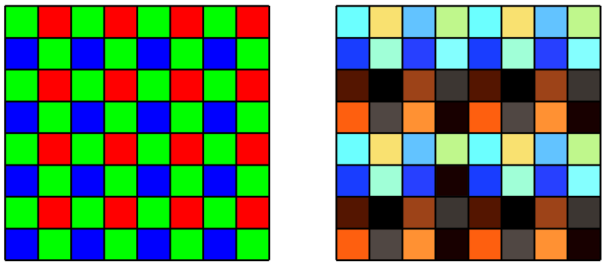


Figure 1. Conventional "GRBG" Bayer pattern CFA (left) compared to the 4×4 MSFA described in (right). Filter colors are roughly correlated to the perceived "color" of each filter to a human observer. Filter configuration for the depicted MSFA is described in Figure 2.

A growing variety of commercial MSFA cameras are becoming available to researchers and industry professionals at increasingly lower costs. Such systems include the IMEC SNAPSHOT spectral camera series [24, 46], the XIMEA Snapshot USB3 camera series [26], the silios CMS camera series [29] and others. A major advantage of MSFA cameras is that they can be implemented in a similar form-factor and operated in a similar manner to conventional RGB cam-

eras. However, to fully utilize the spatial and spectral information recorded by MSFA cameras, efficient and accurate spectral demosaicing methods are required.

The challenge of such demosaicing large MSFAs is twofold: a larger mosaic pattern provides a significantly stronger sub-sampling of the imaged scene and the narrow bands found in most MSFAs have weaker inter-channel correlation than the R-G-B channels of a Bayer filter camera. Previously proposed methods for spectral demosaicing include expansions of long-established interpolation-based methodologies [15, 32], matrix-factorization/recovery-based methods [1, 44], and most recently deep learning approaches [11, 16, 36, 38, 43]. While previously present methods for spectral demosaicing vary in their methodologies, a commonality they share is training and testing over very small data sets - with recent works training over less than 100 images and testing over less than 10 images.

Previously reported methodologies are difficult to compare as they may differ in their target system (MSFA cameras with a different amount of channels and/or different filter configurations) or in their selection of test images even if the latter are drawn from the same data set. To facilitate equal grounds comparison of current and future state-of-the-art multi-spectral filter array demosaicing methods this challenge presents: a large-scale data set of 1,000 natural hyperspectral scenes, a single target 4×4 MSFA camera, and a uniform testing procedure.

This challenge is one of the NTIRE 2022 associated challenges: spectral recovery [4], spectral demosaicing [3], perceptual image quality assessment [13], inpainting [39], night photography rendering [10], efficient super-resolution [30], learning the super-resolution space [31], super-resolution and quality enhancement of compressed video [49], high dynamic range [37], stereo super-resolution [47], burst super-resolution [6].

2. Data Set

Name	Scenes	Spatial Resolution	Spectral Resolution
CAVE [50]	32	512×512	31 bands (400-700nm)
TokyoTech [33]	30	2048×2048	31 bands (420-720nm)
TT59 [34]	40	2048×2048	59 bands (420-1000nm)
Hytexila [20]	112	1024×1024	186 bands (400-1000nm)
ICVL [2]	201	1392×1300	**519 bands (400-1000nm)**
ARAD_1K (primary) [4]	**1,000**	480×512	31 bands (400-700nm)
ARAD_1K (16 band)	**1,000**	480×512	16 bands (400-1000nm)

Table 1. Summary of existing data sets of natural hyperspectral images used to train/evaluate methods for MSFA demosaicing compared to the ARAD_1K data set.

The development and testing of MSFA demosaicing methods requires ground truth hyperspectral information which is difficult to obtain [20]. For this reason, previous works [1] have either used data acquired by airborne hyperspectral platforms, such as the NASA AVIRIS [45], or relied on much smaller data sets of natural hyperspectral images. Table 1 summarizes existing data sets of natural hyperspectral images previously used [11, 14, 16, 38, 43] to train/evaluate methods for MSFA demosaicing and compares them the the ARAD_1K data set presented here.

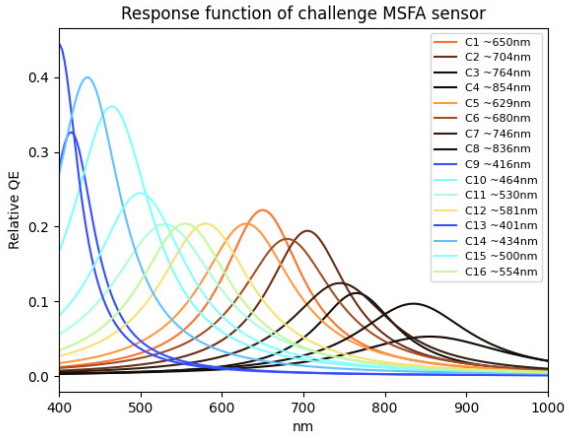


Figure 2. Response function of a prototype MSFA camera sensor provided by one of ODDITY's commercial partners. The sensor provides 16 channels of spectral information over the 400-1000nm range, spatially sub-sampled using a 4×4 multi-spectral filter array. The approximate peaks of each channel are denoted in the legend. Line colors are roughly correlated to the perceived "color" of each filter to a human observer.

This challenge expands on the ARAD_1K spectral image data set presented as part of the NTIRE 2022 Spectral Recovery challenge [4]. While the former provides 31 band hyperspectral images in the 400-700nm range, this challenge presents 16 channel hyperspectral images in the 400-1000nm range - covering a wider range of wavelengths, but at a reduced spectral resolution. Figure 3 depicts a set of sample images from the ARAD_1K data set.

The MSFA demosaic expansion of the ARAD_1K provides the same 1,000 scenes presented in the ARAD_1K data set, divided similarly to 900 training images, 50 validation images, and 50 confidential test images. Ground truth hyperspectral information is provided as 480×512 spatial resolution images across the 16 spectral bands depicted in Figure 2.

Additional information regarding the data set, its relation to the previously published ARAD data set, instructions for data access, and relevant code is available at the following GitHub repository: https://github.com/boazarad/ARAD_1K

Figure 3. Sample images from the ARAD_1K hyperspectral image data set. Note the variety of settings and viewpoints (images modified for optimal display).

2.1. MSFA Camera Simulation

As a first-of-its-kind large-scale evaluation of MSFA demosaicing methods, this challenge aims to establish an initial baseline for the spectral demosaicing task, without adding compounding factors such as sensor acquisition noise. To this end, the camera simulation pipeline makes the following assumptions:

1. The camera's spectral response function is known.
2. The camera determines its exposure settings automatically - the exposure algorithm is known, but parameters used to compute it for each scene are not (e.g. average scene brightness).
3. The camera has zero noise.
4. No post-processing is applied to acquired images except for highlight clipping.
5. Images are saved in uncompressed "RAW".

Participants were provided with training images produced by the challenge MSFA camera simulation pipeline, camera simulation pipeline code, and the MSFA camera response function used in the simulation. Figure 2 depicts the response function used for MSFA camera simulation in this challenge. Pipeline code and the MSFA camera response function were provided to participants.

3. Challenge

The NTIRE 2022 Spectral Demosaic Challenge was presented as a competition on the CodaLab [1] platform which consisted of two phases:

1. **Development** participants were provided with 900 training and 50 validation RGB images generated by the camera simulation pipeline (c.f. Sec. 2.1). Corresponding ground truth hyperspectral images were provided for the 900 training images. A test server was made available where participants could upload recovered spectral information for the 50 validation images and receive immediate feedback on their performance in terms of PSNR and SAM per-image (c.f. Sec. 3.1). During the development phase, there were no limits on the amount of submission per team.
2. **Testing** Ground truth hyperspectral images for the 50 validation images were released, alongside 50 test RGB images. Similarly to the development phase, a test server was made available where participants could upload their results and receive feedback on their performance, but each team was limited to a total of three submissions. This feedback allowed participants to select their best model, while limiting the possibility of overfitting to the test set.

Code and other data provided to participants is curated in the following GitHub repository:https://github.com/boazarad/NTIRE2022_spectral

3.1. Evaluation Metrics

For this challenge, Peak signal-to-noise ratio (PSNR) computed between the submitted reconstruction results and the ground truth images was selected as the primary quantitative measure. Spectral Angle Mapper (SAM) [23] was reported as well, but not used to rank results. PSNR is defined as:

$$PSNR = 10 \cdot log_{10}\left(\frac{PEAK^2}{MSE}\right) \tag{1}$$

where $PEAK$ denotes the maximum possible pixel value to the image (For the ARAD_1K data set $PEAK = 1$) and MSE and and SAM are defined as:

$$MSE = \frac{1}{|P_{gt}|}\sum_{i,c}\left(P_{gt_{i,c}} - P_{rec_{i,c}}\right)^2 \tag{2}$$

$$SAM = \frac{1}{|P_{gt}|}\sum_{i} cos^{-1}\left(\frac{\sum_c P_{gt_{i,c}} P_{rec_{i,c}}}{\sqrt{\sum_c P^2_{gt_{i,c}}}\sqrt{\sum_c P^2_{rec_{i,c}}}}\right) \tag{3}$$

[1] https://codalab.lisn.upsaclay.fr/competitions/722

Where $P_{gt_{i_c}}$ and $P_{rec_{i_c}}$ denote the value of the c spectral channel of the i-th pixel in the ground truth and the reconstructed image, respectively, and $|P_{gt}|$ is the size of the ground truth image (pixel count $\times$ number of spectral channels).

3.2. Evaluation Protocol

Similarly to the NTIRE 2022 Spectral recovery challenge [4] a 50 image test set, including a large variety of images from multiple settings (c.f. Sec. 2) were provided for evaluation. Due to space and bandwidth constraints limitations of the CodaLab platform evaluation was performed over a cropped central region of the test images. Participants were provided with code to prepare images for evaluation over a central 226×256 region cropped from the original 480×512 spatial resolution. Final results were scored for PSNR and SAM over the selected central region of the test images.

4. Challenge Results

Table 2 details the final rankings of all participants over the primary evaluation metrics. The highest PSNR achieved was 47.74 and the lowest SAM achieved was 0.00973. SAM rankings differed significantly from PSNR rankings. The top performing method in terms of inference time was able to achieve $\sim$ 71 FPS which could be considered "real-time" performance, though this frame-rate would only be possible with an advanced GPU (NVIDIA RTX 2080Ti) and for $\sim$ 0.25MP images. The high frame-rate solution also comes at a cost of a $\sim$ 4.3db PSNR loss. Section 6 describes the methodologies used by top-performing teams in this challenge, as described by their authors.

4.1. Performance on "Out-of-Scope" Image

Similarly to the NTIRE 2022 Spectral recovery challenge [4], finalists were presented with an "out-of-scope" image to recover. Figure 4 depicts the out-of-scope image selected for this challenge: it features a prominent human subjects and calibration target - objects which are very rare in the training data set. Furthermore, the image was taken under photographic studio lights, while the majority of training images were captured under natural illumination or conventional indoor lighting. While performance on the out-of-scope image may be indicative of a methods extrapolation power, these measurements did not affect participants final ranking in the challenge. Table 4 details the performance of most submitted methods over the out-of-scope image.

The out-of-scope image contains a large amount of relatively uniform surfaces. Spatially uniform surfaces should present an easier target for demosaicing, a even naive interpolation should produce good results over uniform areas. It is therefore surprising to see degradation in PSNR performance of up to 10db for some of the top performing methodologies. Conversely, other methodologies saw PSNR gains of almost 10db. This variability in performance indicates that both the methodologies, as well as the train/test data set used in this challenge have significant room for improvement.

Figure 4. "Out-of-scope" image used to gauge the extrapolation ability of methods presented in this challenge. This image contains a prominent human subject, a calibration target, and was taken under studio lighting (image modified for optimal display).

5. Conclusion

The NTIRE 2022 Spectral Demosaic Challenge presents a first-of-its-kind large-scale evaluation of MSFA demosaicing methods. A larger-than-ever natural hyperspectral image data set for both training and evaluation of MSFA demosaicing methods is presented. Top performing methodologies, selected from a total of 157 participating teams are presented as a baseline for future evaluations and/or challenges.

Top performing methodologies were all bases on neural nets and require high-end GPUs for inference. While the 4th ranked method presented performance that could be considered real-time for low-resolution images ($\sim$ 0.2mp), real time performance on modern $>$ 2mp snapshot mosaic HISs seems unattainable for edge devices or even with a high-end GPU.

The high PSNR values achieved by top performing methods provide motivation to explore a more realistic camera simulation, which includes sensor noise, in future challenges. It is our hope that this data set and the method-

Rank	Team	Username	PSNR	SAM
1	HITZST01	Chen01	47.74	$0.01114_{(3)}$
2	MIALGO	mialgo_ls	46.89	$0.01090_{(2)}$
3	IFL	Ptdoge	45.39	$0.00973_{(1)}$
4	NPUMPI	fengkainpu	43.39	$0.01427_{(6)}$
5	SIP	xleft	42.81	$0.01323_{(4)}$
6	ZJU231	ZJU231	41.31	$0.03204_{(7)}$
7	OnRoad (SRC-B)	benfen	41.07	$0.01383_{(5)}$

Table 2. NTIRE 2022 Spectral Reconstruction Challenge results and final rankings on the ARAD_1K HS test data. Secondary (SAM) ranks are denoted in parenthesis.

Team Name	CPU	GPU	Platform	Train Time	Inference Time
HITZST01	Intel(R) Xeon(R) CPU E5-2697A v4 @ 2.60GHz	NVIDIA TITAN Xp	PyTorch	96 Hours	4s
MIALGO	Intel(R) Xeon(R) Gold 6240 CPU @ 2.60GHz * 2	8 x NVIDIA Tesla V100 32GB	PyTorch	5 Days	0.43s
IFL	Intel(R) Xeon(R) CPU E5-2680 v4 @ 2.40GHz	NVIDIA RTX3090	PyTorch	40 Hours	0.057
NPUMPI	Intel(R) Xeon(R) Gold 6240 CPU @ 2.60GHz	NVIDIA RTX 2080Ti	PyTorch	17 Hours	**0.014s**
SIP		NVIDIA GeForceRTX 3090	PyTorch		0.61s
ZJU231	Intel(R) Xeon(R) Gold 5218 CPU @ 2.30GHz	NVIDIA Tesla V100	PyTorch	52 Hours	0.45s
OnRoad (SRC-B)		NVIDIA A100-40GB	PyTorch		0.43s

Table 3. Self-reported training and inference runtimes for proposed methods.

Rank	Team	PSNR	SAM
1	$IFL_{(3)}$	55.06	0.005457
2	$MIALGO_{(2)}$	52.92	0.006016
3	$OnRoad\ (SRC\text{-}B)_{(7)}$	48.84	0.006548
4	$HITZST01_{(1)}$	46.71	0.007060
5	$ZJU231_{(6)}$	33.24	0.027480
6	$NPUMPI_{(4)}$	33.06	0.014158

Table 4. Performance of proposed methodologies for "out-of-scope" image, ranking on the primary test set is denoted in subscript beside the team name.

ologies described here will facilitate future improvement in MSFA demosaicing.

6. Methods and Teams

6.1. HITZST01: Domain Adapted Multi-scale Channel-attention Network (DAMCNet) for multi-spectral demosaicing.

We present the DAMCNet for multi-spectral demosaicing from 16 channel mosaic, that is, the task of restoration of 16 channel multi-spectral images (label) from simulated "RAW" 4×4 MSFA input (input). Figure 5 describes the overall architecture of our proposed network. We divide the demosaicing task into two sub-tasks, which are source domain adaptation stage and target domain demosaicing stage, respectively. In the first stage, we convert MSFA inputs from the source domain to the target domain to compensate the quantified loss and the numerical distribution shifts caused by ISP operations. In the second stage, we propose the MCNet which utilizes the cross-channel spatial and spectral information globally and locally to reconstruct the full-channel multi-spectral images. It is worth noting that our DAMCNet is of strong nonlinear modeling ability, which can be regarded as a general model for different demosaicing missions with different raw inputs. As shown in Figure 5, the network (DAMCNet) is composed of four sub-modules: domain adaptation, source encoder module, feature refinement module and final prediction module.

For the second stage, **Source Encoder Module** consists of two convolution layers. Its role is to reconstruct image texture information from the MSFA as the input (I_{raw}) of feature refinement module. **Feature Refinement Module** takes I_{encode} as input to fully select, refine and enhance useful information in raw domain. It is formed by sequential connection of MCCA [21] [51] blocks of different scale. After getting a great representation I_{refine} of the raw image from the feature refinement network, **Final Prediction Module** takes the I_{refine} as input and perform demosaicing to reconstruct a full-channel multi-spectral image.

Our loss function is defined as:

$$loss = \|I_R - I_G\|_1 \tag{4}$$

Here I_R and I_G present the reconstructed image from mosaic image and corresponding ground truth image.

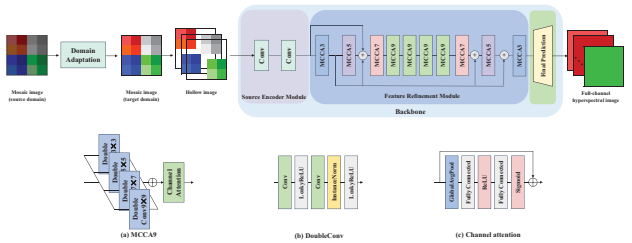


Figure 5. The DAMCNet for multi-spectral demosaicing network architecture.

- **Training**

  The training details of the demosaicing network are as follows: model is implemented in Pytorch and runs on 8 Nvidia Titan Xp graphical processing units (GPU). The model is optimized with an Adam optimizer as $\beta_1 = 0.9$, $\beta_2 = 0.99$ and learning rate = 1e-4 with a batch size of 48.

- **Testing**

  During the testing phase, we first predict the brightness offset by domain adaptation. Then we input the mosaic image with restored brightness into the trained model to realize the restoration of multi-spectral image.

- **Overexposure Correction**

  Due to the loss of information in the overexposed area, the quantization loss of 12-bit data, and the scene brightness shifting caused by camera's auto exposure in the source domain data, it is difficult to recover the multi-spectral image directly from the source domain data. The data in the target domain is easy to establish a mapping relationship to the multi-spectral image due to the following advantages: a) the average brightness is consistent with ground truth; b) the pixel value distribution of MSFA is consistent with ground truth. Considering these situations, we adopt domain adaptation to transfer mosaic image from the source domain to the target domain, thus realizing the brightness correction of mosaic image.

- **Data augmentation**

  Because of the specific pattern of MSFA, some data augmentation methods such as flipping and rotation cannot be directly performed, otherwise the arrangement of MSFA may be affected. For this dilemma, we directly sample mosaic image from 16 channel multi-spectral image which greatly increases the number of training sets. Random geometric transformation (such as translation, flip, rotation) is performed so as to obtain the transformed multi-spectral image patch (label), and then sample it based on MSFA to obtain mosaic image patch (input). Through this process, on the one hand, the input that strictly satisfies the MSFA pattern can be obtained, and on the other hand, the training data can be greatly expanded.

- **Data preprocessing**

  Considering the special pattern of the 4×4 MSFA used in this challenge, we preprocess the input mosaic image, that is, the adjacent pixels with similar wavelength ranges are processed into one channel, so as to convert the input mosaic image from single channel to 3 channels, which is inspired by quadbayer color filter array (quadbayer CFA) and has been proved to make lower crosstalk between different color channels.

6.2. MIALGO: Enhanced Holistic Attention Network for Spectral Reconstruction

Spectral reconstruction, as a typical reconstruction task is highly similar to the image super-resolution. We utilize Holistic Attention Network(HAN [35]), a SOTA method in the super-resolution tasks as the backbone to solve it. To be specific, we notice that the brightness(mean) of the input images is set to a fixed value(typical scene reflectivity, 0.18), it is particularly important to estimate the brightness of the target, and thus we divide the RGB/Mosaic image into two cases based on the maximum value. Followings are detailed explanations for the two cases:

1. The maximum value is **less than** the upper limit (255 for rgb or 4095 for mosaic). In this case, the maximum value of the input corresponds to the maximum value of GT (ignoring the effects of mosaic processing and quantization errors), so we add a simple normalization layer before the backbone, after that, the brightness of the image is basically same with GT. This case is relatively simple, and the network can handle it well.
2. The maximum value of the input is **equal** the upper limit. In this case, the clip operation during the generation of input causes a lot of energy loss, so the brightness cannot be estimated by referring to the maximum value like case 1. To deal with this ill-conditioned and difficult problem, we use a lot of augmented data for training.

We also remove the upsampling layer of HAN to keep the size of the input, and add a normalization layer after the backbone to avoid the loss caused by the clip operation.

Figure 6 describes the high-levle architecture of the solution.

- **Total Method Complexity**

  the total number of GMACS is 1,822, and the total number of parameters is 7,457,168.

Figure 6. Architecture of the Enhanced Holistic Attention Network for Spectral Reconstruction.

- **Additional Training Data**

  We found that the bottleneck of the task is the brightness estimation, that is, the richness of the data, so we tried to use a lot of additional data, including ICLV citearad2016sparse, CAVE [50] and Harvardčitechakrabarti2011statistics.

- **Training**

  According to the code provided by the organizer, we generate and augment the input data ourselves, including random brightness, random noise, random padding, flip, rotation, etc. We first train on all the data for 100k iterations, and then train separately on each case's data for 100k iterations. In the later stages of training, we increase the proportion of hard samples. L1 and SSIM were used as training loss and in late training, we keep only the luminance component of SSIM.

- **Testing**

  The model is switched according to the maximum value of the input, which corresponds to the two cases in the training phase.

6.3. IFL: Non-Local Residual Attention Network (NLRAN)

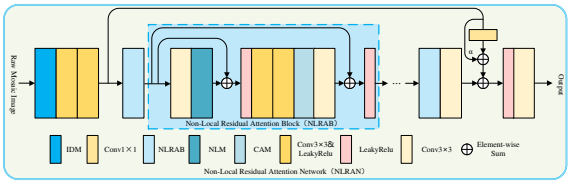


Figure 7. Overall Architecture of NLRAN.

In this challenge, we propose a non-local residual attention network (NLRAN) for multi-spectral filter array demosaicing. Fig. 7 shows the detail architecture of NLRAN. Non-local residual attention block (NLRAB) is the basic unit of NLRAN, which includes two key components, the non-local module (NLM) and channel attention module (CAM). Note that the structure of NLRAB benefits from [27]. NLM and CAM [19] are introduced to capture spatial long-range similarity and channel interdependence within intermediate features respectively. Concretely, as shown in Fig. 8, NLM adopts classic non-local operations. To reduce computational burden, different from [48], NLM regards one patch as one pixel, which making it possible to embed non-local operations in each basic unit of network in the case of limited memory resources. Besides, we design a initial demosaicing module (IDM), which is based on neighborhood spatial similarity, spectral correlation and periodic repeat of mosaic image. As shown in Fig. 9, IDM restores missing spectral bands at the current position by weighted fusion neighborhood information simply, and weights of different positions in the filter array are different. The neighborhood size designed in IDM is 3×3. In addition to the above well-designed network, we explore the data processing strategy seriously. According to the public code and resource provided by this challenge, we propose the data pre-processing strategy shown in Fig. 10. The data used to supervised the output of our network is not normalized. Experiments prove that the strategy is effective in this challenge. Besides, L1loss and mean relative absolute error (MARE) loss [42] are used to train our model respectively.

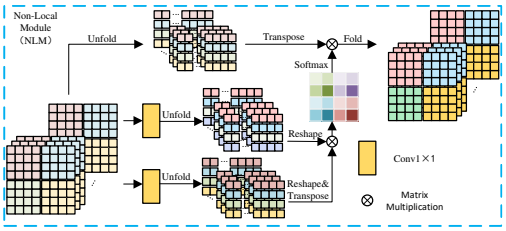


Figure 8. Diagram of Non-local Module.

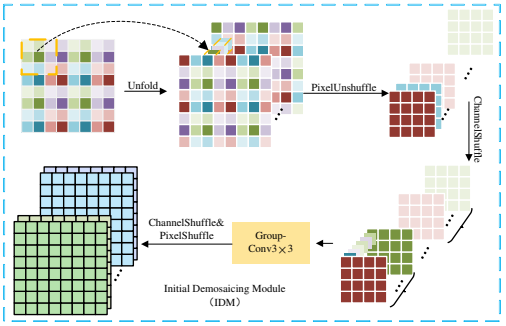


Figure 9. Diagram of Initial Demosaicing Module

- **Training**

  During the development phase, we train the proposed model for five rounds with different settings. Those models obtained are evaluated when validation data is public. Table 5 shows the performance, parameters setting, and training strategy of all models retained. In addition, during the development phase, the batch size of our model is set to 32 and 16 for models ensemble,

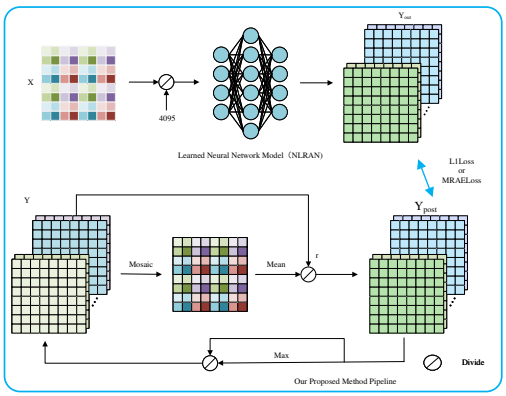


Figure 10. The Proposed NLRAN Pipeline.

and the Adam optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$ is adopted. 64 × 64 RGB-HSI pairs are cropped with a stride of 32 from the original data set for training [27]. The learning rate is initialized to 0.0001 and the linear function is set as the decay strategy. Then PyTorch framework is used to realize proposed models. The optimization of models is implemented on NVIDIA GPU.

- **Testing**

  During the testing phase, the whole mosaic image is input into our trained model. Note that the corresponding output needs to be normalized by dividing its maximum value. The output normalized, i.e. the reconstructed multi-sepctral iamge, is submitted on the codalab platform.

Model Name	Batch Size	Number of NLRAB	Number of Feature Channle	Loss Function	PSNR on Cropped Image
BEST_v1	32	10	128	L1	44.30487
BEST_v2	32	10	128	MRAE	46.210949
BEST_v3	16	10	128	L1	46.103226
BEST_v4	16	10	128	MRAE	46.274715
BEST_v5	32	8	176	MRAE	45.946461
Ensemble	-	-	-	-	46.529556

Table 5. Performance of individual models and model ensemble over the challenge's validation data.

- **Ensembles and fusion strategies**

  Five models trained with different parameters setting are used to restore multi-spectral image from the filter array. Then all results are weighted averaged as the final result. Compared with the single model, as shown in Tab. 5, model-ensemble strategy can further improve the accuracy of demosaicing.

6.4. NPUMPI: Deep Joint Multispectral Demosaicing and Anti-clipping (DJMDA)

We divide this challenge into spectral demosaicing and spectral anti-clipping and remake the ground truth spectral cubes. We think there is a difference between clipped and non-clipped data. Therefore, we train two mosaic convolution-attention models [11] for clipped and non-clipped data respectively. Finally, we normalize the maximum value of the cube. Our training and evaluation pipeline is as shown in 11.

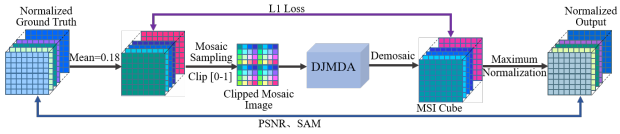


Figure 11. A schematic diagram of our training and evaluation pipeline.

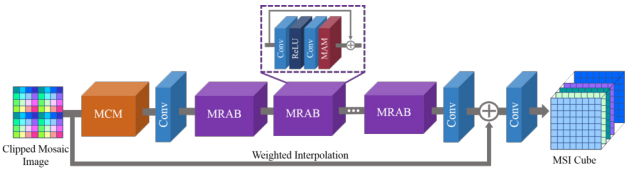


Figure 12. A schematic diagram of the Deep Joint Multispectral Demosaicing and Anti-clipping (DJMDA).

We think there is a difference between clipped and non-clipped data, and the non-clipped scenes in the data set is small. If they are directly mixed together for training, it is not good for the non-clipped scenes. Therefore, we first train a model using all the data for demosaicing clipped scenes. Then, we take the model trained by all data as the pre-trained model and fine-tune it on the non-clipped scenes. In order to expand the amount of no-clipped data, we still feed all kinds of data to the network but only calulate $L1$ loss on no-clipped regions of clipped cubes under the guidance of mask M:

$$L_{non-clip} = L_1(Out, GT) \odot M \tag{5}$$

where M is a binary mask that indicates non-clipped areas computed on the coarse weighted interpolation results of raw mosaic images.
We use the MCAN [11] as our base network architecture, increase the number of mosaic residual attention blocks (MRABs) to 10 and use a convolution layer to fuse the residual blocks output and weighted interpolation results, as shown in Fig. 12. The detailed implementation can be viewed in [11].

- **Training**

  For the pre-processing of the training data, we first tune the average value of fully-defined ground-truth cube to 0.18. Then we randomly spatially flip, rotate, and crop these cubes to generate the corresponding spectral mosaic images as inputs. Finally we use

32 as our training batch size and 128×128 as our training patch size.
For the training setup, we use $L1$ loss function and Adam optimizer.

- **Testing**

  We use the full spectral mosaic image provided as input of network. Then we normalize the maximum value of the cube outputted by the network. Finally, we compute PNSR and SAM between the normalized cube and the ground truth cube.

6.5. SIP - Spectral Image Processing: Multi-Spectral Filter Array Demosaicing based on Res2-Unet

Based on the spatial and spectral correlation existing in multi-spectral filter array mosaic images, we propose a deep neural network based on Res2-Unet for Multi-Spectral Filter Array Demosaicing. As shown in Figure 13, the whole network is based on Unet [40] framework, combined with band separation and 7×7convolution, used Res2net-SE module [12] [19] to construct backbone network. PixelShuffle and PixelUnShuffle [41] are used to connect network layers.

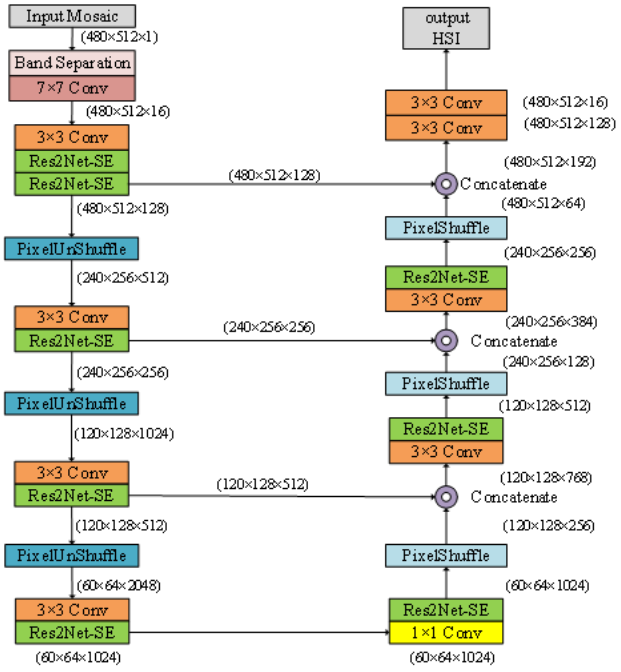


Figure 13. Network architecture of Res2-Unet.

(1)Band separation

Specifically, as shown in Figure 14, sixteen sampling matrices is used to separate the band of the input image from 480×512×1 to 480×512×16.
(2)Weighted bilinear interpolation reconstruction

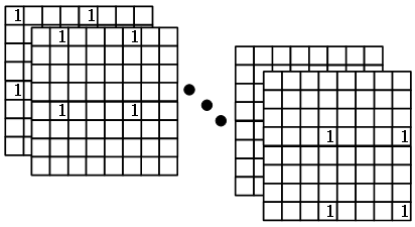


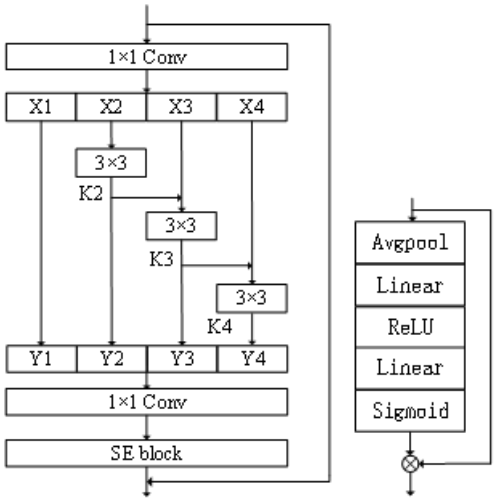
Figure 14. Sixteen sampling matrices.

7×7 convolution filter is proposed as shown in Equation 6 to carry out preliminary hyperspectral image reconstruction.

$$F = 1\frac{1}{6}\begin{bmatrix} 1 & 2 & 3 & 4 & 3 & 2 & 1 \\ 2 & 4 & 6 & 8 & 6 & 4 & 2 \\ 3 & 6 & 9 & 12 & 9 & 6 & 3 \\ 4 & 8 & 12 & 16 & 12 & 8 & 4 \\ 3 & 6 & 9 & 12 & 9 & 6 & 3 \\ 2 & 4 & 6 & 8 & 6 & 4 & 2 \\ 1 & 2 & 3 & 4 & 3 & 2 & 1 \end{bmatrix} \tag{6}$$

(3)Network reconstruction

As shown in Figure 15, feature extraction is carried out by combining 3x3 convolution with Res2net-SE module. The Res2net-SE module has the character of residual connection and multi-scale feature fusion, which can extract images' local and global features at a finer granularity level. The SE module with channel attention mechanism is added at the end of the module, which can adaptively adjust channel feature response and protect important channel features.



(a) Res2Net-SE module (b) SE block

Figure 15. Res2Net-SE module and SE block.

Define the loss in the training process as:

$$Loss = E[||G(x) - y)||_1] + 0.01 * MRAE \tag{7}$$

$$MRAE = E[|G(x) - y|\frac{y}{+}10^{-6}] \quad (8)$$

Where, x is the input Mosaic image, y reference hyperspectral image and $G()$ is the reconstruction network Res2-Unet proposed above.

- **Training**

  The whole image is taken as the input of the network, and Equasion 9 is adopted for normalization processing:

  $$HSI_{norm} = Mos * mean(HSI)\frac{(}{4}095 * 0.18) \quad (9)$$

  Where, Mos represents the input Mosaic image, HSI is the corresponding ground-truth hyperspectral image, and HSI_{norm} is the normalized result.

  The whole network was trained for 5000 epochs in total, the initial learning rate was $1e^{-4}$ and halved every 1000 epochs. We use the Adam optimizer with $\beta_1 = 0.5, \beta_2 = 0.999$,and the batchsize is set to 4. LeakyReLU activation function was used after each convolution layer. In order to enhance the network generalization ability, the data is randomly flipped vertically and horizontally.

- **Testing**

  In the test phase, as shown in Equation 10,the input image is divided by 4095 for normalization firstly, and then divided by the maximum value of reconstructed image as the final result.

  $$HSI^{'} = G(Mos\frac{4}{0}95)\frac{m}{a}x[G(Mos\frac{4}{0}95)] \quad (10)$$

  Where, Mos is the input Mosaic image, $G()$ is the reconstruction network Res2-Unet proposed above, and $HSI^{'}$ reference the reconstructed hyperspectral image.

6.6. ZJU231: Multi-Scale Mosaic Channel Attention Network for Multi-Spectral Filter Array Demosaicing

As illustrated in Fig. 16, we present a multi-scale mosaic channel attention network (MS-MCAN) for multi-spectral filter array demosaicing. We use residual channel attention network (RCAN) which is the classical image super-resolution model as our baseline [51].

We adopt the mosaic convolution module (MCM) to softly split the periodic spectral mosaic in the raw image during learning [11]. MCM assigns the same weight to pixels that belong to the same spectral band and softly splits the spectral bands into full spatial resolution spectral feature maps. We propose a multi-scale spectral feature extraction method by using three MCM modules with different kernel size, which can better extract spectral features at different scales.

All spectral feature maps extracted by multi-scale MCM will be concatenated and fed to adaptive weighted channel attention module (AWCA) for selectively emphasizing the informative features by exploring adaptive weighted feature statistics [28]. AWCA module can adaptively recalibrate channel-wise feature responses by exploiting the adaptive weighted feature statistics instead of average-pooled ones.

We use the same residual in residual (RIR) structure as RCAN, where the residual group (RG) serves as the basic module and long skip connection (LSC) allows residual learning in a coarse level. In each RG module, we stack several simplified residual block with short skip connection (SSC). It's worth noticing that we replace the channel attention (CA) with AWCA module in each residual channel attention block (RCAB). The details of RCAB and AWCA module are shown in Fig. 17,

Taking the global skip connection into consideration, the final demosaicing multi-spectral image (MSI) cube can be obtained by concatenating the bilinear interpolated cube with the reconstructed residual image cube.

- **Training**

  We use 10 RG in the RIR structure. In each RG, we set RCAB number as 20. Our model is optimized using Adam solver with default settings. The training batch size is set to 2. For data augmentation, we randomly rotate, flip and crop the HS images with the size of 256×256 as labels then create mosaic images as inputs. Besides, the initial learning rate is set to 10^{-4} and is halved every 100 epochs. In total, 300 epochs are adopted during the training phase. We train our model using the L1 Charbonnier loss function.

- **Testing**

  We adopt the model snapshots ($\times 5$) strategies to further improve our model performance.

- **Overexposure Correction**

  Due to the loss of information caused by mask sampling and clip operation, it becomes difficult to restore the scale coefficient, which has a huge impact on the final result. We notice that HS images are normalized. Based on this prior knowledge, we simply find the maximum value M_{base} in mosaic images and process it in two ways:

  1. If $M_{base} < 4095$, it means that the mosaic image has not been clipped. Obviously, due to mask sampling, the maximum value of the original HS images is not necessarily retained in the mosaic images, so

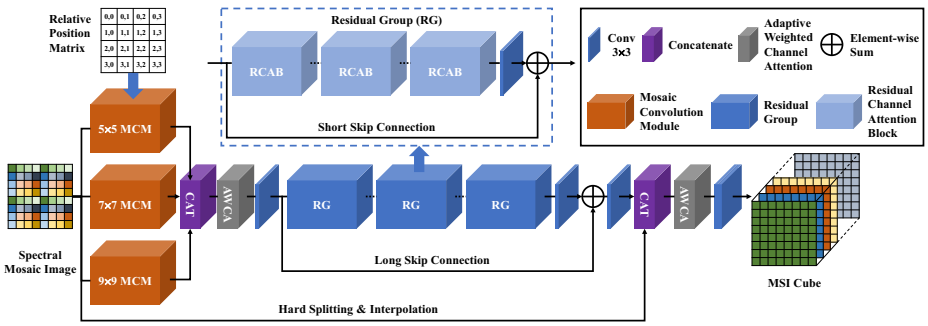


Figure 16. Network architecture of our multi-scale mosaic channel attention network (MS-MCAN).

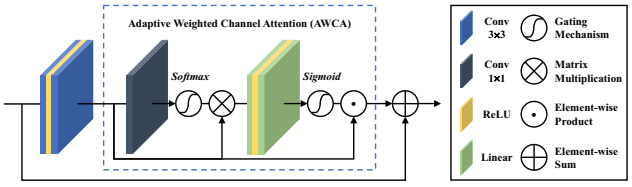


Figure 17. Residual channel attention block (RCAB) and adaptive weighted channel attention (AWCA) module.

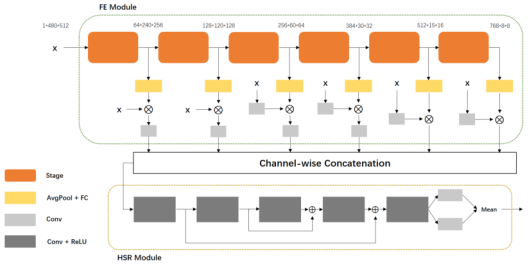


Figure 18. Overview of our model and x means the input mosaic.

the real maximum value $M_{real} \geq M_{base}$. We build a dictionary to save the maximum value of each channel $(M_{C_0}, M_{C_1}, ..., M_{C_{15}}) \in M_C$ of all non-clipped mosaic images and their corresponding difference M_d between M_{real} and M_{base}. For a new non-clipped mosaic image, we calculate its M_{base} and M_C, look up the dictionary to rank by L1 distance of M_C between items of dictionary and the new one then pick the top-m M_d. The final estimated maximum value can be calculated through Eq. 11.

$$M_{real} = M_{base} + \frac{\sum_{i=1}^{m} M_{di}}{m} \tag{11}$$

2. If $M_{base} = 4095$, it means that some pixels of the mosaic image have been clipped. We build a dictionary to save the clip number N_{clip} (number of pixel value that equals to 4095), sum of the non-clipped pixel value S_n and the corresponding M_{real} of all clipped mosaic images. Besides, we divide the mosaic images into 4×4 patches and calculate the sum of each patch, select the one with the largest sum called MAX_PATCH and save it into the dictionary. For a new clipped mosaic image, we look up the dictionary, rank by the different of N_{clip}, S_n and L1 distance of MAX_PATCH between items of dictionary and the new one, then pick the top-n M_{real} and calculate their mean as the final result.

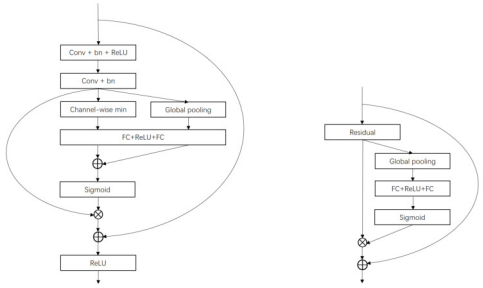


Figure 19. Left is extended channel attention block and right is channel attention block. ⊗ refers to channel-wise multiplication.

6.7. OnRoad (SRC-B): Multi-Scale Image Reconstruction Network for Spectral Demosaic

As shown in Fig. 18, our method consists of two parts and we call them feature extraction module(FE module) and high spectral reconstruction module(HSR module) respectively. FE module extract multi-scale features through 6 stages and then features are mapping to scalars respectively. Mosaic and its convoluted feature maps will be multiplied by scalars. For better reconstructing high spectral images, we concatenate these feature maps which consists of multi-scale information and send them to a straight network. Spatial resolution of feature maps in the straight network will not change.

- **Multi-scale Feature Concatenation**

  To better utilize the semantic information extracted by FE module, we mapping outputs of each stage to scalars and then mosaic and its convoluted feature maps are multiplied by the scalars. In experiments, we find multi-scale information is important for high spectral image reconstruction. We infer that our multi-scale feature concatenation has similar effect to pyramid network.

- **Extended Channel Attention Block**

  Every stage in FE module consists of multiple extended channel attention blocks(ECA blocks). As shown in Figure 19, our ECA block is on the left while the channel attention block(CA block) is on the right. Compared with CA block, our ECA block increases a branch of minimum value per channel. Experiments show that ECA blocks achieve better PSNR. The number of ECA blocks in each stage is 3.

- **HSR Module**

  Compared with U-net, experiments show that straight net that keeps the spatial resolution of feature maps unchanged is more suitable for quality enhancement tasks. At the end of the straight net, we find that predicting two different outputs and then calculate their mean as final output can achieve a better reconstruction effect.

- **Training**

  We train our model with weight decay and data augmentation. The weight decay coefficient is $1e-2$. Our used data augmentation methods consist of horizontal flip, vertical flip, rotation, mixing up, blur, resize and crop and randomly scale. We train our model around 3000 epochs to get the best model in validation data set. The optimizer we used is AdamW optimizer and the batch size is 10. We implement our model and training process with Pytorch. We used a single NVIDIA A100-40GB to train and test.

- **Testing** Our final testing result is the fusion of 5 model results. With the use of a single NVIDIA A100-40GB gpu, we can reconstruct a high spectral image per 430 ms.

- **Ensembles and Fusion Strategies**

  We only fuse model results While testing. We fuse the model results of 5 models to get the mean reconstruction results rather than one model. In experiments we find that one model can only learn a part of data distribution and sometimes model may overfit training set. Multiple model ensemble can significantly solve this problem.

Acknowledgments

We thank the NTIRE 2022 sponsors: Huawei, Reality Labs, Bending Spoons, MediaTek, OPPO, Oddity, Voyage81, ETH Zurich (Computer Vision Lab) and University of Wurzburg (CAIDAS).

A. Teams and affiliations

NTIRE2022 team

Members: *Boaz Arad*[1,2]*(boazar@post.bgu.ac.il)*, *Radu Timofte*[3,4] *(radu.timofte@uni-wuerzburg.de)*, Rony Yahel[1,2,5], Nimrod Morag[1,2,6], Amir Bernat[1,2] ***Affiliations:***

[1] Oddity tech Ltd.
[2] Voyage81 Ltd.
[3] Center for Artificial Intelligence and Data Science, University of Würzburg
[4] Computation Vision Lab, ETH Zürich
[5] The Academic College of Tel Aviv–Yaffo
[6] Tel Aviv University

HITZST01

Members: *Yaqi Wu*[1]*(titimasta@163.com)*, Xun *Wu*[2], *Zhihao Fan*[3], *Chenjie Xia*[4], *Feng Zhang*

Affiliations:

[1] Harbin Institute of Technology, Harbin, 150001, China
[2] Tsinghua University, Beijing, 100084, China
[3] University of Shanghai for Science and Technology, Shanghai, 200093, China
[4] Zhejiang University, Hangzhou, 310027, China

MIALGO

Members: *Shuai Liu*[1]*(liushuai21@xiaomi.com)*, *Yongqiang Li*, *Chaoyu Feng*, *Lei Lei*

Affiliations:

[1] Xiaomi Inc., China

IFL

Members: *Mingwei Zhang*[1]*(dlaizmw@gmail.com)*

Affiliations:

[1] Northwestern Polytechnical University, Xi'an, China

NPUMPI

Members: *Kai Feng*[1]*(fengkainpu@gmail.com)*, *Xun Zhang*, *Jiaxin Yao*, *Yongqiang Zhao*

Affiliations:

[1] Northwestern Polytechnical University, Xi'an, Shaanxi, China

SIP

Members: *Suina Ma[1](1610764697@qq.com), Fan He[1], Yangyang Dong[1], Shufang Yu[1], Difa Qiu[1], Jinhui Liu[1], Mengzhao Bi[1], Beibei Song[1], WenFang Sun[2]*
Affiliations:
[1] Chang'an University,China
[2] Xidian University,China

ZJU231

Members: *Jiesi Zheng[1,2](jaszheng@zju.edu.cn), Bowen Zhao[1,2](bowenzhao@zju.edu.cn), Yanpeng Cao[1,2], Jiangxin Yang[1,2], Yanlong Cao[1,2]*
Affiliations:
[1] State Key Laboratory of Fluid Power and Mechatronic Systems, School of Mechanical Engineering, Zhejiang University, Hangzhou, 310027, China
[2] Key Laboratory of Advanced Manufacturing Technology of Zhejiang Province, School of Mechanical Engineering, Zhejiang University, Hangzhou, 310027, China

OnRoad (SRC-B)

Members: *Xiangyu Kong[1](85094407@qq.com), Jingbo Yu, Yuanyang Xue, Zheng Xie*
Affiliations:
[1] Affiliation

References

[1] Giancarlo A Antonucci, Simon Vary, David Humphreys, Robert A Lamb, Jonathan Piper, and Jared Tanner. Multispectral snapshot demosaicing via non-convex matrix completion. In *2019 IEEE Data Science Workshop (DSW)*, pages 227–231. IEEE, 2019. 2

[2] 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

[3] Boaz Arad, Radu Timofte, Rony Yahel, Nimrod Morag, Amir Bernat, et al. NTIRE 2022 spectral demosaicing challenge and dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 2

[4] 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. 2, 4

[5] Elena Beletkaia and Jose Pozo. More than meets the eye: Applications enabled by the non-stop development of hyperspectral imaging technology. *PhotonicsViews*, 17(1):24–26, 2020. 1

[6] Goutam Bhat, Martin Danelljan, Radu Timofte, et al. NTIRE 2022 burst super-resolution challenge. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 2

[7] Chein-I Chang. *Hyperspectral data exploitation: theory and applications*. John Wiley & Sons, 2007. 1

[8] Qi Cui, Jongchan Park, R Theodore Smith, and Liang Gao. Snapshot hyperspectral light field imaging using image mapping spectrometry. *Optics letters*, 45(3):772–775, 2020. 1

[9] Gabriella M Dalton, Noelle M Collins, Joshua M Clifford, Emily L Kemp, Ben Limpanukorn, and Edward S Jimenez. Monte-carlo modeling and design of a high-resolution hyperspectral computed tomography system with multi-material patterned anodes for material identification applications. In *Developments in X-Ray Tomography XIII*, volume 11840, pages 78–94. SPIE, 2021. 1

[10] Egor Ershov, Alex Savchik, Denis Shepelev, Nikola Banic, Michael S Brown, Radu Timofte, et al. NTIRE 2022 challenge on night photography rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 2

[11] 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. 2, 8, 10

[12] Shang-Hua Gao, Ming-Ming Cheng, Kai Zhao, Xin-Yu Zhang, Ming-Hsuan Yang, and Philip Torr. Res2net: A new multi-scale backbone architecture. *IEEE transactions on pattern analysis and machine intelligence*, 43(2):652–662, 2019. 9

[13] Jinjin Gu, Haoming Cai, Chao Dong, Jimmy Ren, Radu Timofte, et al. NTIRE 2022 challenge on perceptual image quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 2

[14] Medha Gupta. Generalizing spectral difference method for multispectral image demosaicking and analyzing the role of msfa patterns. In *2020 9th International Conference System Modeling and Advancement in Research Trends (SMART)*, pages 416–420. IEEE, 2020. 2

[15] Medha Gupta and Mangey Ram. Weighted bilinear interpolation based generic multispectral image demosaicking method. *Journal of Graphic Era University*, pages 108–118, 2019. 2

[16] Tewodros Amberbir Habtegebrial, Gerd Reis, and Didier Stricker. Deep convolutional networks for snapshot hypercpectral demosaicking. In *2019 10th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pages 1–5. IEEE, 2019. 2

[17] Nathan A Hagen and Michael W Kudenov. Review of snapshot spectral imaging technologies. *Optical Engineering*, 52(9):090901, 2013. 1

[18] Robin Hahn, Freya-Elin Hämmerling, Tobias Haist, David Fleischle, Oliver Schwanke, Otto Hauler, Karsten Rebner, Marc Brecht, and Wolfgang Osten. Detailed characterization of a mosaic based hyperspectral snapshot imager. *Optical Engineering*, 59(12):125102, 2020. 1

[19] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pages 7132–7141, 2018. 7, 9

[20] Haris Ahmad Khan, Sofiane Mihoubi, Benjamin Mathon, Jean-Baptiste Thomas, and Jon Yngve Hardeberg. Hytexila: High resolution visible and near infrared hyperspectral texture images. *Sensors*, 18(7):2045, 2018. 2

[21] Byung-Hoon Kim, Joonyoung Song, Jong Chul Ye, and Jae-Hyun Baek. Pynet-ca: enhanced pynet with channel attention for end-to-end mobile image signal processing. In *European Conference on Computer Vision*, pages 202–212. Springer, 2020. 5

[22] Srivathsan Koundinyan, Edward Steven Jimenez, Kyle R Thompson, and April Suknot. Material identification and classification using machine learning techniques with hyperspectral computed tomography. Technical report, Sandia National Lab.(SNL-NM), Albuquerque, NM (United States), 2018. 1

[23] Fred A Kruse, AB Lefkoff, JW Boardman, KB Heidebrecht, AT Shapiro, PJ Barloon, and AFH Goetz. The spectral image processing system (sips)—interactive visualization and analysis of imaging spectrometer data. *Remote sensing of environment*, 44(2-3):145–163, 1993. 3

[24] Andy Lambrechts, Pilar Gonzalez, Bert Geelen, Philippe Soussan, Klaas Tack, and Murali Jayapala. A cmos-compatible, integrated approach to hyper-and multispectral imaging. In *2014 IEEE International Electron Devices Meeting*, pages 10–5. IEEE, 2014. 1

[25] 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. 1

[26] Sophie Lemmens, Toon Van Craenendonck, Jan Van Eijgen, Lies De Groef, Rose Bruffaerts, Danilo Andrade de Jesus, Wouter Charle, Murali Jayapala, Gordana Sunaric-Mégevand, Arnout Standaert, et al. Combination of snapshot hyperspectral retinal imaging and optical coherence tomography to identify alzheimer's disease patients. *Alzheimer's research & therapy*, 12(1):1–13, 2020. 1

[27] 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 *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, pages 462–463, 2020. 7, 8

[28] 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. 10

[29] Yasheng Li, Ningfang Liao, Xueqiong Bai, Haobo Cheng, Wenming Yang, and Chenyang Deng. An on-line color defect detection method for printed matter based on snapshot multispectral camera. In *Advanced Optical Imaging Technologies*, volume 10816, page 1081612. International Society for Optics and Photonics, 2018. 1

[30] Yawei Li, Kai Zhang, Radu Timofte, Luc Van Gool, et al. NTIRE 2022 challenge on efficient super-resolution: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 2

[31] Andreas Lugmayr, Martin Danelljan, Radu Timofte, et al. NTIRE 2022 challenge on learning the super-resolution space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 2

[32] Sofiane Mihoubi. *Snapshot multispectral image demosaicing and classification*. PhD thesis, Université de Lille, 2018. 2

[33] 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. 2

[34] Yusuke Monno, Hayato Teranaka, Kazunori Yoshizaki, Masayuki Tanaka, and Masatoshi Okutomi. Single-sensor rgb-nir imaging: High-quality system design and prototype implementation. *IEEE Sensors Journal*, 19(2):497–507, 2018. 2

[35] Ben Niu, Weilei Wen, Wenqi Ren, Xiangde Zhang, Lianping Yang, Shuzhen Wang, Kaihao Zhang, Xiaochun Cao, and Haifeng Shen. Single image super-resolution via a holistic attention network. In *European conference on computer vision*, pages 191–207. Springer, 2020. 6

[36] Zhihong Pan, Baopu Li, Yingze Bao, and Hsuchun Cheng. Deep panchromatic image guided residual interpolation for multispectral image demosaicking. In *2019 10th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pages 1–5, 2019. 2

[37] Eduardo Pérez-Pellitero, Sibi Catley-Chandar, Richard Shaw, Ales Leonardis, Radu Timofte, et al. NTIRE 2022 challenge on high dynamic range imaging: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 2

[38] Vishwas Rathi and Puneet Goyal. Convolution filter based efficient multispectral image demosaicking for compact ms-fas. In *VISIGRAPP (4: VISAPP)*, pages 112–121, 2021. 2

[39] Andres Romero, Angela Castillo, Jose M Abril-Nova, Radu Timofte, et al. NTIRE 2022 image inpainting challenge: Report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 2

[40] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 9

[41] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1874–1883, 2016. 9

[42] Zhan Shi, Chang Chen, Zhiwei Xiong, Dong Liu, and Feng Wu. Hscnn+: Advanced cnn-based hyperspectral recovery from rgb images. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, pages 939–947, 2018. 7

[43] Kazuma Shinoda, Shoichiro Yoshiba, and Madoka Hasegawa. Deep demosaicking for multispectral filter arrays. *arXiv preprint arXiv:1808.08021*, 2018. 2

[44] Grigorios Tsagkatakis, Maarten Bloemen, Bert Geelen, Murali Jayapala, and Panagiotis Tsakalides. Graph and rank regularized matrix recovery for snapshot spectral image demosaicing. *IEEE Transactions on Computational Imaging*, 5(2):301–316, 2018. 2

[45] Gregg Vane, Robert O Green, Thomas G Chrien, Harry T Enmark, Earl G Hansen, and Wallace M Porter. The airborne visible/infrared imaging spectrometer (aviris). *Remote sensing of environment*, 44(2-3):127–143, 1993. 2

[46] Kathleen Vunckx and Wouter Charle. Accurate video-rate multi-spectral imaging using imec snapshot sensors. In *2021 11th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pages 1–7. IEEE, 2021. 1

[47] Longguang Wang, Yulan Guo, Yingqian Wang, Juncheng Li, Shuhang Gu, Radu Timofte, et al. NTIRE 2022 challenge on stereo image super-resolution: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 2

[48] Xiaolong Wang, Ross Girshick, Abhinav Gupta, and Kaiming He. Non-local neural networks. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pages 7794–7803, 2018. 7

[49] Ren Yang, Radu Timofte, et al. NTIRE 2022 challenge on super-resolution and quality enhancement of compressed video: Dataset, methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2022. 2

[50] 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. 2, 7

[51] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 286–301, 2018. 5, 10