IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 11 December 2024, accepted 28 December 2024, date of publication 13 January 2025, date of current version 27 January 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3528753

RESEARCH ARTICLE

Deep Joint Demosaicking and Super-Resolution for Spectral Filter Array Images

ABDELHAMID N. FSIAN[1], (Student Member, IEEE), JEAN-BAPTISTE THOMAS[1,2], JON Y. HARDEBERG[2], (Senior Member, IEEE), AND PIERRE GOUTON[1]

[1]Imagerie et Vision Artificielle (ImVIA) Laboratory, Department Informatique, Electronique, Mécanique (IEM), Université de Bourgogne, 21000 Dijon, France
[2]Colourlab, Department of Computer Science, Norwegian University of Science and Technology (NTNU), 2815 Gjøvik, Norway

Corresponding author: Jon Y. Hardeberg (jon.hardeberg@ntnu.no)

ABSTRACT In spectral imaging, the constraints imposed by hardware often lead to a limited spatial resolution within spectral filter array images. On the other hand, the process of demosaicking is challenging due to intricate filter patterns and a strong spectral cross correlation. Moreover, demosaicking and super resolution are usually approached independently, overlooking the potential advantages of a joint solution. To this end, we use a two-branch framework, namely a pseudo-panchromatic image network and a pre-demosaicking sub-branch coupled with a novel deep residual demosaicking and super resolution module. This holistic approach ensures a more coherent and optimized restoration process, mitigating the risk of error accumulation and preserving image quality throughout the reconstruction pipeline. Our experimental results underscore the efficacy of the proposed network, showcasing an improvement of performance both qualitatively and quantitatively when compared to the sequential combination of state-of-the-art demosaicking and super resolution. With our proposed method, we obtained on the ARAD-1K dataset an average PSNR of 48.02 (dB) for domosaicking only, equivalent to the best method of the state-of-the-art. Moreover, for joint demosaicking and super resolution our model averages 35.26 (dB) and 26.29 (dB), respectively for $\times 2$ and $\times 4$ upscale, outperforming state-of-the-art sequential approach.The codes and datasets are available at https://github.com/HamidFsian/DRDmSR.

INDEX TERMS Spectral imaging, demosaicking, super resolution, deep learning, pseudo-panchromatic image, spectral filter array.

I. INTRODUCTION

Spectral imaging has emerged as a promising candidate to address the inherent limitations of color imaging. While color cameras have traditionally been invaluable tools for image capture across a myriad of applications, they are limited in spectral separability and to provide a comprehensive spectral information. In contrast to traditional color cameras, spectral cameras offer a more comprehensive source of spectral data. This superior capability proves especially advantageous in applications such as food safety inspection [1], land cover categorization [2] and object tracking [3].

Conventional spectral imaging is a spatial, spectral and temporal sampling on the image plane. Implementing a snapshot imaging technique increase the temporal resolution at the expense of the spectral and spatial information [4]. A good temporal resolution is relevant to analyse dynamic systems. spectral filter array (SFA) [4] imaging techniques are based on the color filter array (CFA) and obtain both spatial and spectral information from a single image exposure. Its tiny sensor size and capacity to capture snapshots present a wealth of possibilities for a variety of uses.

An SFA-based device is used to create a spectral mosaic image, similar to color imaging sensors with CFA. For traditional methods, extending CFA demosaicking algorithms towards SFA algorithm has been explored [5]. However, this straightforward rearrangement results in aliasing distortion in both the spectral and spatial domains and do not systematically incorporate high-frequency information due to the increased number of bands and the absence of a dominant

The associate editor coordinating the review of this manuscript and approving it for publication was Olarik Surinta.

band. Therefore, spectral demosaicking, which refers to the process of creating a fully-defined multispectral image (MSI) free of spatial and spectral distortions, is a crucial stage in SFA-based imaging approaches. Moreover, and due to hardware limitations, most SFA devices have substantially lower spatial resolution compared to CFA-based images. To address this limitation, super resolution techniques emerge as indispensable tools in enhancing the spatial resolution of SFA images.

The exploration and application of demosaicking and super resolution techniques have been central to research and practical uses for decades. However, treating demosaicking and super resolution as separate processes can be suboptimal, leading to error accumulation, as mentionned by [6]. One significant concern in this regard is the potential propagation and magnification of artifacts introduced during demosaicking in subsequent super resolution processing stages. For instance, super resolution algorithms may interpret demosaicking artifacts, such as color zippering, as valid components of the input image signal, contributing to an increase in overall image inaccuracies. This issue complicates the accurate reconstruction of spectral images, emphasizing the need for a more integrated and cohesive approach to address both demosaicking and super resolution tasks simultaneously. Despite the well-established understanding that the sequential application of demosaicking and super resolution for both SFA and CFA cameras is sub-optimal [7], [8], there has been relatively less attention given to the development of a joint solution compared to sequential approach. Acknowledging this gap in the existing research, our paper aims to contribute to the field by proposing a novel approach – a joint demosaicking super resolution model for SFA images. Through this joint framework, we aim to enhance the overall quality of reconstructed images from spectral filter array cameras, addressing the limitations associated with the traditional sequential application of these image processing techniques [7]. To summarize, The main contributions of our paper are delineated as follows:

1) To the best of our knowledge, our paper introduces the first joint demosaicking and super resolution network specifically designed for SFA-based images. This innovative approach aims to overcome the limitations associated with the traditional sequential application [7], [9] of these two essential image processing techniques.
2) We enhance the capabilities of our network through the incorporation of a novel module named Deep Residual Demosaicking and Super Resolution (DRDmSR). This module, featuring Residual in Residual (RIR) structure to obtain a very deep trainable network. Skip connections, both long and short, in RIR aid in bypassing prevalent low-frequency information, enabling the main network to effectively learn more valuable information. In addition to channel attention (CA) mechanism which adjusts feature scale by considering the interdependencies among feature channels.
3) Through rigorous experimentation and evaluation, our proposed architecture demonstrates superior performance compared to existing methodologies in the field of spectral demosaicking and super resolution for both synthetic and real data. The empirical results showcase the efficacy of our approach, establishing a new benchmark for image reconstruction in SFA-based cameras. Furthermore, spectral reconstruction from RGB data to train our framework has shown great utility in the generalization of our framework (see section V-C).

Our paper is structured as follows: In Section II, we delve into the existing body of work related to demosaicking and super resolution, presenting a review of prior research efforts and methodologies in each domain. Following this, Section III provides an insight into our imaging system model, explaining key components and considerations that form the foundation of our experimental setup. Moving to the core of our contribution, Section IV introduces our proposed joint demosaicking and super resolution framework. We detail the architecture of our model, emphasizing the integration of the DRDmSR and its coupling with the Pseudo Panchromatic Network (PPI-Net). In Section V, we transition to the experimentation phase, providing an extensive exploration of our experimental setup, dataset, and performance metrics. The empirical results and comparisons with existing methodologies are presented to validate the effectiveness of our proposed approach. Finally, Section VI concludes the paper, summarizing our main contributions, discussing the implications of our results, and suggesting potential avenues for future research in the context of spectral filter array cameras.

II. RELATED WORK

In this section, we offer an overview of previous research and discuss the challenges related to demosaicking and super resolution. Then, we explore the existing literature concerning joint solutions for these tasks.

A. DEMOSAICKING

Image demosaicking (DM), is an ill-posed problem that involves interpolating full-resolution spectral images from mosaic images. Model-based and learning-based approaches are the two basic groups into which existing methods can be divided. In order to facilitate the recovery of missing data, model-based techniques [10], [11] concentrate on building mathematical models and image priors in the spatial-spectral domain. Learning-based methods [11], [12] use a wealth of training data to learn how to construct the process mapping.

Previous studies on SFA image demosaicking have looked into a variety of handcrafted techniques [13], [14], [15], [16]. Brauers and Aach [17] introduced the weighted bilinear (WB) interpolation technique and expanded the CFA demosaicking to SFA. The pseudo-panchromatic image (PPI), or average

image of all spectral channels, was initially introduced by Chini et al. [18]. The PPI difference (PPID) method was introduced by Mihoubi et al. [19], [20] to facilitate SFA demosaicking by sharpening the PPI using a channel residual structure, taking into account the spatial and spectral correlation of SFA image. A vast number researchers tried to develop data-driven methods to achieve high-accuracy SFA image demosaicking [13], [21], [22], [23], [24], motivated by the recent success of deep convolutional neural networks (CNNs) in various SFA-based image applications, such as object tracking [25], image denoising [26] and image deblurring [27].

In a similar vein, several CNN-based deep learning models for demosaicking have been proposed. In comparison to the PPID technique [20], Shinoda et al. [28] shows better results using a deep demosaicking network that makes use of three-dimensional (3D) convolutions and a deep residual network ResNet. Nevertheless, images produced through this methodology may manifest false color artifacts in regions characterized by high contrast and luminosity. Feng et al. [29] introduced a deep CNN using mosaic convolution-attention network showing the importance of initial feature extraction from the raw mosaic image. But, their mosaic convolution module do ignore the fluctuation of spatial locations, which leads to checkerboard distortion. Pan et al. [30] introduces a modification to the demosaicking framework by incorporating PPIs, estimated through a CNN, and employing a traditional two-branch residual interpolation method for demosaicking. The first branch utilizes the CNN-generated PPI to guide the handling of residuals between each subsampled band and the corresponding PPI. In the second branch, the initially demosaicked band is employed to further mitigate residuals between itself and the subsampled mosaic image. Nevertheless, the incorporation of guided filters in this process gives rise to a halo effect in the demosaicking results. Following the success of Pan et al. [30], Zhao et al. [31] introduced a Residual Network and an improved PPI generation by infusing edge-related information and employing adaptive spatial and spectral compensation within the network to improve demosaicking results, while achieving state-of-the-art metrics.

B. SUPER RESOLUTION

Single Image super resolution (SISR) is the process of producing a high resolution (HR) image from it corresponding low resolution (LR) image. We can categorize SR methods into two groups: Model based techniques [32], [33], [34] and learning based methods [35]. Model-based techniques for SISR are infamous for their aliasing artifacts and edge blurring [32], [33]. Deep learning-based methods have made significant progress recently. In order to tackle the SISR challenge, SRCNN [35] first presented a deep learning based model that performed significantly better than model based methods. Benefiting from the ResNet [36] approach, VDSR [37] trained 20 layers deep networks with long residual connections. These networks could only learn more high-frequency data and accelerate convergence. In order to further enhance SISR performance, EDSR [38] suggested integrating several resblocks and removing the batch-normalization layer. This can save GPU RAM, stack more layers, and widen networks [39]. LapSRN [40] suggested repeatedly super-resolving LR images in order to reduce GPU memory usage and improve performance. Moreover, Sidorov and Hardeberg [41] proposed a fully-convolutional encoder-decoder network designed to reconstruct images from noisy inputs by leveraging the intrinsic prior contained in the network structure without any training.

The most recent work include RDN [42], produced residual dense blocks (RDB) by combining ResNet with DenseNet [43]. The proposed RDB can allow higher growth rate to improve performance through local feature fusion. In order to calibrate feature maps and propose RIR structure to produce a very deep convolutional networks that achieved new state-of-the-art performance for SISR task, RCAN [44] first included attention mechanism, which was influenced by SENet [45]. However, the joint solution of demosaicking and super resolution has been underesearched so far with some exceptions in the CFA community [6], [7], [9].

C. JOINT SOLUTION

Even though there have been a lot of CDM and SISR methods presented in the previous several decades, in practical applications, the two approaches are often explored separately and used in order. Although it is much easier to examine the two problems individually than to consider them combined, there are three major issues with the former plan:

1) Sub-optimal process. Both tasks are closely connected and may be thought of as a similar inverse problem since they are based on the limits of sensors for determining spatial and spectral information. Thus, it is not ideal from a mathematical perspective to split the joint issue into two subproblems and solve them one after the other.
2) Accumulation of error. Since DM is an inverse problem. Artifacts and errors may be introduced. Consequently, these artifacts may mislead the SISR method, resulting in the accumulation and spread of these artifacts in the final image.
3) An inefficient use of computational and memory resources. The well-established "CDM-followed-by-SR" pipeline for model-based algorithms [13], [32] suggests separately investigating prior knowledge for the two tasks on the same image content, which may result in needless repetition of computation. The convolutional layers used for feature extraction and refinement in CNN-based techniques [29], [42] cannot be effectively used by both tasks in the sequential pipeline.

While joint demosaicking and super resolution (JDSR) seems to be an appealing idea, the bibliography for a JDSR

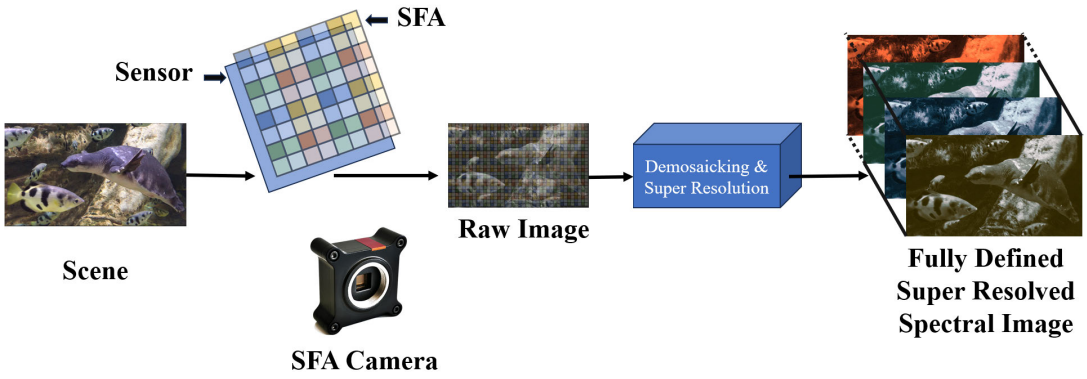


FIGURE 1. The raw data acquisition process of a scene using snapshot spectral camera and the reconstruction of the fully-defined super resolved image.

approaches is short. Moreover, it consist only CFA-based solution, and no work have been proposed for SFA based imaging. The very first model was proposed by [7], where several residual blocks where directly linked to establish a nonlinear mapping from the low resolution image input to the high resolution output. Moreover, Xu et al. [6] introduced a JDSR network called RDSEN, where the foundational unit integrates dense connections into the RCAB [44]. Preceding RDSEN, a pre-demosaicking network (PDNet) was devised to generate an intermediate demosaicked image. Through joint end-to-end training of PDNet and RDSEN, the model presented in [6] achieves state-of-the-art performance. In a recent development, Chang et al. [9] propsed the TSCNN model which differs from RDSEN [6] by maintaining the same resolution as the Bayer-sampled LR image for initial feature extraction, utilizing the green channel for better performance. It also includes densely-connected dual-path enhancement blocks (DDEB) and a dual-branch feature refinemen (DFR) module to decompose input features into high and low-frequency components, enhancing multi-frequency information in images for an efficient CNN model. These models are specifically tailored for CFA based camera, using a bayer filter and taking advantage of a dominant band (Green band). However, they are not suitable for SFA-based images, incorporating a spectral filter array with more spectral bands and may or may not exhibit a Bayer-like Green dominant channel.

III. MODEL AND PRELIMINARIES

A. SFA MOSAICKING MODEL

First, the incident light is directed into the spectral filter array, and then into the single sensor, which in turn provides a mosaic image, denoted as I_{raw}, with a spatial resolution of $W \times H$ pixels. Moreover, at every pixel of the mosaic image I_{raw}, only a single band is available out of the C bands. Mathematically, by taking into account that a fully-defined MSI with C bands (which is not available in practice) $\{I_c\}_{c=1}^{C}$, is modulated by $\{S_c\}_{c=1}^{C}$. Therefore, the mosaic image is formulated as

$$I_{raw} = \sum_{c=1}^{C} S_c \odot I_c, \tag{1}$$

where S_c is a sparse band-wise binary mask, containing values solely at positions that align with the c band on the SFA. Moreover, $\odot$ is the element-wise product. Demosaicking is performed on each sparse channel of S_c to obtain a reconstructed image $\hat{I}$ with C fully defined channels. Figure 1 shows the imaging reconstruction process scheme.

B. PPI: METHODOLOGY AND ESTIMATION

The PPI is determined at each pixel by calculating the mean value over all channels of a fully defined spectral image [18]

$$I_{PPI} = \frac{1}{N} \sum_{c=1}^{C} I_c. \tag{2}$$

Following the assumption in [18], when channels exhibit spectral separation, signifying a notable distance between the centers of bands associated with these channels, they demonstrate a higher correlation with the pseudo-panchromatic image (PPI) than with each other. Mihoubi et al. [19] introduced a straightforward approach for estimating an initial PPI by convolving the SFA with a weighted average filter, denoted as M. Given the impracticality of obtaining all I_c values for $c \in \{1, C\}$ at pixel position c, M is configured to consider surrounding I values around pixel position. This design is based on the assumption that neighboring pixels of the PPI exhibit strong correlations. Additionally, the coefficients of M are structured to ensure that the sum of coefficients for each channel is identical, emphasizing equal importance for each channel in the PPI generation process. Following the notation from [46] for an $m \times n$ SFA mosaic pattern with no dominant band, the M matrix is expressed as

$$M = \frac{M_m \times M_n}{m \times n}, \tag{3}$$

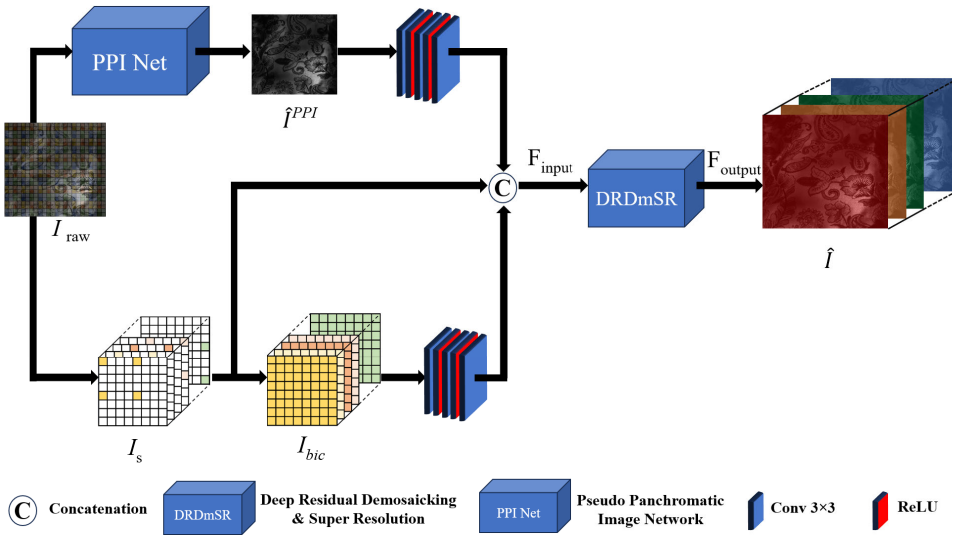


FIGURE 2. Overview of our model framework. I_{raw}, I_s, I_{bic}, $\hat{I}^{PPI}$ and $\hat{I}$ denotes respectively, input raw image, sparse image, pre-demosaicked image, estimated pseudo panchromatic image and the estimated demosaicked super resolved image. Our framework contains two main modules namely PPI-net and the DRDmSR module.

where

$$M_m = \begin{cases} K_{1,m} & \text{if m is odd,} \\ [\frac{1}{2}, K_{1,m-1}, \frac{1}{2}] & \text{otherwise,} \end{cases} \tag{4}$$

$$M_n = \begin{cases} {K_{1,n}}^T & \text{if n is odd,} \\ [\frac{1}{2}, K_{1,n-1}, \frac{1}{2}]^T & \text{otherwise,} \end{cases} \tag{5}$$

where the term $K_{1,m}$ denotes a matrix consisting of all ones, with dimensions $1 \times$ m.

An initial PPI noted as $\hat{I}^M$ is derived by computing the convolution between I_{raw} and M

$$\hat{I}^M = I_{raw} * M. \tag{6}$$

Building upon the assumption articulated by Mihoubi et al. [20], asserting the strong correlation among neighboring pixels, the filter matrix M is structured specifically for a 4×4 SFA mosaic pattern

$$M = \frac{1}{64} \begin{bmatrix} 1 & 2 & 2 & 2 & 1 \\ 2 & 4 & 4 & 4 & 2 \\ 2 & 4 & 4 & 4 & 2 \\ 2 & 4 & 4 & 4 & 2 \\ 1 & 2 & 2 & 2 & 1 \end{bmatrix}. \tag{7}$$

IV. DEEP JOINT DEMOSAICKING AND SUPER RESOLUTION NETWORK

In this section, we present a comprehensive framework featuring a two-branch Convolutional Neural Network (CNN) designed specifically for the demosaicking and Super Resolution of SFA images, as illustrated in Figure 2. Our network architecture is characterized by the integration of two distinct branches: the first branch comprises a Deep Pseudo Panchromatic Image network (PPI-Net) [46]. While the second branch performs pre-demosaicking operations using interpolation method. These branches are seamlessly fused into a unified module termed the Deep Residual Demosaicking and Super Resolution (DRDmSR) module.

In the subsequent sections, we introduce the first branch, namely the deep PPI-net, which serves as the cornerstone of our approach. Following this, we delve into the details of the DRDmSR module, elucidating its underlying principles and operational mechanisms. Through this sequential presentation, we aim to provide a comprehensive understanding of the intricate interplay between the constituent components of our proposed framework.

A. DEEP PPI NETWORK

Considering the pronounced positive linear correlation observed between the reconstructed spectral image and the high-frequency details preserved in the PPI [46], we advocate for the utilization of elementwise summation between $\bar{I}^M$ and $\hat{I}^M$, as illustrated in Figure 3. Furthermore, recognizing that the initial PPI generated by Equation 6 can be interpreted as a smoothed version of I_{PPI} [46], we adopt an architecture proposed by [46], inspired by methodologies in the image deblurring domain. Subsequently, we leverage the PPI-net to enhance the sharpness of $\bar{I}^M$.

Figure 3, illustrates the PPI-Net, which integrates a CNN architecture with a conventional PPI estimation method. Initially, a preliminary $\hat{I}^M$ is obtained by applying the PPI filter M to the raw mosaic image I_{raw}. Subsequently, four convolutional layers with varying kernel sizes ($C \times 9 \times 9$, $C \times 7 \times 7$, $C \times 5 \times 5$, and $1 \times 5 \times 5$, where C denotes the number of channels) are used to determine the residual between the preliminary $\bar{I}^M$ and the ground truth I_{PPI} defined

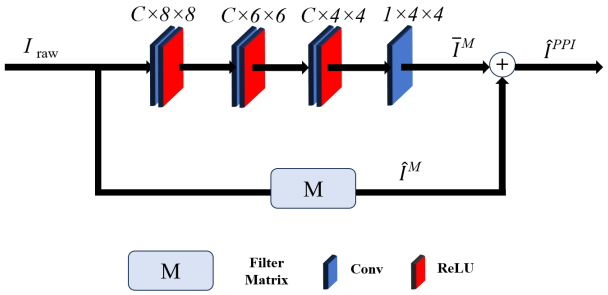


FIGURE 3. Overview of PPI-net module. where ⊕ denotes elementwise sum and *M* the filter matrix.

in Equation 2 in order to estimate $\hat{I}^{PPI}$.

$$L_{PPI} = \sum_{w=1}^{W}\sum_{h=1}^{H}\left|\left|\hat{I}^{PPI}(h,w) - I_{PPI}(h,w)\right|\right|, \tag{8}$$

here, L_{PPI} denotes the loss function targeted for minimization within our PPI-net module, and I_{PPI} is the simulated ground truth PPI defined in Equation 2. Additionally, H and W represent the height and width of the PPI, while h and w indicate the pixel coordinates.

B. DRDmSR NETWORK

After obtaining the estimated PPI, denoted as $\hat{I}^{PPI}$, and the pre-demosaicked image, achieved through bicubic interpolation denoted as I_{bic} in Figure 2, both outputs are fed into identical blocks. They are then concatenated along with the sparsity image for input into the DRDmSR module.

The key components of our DRDmSR module include the upscale module, reconstruction section, shallow feature extraction, RIR deep feature extraction, as depicted in Figure 4.

Let's refer to F_{input} and F_{output} as DRDmSR's input and output, respectively. As examined in [44], we derive the shallow feature F_0 from the F_{input} input using a single convolutional layer (Conv)

$$F_0 = H_{Conv}(F_{input}), \tag{9}$$

where the convolution operation is indicated by $H_{Conv}(.)$. The RIR module is then used to extract deep features using F_0. Therefore, we can write

$$F_{RIR} = H_{RIR}(F_0), \tag{10}$$

here, $H_{RIR}(.)$ indicates our deep residual in residual structure, which contains two residual groups (RG). This H_{RIR} variation, influenced by the success of RCAN [44], incorporates another rendition of the channel attention mechanism similar to the one employed in SE-Resnet [45]. Our suggested RIR offers an especially large receptive field size. As a result, we handle its output F_{RIR} as a deep feature, which is subsequently processed by an upscale module to increase the spatial resolution of the spectral image.

$$F_{output} = F_{UM} = H_{UM}(F_{RIR}), \tag{11}$$

here, F_{UM} is the upscaled output and the $H_{UM}(.)$ denotes the upscaling module depicted in Figure 4.

Various choices exist for upscale modules, each offering distinct advantages and trade-offs. Common upscale modules include the deconvolution layer, also known as the transposed convolution layer [47], which aims to upsample the feature maps by learning an inverse operation to convolution. Another approach involves nearest-neighbor upsampling followed by convolution [48], which utilizes simple interpolation techniques before applying convolutional operations. Additionally, the Efficient Sub-Pixel Convolutional Network (ESPCN) [49] presents an alternative method, where the network directly learns to upscale feature maps by utilizing sub-pixel convolutional layers. In our implementation, following the success of ESPCN, we adopt a similar upscaling strategy, employing convolutional layers followed by a pixel shuffle operation to enhance spatial resolution.

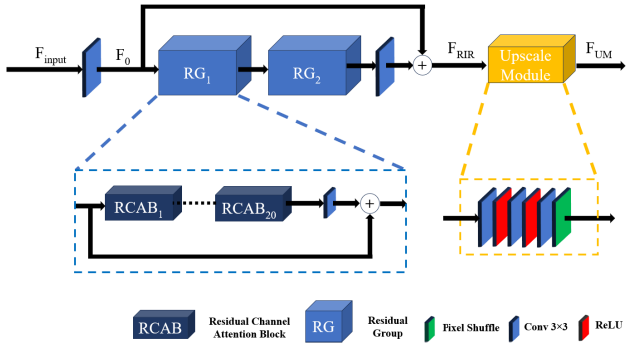


FIGURE 4. Overview of the DRDmSR module where it consists of 3 main blocks, Residual Group (RG), Residual channel attention block (RCAB) and upscaling module.

C. RESIDUAL IN RESIDUAL STRUCTURE

The RIR structure is a pivotal element in our approach to joint demosaicking and super resolution. It enables training of deep convolutional neural networks (CNNs), resulting in superior performance. Comprising RG (see Figure 4) containing Residual Channel Attention Blocks (RCAB) and Long Skip Connections (LSC). Moreover, RCABs (see Figure 5) introduce an adaptive channel attention mechanism within each RG, rescaling channel-wise features based on interdependencies among channels. This enhances CNN's representational ability, focusing on high-frequency information while bypassing low-frequency details through Short Skip Connections (SSC). LSCs establish long-range connections between RGs, ensuring effective gradient flow during training and efficient propagation of useful information and allow the RIR structure learning residual information in a coarse level.

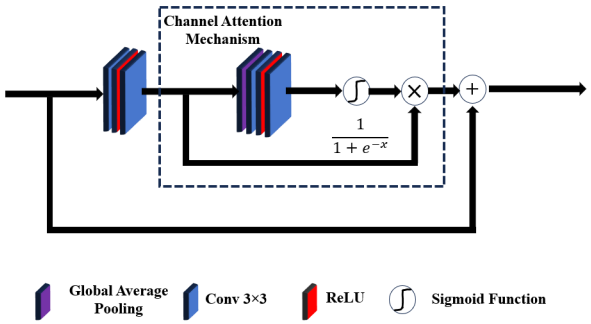


FIGURE 5. Overview of RCAB module, where ⊕ and ⊗ denotes respectively, elementwise sum and elementwise multiplication.

D. LOSS FUNCTION

We employ a composite loss function, denoted by L, to train the entire framework. Its objective is to concurrently minimize the signal reconstruction errors of both the PPI and joint demosaicked super-resolved image.

$$L = L_{PPI} + L_{DRDmSR}, \tag{12}$$

where L_{PPI} is the loss function of the PPI-Net defined in Equation 8, L_{DRDmSR} is the loss function for the DRDmSR module, where this latter is a combination as,

$$L_{DRDmSR} = L_{mse} + L_{wavelet}. \tag{13}$$

Mean Squared Error (MSE) is commonly employed as a loss function in the field of image processing. Its purpose is to drive the pixel values of the estimated image closer to those of the ground truth image in their entirety. Thus, we used the MSE as a loss function, comparing both the ground truth image and the estimated demosaicked and super-resolved image.

$$L_{mse} = \frac{1}{P}\sum_{p=1}^{P}\left|\left|I_p^{GT} - \hat{I}_p\right|\right|_2^2, \tag{14}$$

here, P denotes the total number of pixels, with p representing the index of the pixel.

In order to enhance the sharpness and texture richness of the estimated demosaicked and super-resolved image, we incorporate an additional edge loss, denoted as $L_{wavelet}$. This involves transforming both the ground truth image I^{GT} and the estimated image $\hat{I}$ into the wavelet domain, followed by computing the MSE between the transformed I^{GT} and $\hat{I}$ within the high-frequency sub-bands. In this study, the edge loss is evaluated within the wavelet domain, which can be represented as follows:

$$L_{wavelet} = \frac{1}{P_w}\sum_{p=1}^{P_w}\left|\left|w_p^{GT} - \hat{w}_p\right|\right|_2^2, \tag{15}$$

here, w_p^{GT} and $\hat{w}_p$ represent the coefficient of I^{GT} and $\hat{I}$, respectively. P_w denotes the number of high-frequency wavelet coefficients obtained from the image decomposition using the stationary wavelet transform. For our analysis, we employ the Haar filter and configure the transform level to 2.

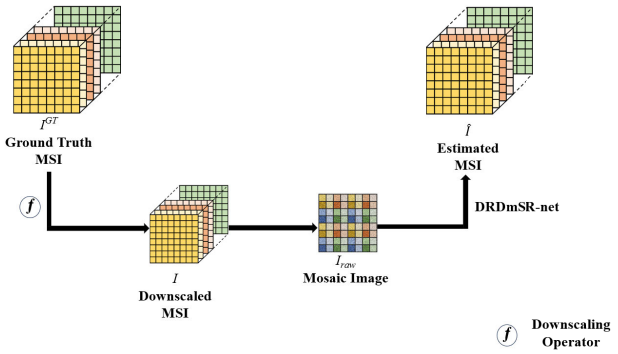


FIGURE 6. Ilustration of data pipeline. From high resolution fully-defined spectral image denoted I^{GT} to demosaicked super resolved spectral image denotes as $\hat{I}$.

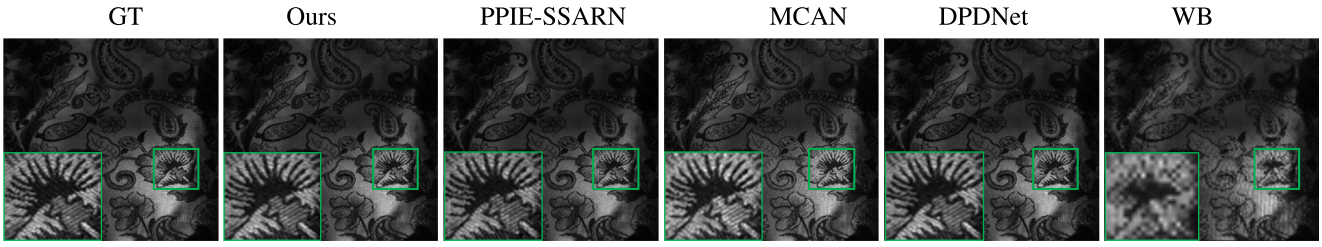
V. EXPERIMENTS

A. SETTINGS

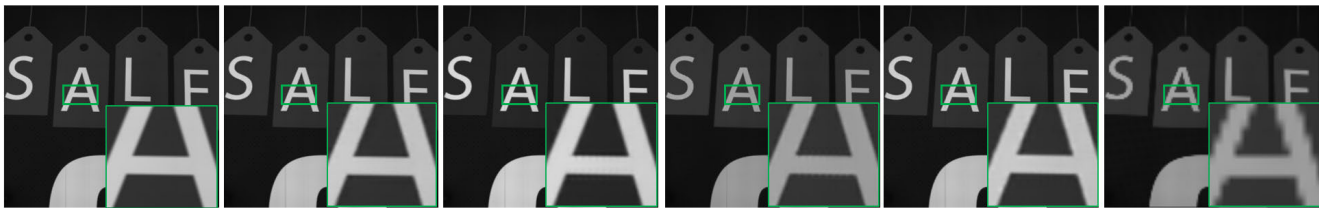
We provide clarification on the training settings, evaluation metric, degradation models, and datasets used in the experiment.

1) DATASETS AND DEGRADATION MODELS

We assess the performance of our framework using two distinct datasets. The first dataset comprises natural spectral images sourced from the ARAD-1K dataset [50]. The second dataset consists of reconstructed spectral images from RGB images obtained from the Vimeo dataset [51]. Spectral reconstruction can be used to increase the quantity of available spectral images for training [52]. To reconstruct the RGB images from the Vimeo dataset we used the MST++ [53]. Furthermore, as part of the preparation for joint demosaicking and super resolution tasks, we apply bicubic interpolation to downscale the spectral images as depicted in Figure 6. The NTIRE 2022 Spectral demosaicking Challenge presents the ARAD-1K dataset, surpassing existing datasets like CAVE [54], TT59 [55], and TokyoTech [27]. ARAD-1K represents a pioneering large-scale dataset tailored specifically for SFA demosaicking of natural scenes, comprising 1000 images with 16 spectral bands covering wavelengths from 400 to 1000 nm. These images provide 16-channel full-resolution spectral data with a spatial resolution of 480 × 512, serving as ground truth (GT). Raw mosaic images are generated from the GT using a 4 × 4 SFA pattern with no dominant band. The dataset is partitioned into 900 training images, 50 validation images, and 50 confidential test images without corresponding GT. For our joint demosaicking and super resolution approach. We utilize the 900 images with GT for training and 50 images for testing to quantitatively evaluate our model.

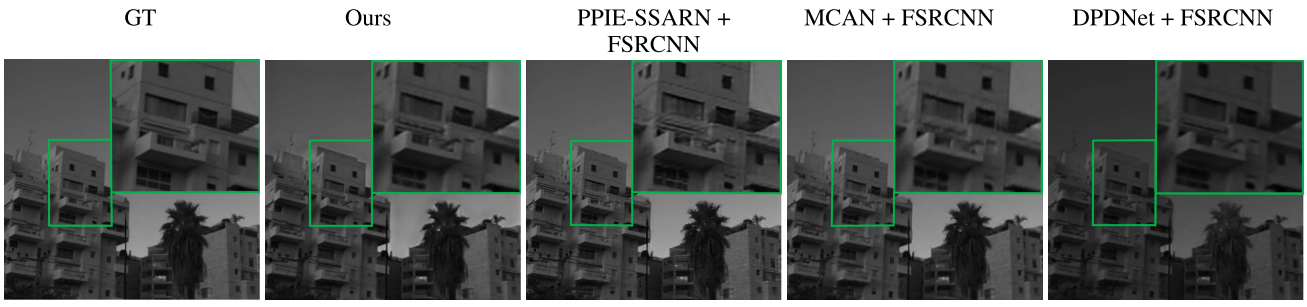


(a) Demosaicking results at 680nm



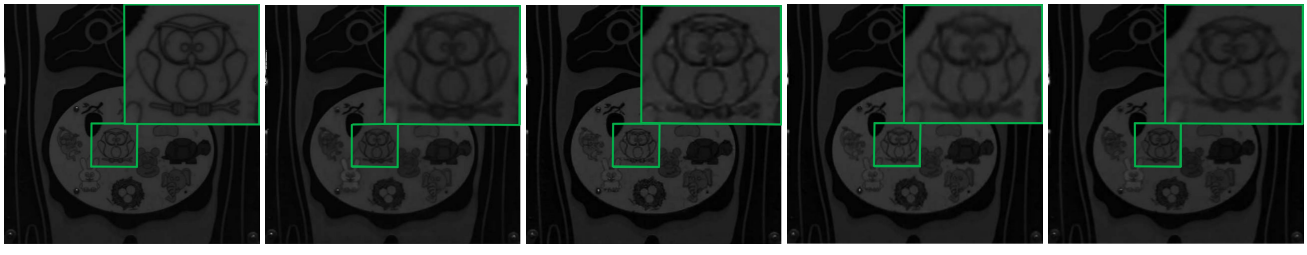
(b) Demosaicking results at 650nm

FIGURE 7. Visual comparison of demosaicking results on the ARAD-1K dataset, showcasing our method alongside PPIE-SSARN [31], MCAN [29], DPD-Net [46], and Weighted Bilinear [17]. The first column presents the ground truth (GT), followed by zoomed-in regions for detailed analysis. While both our method and recent competitors do not exhibit visible artifacts, our method demonstrates superior spectral fidelity, closely matching the ground truth. This aligns with the quantitative results in Table 1.



(a) Results at 400nm

(b) Results at 620nm.



(c) Results at 530nm

FIGURE 8. Comparison of visual demosaicking and super resolution (×2 upscale) results using the ARAD-1K dataset between our method, PPIE-SSARN [31], MCAN [29], DPD-Net [46] each coupled with FSRCNN. The initial column displays the ground truth (GT), supplemented by enlarged views of specific regions. Our method demonstrates the least visible artifacts and superior visual quality, with spectral fidelity closely matching the ground truth. These qualitative results complement the quantitative metrics in Table 2. For finer details, please zoom in accordingly.

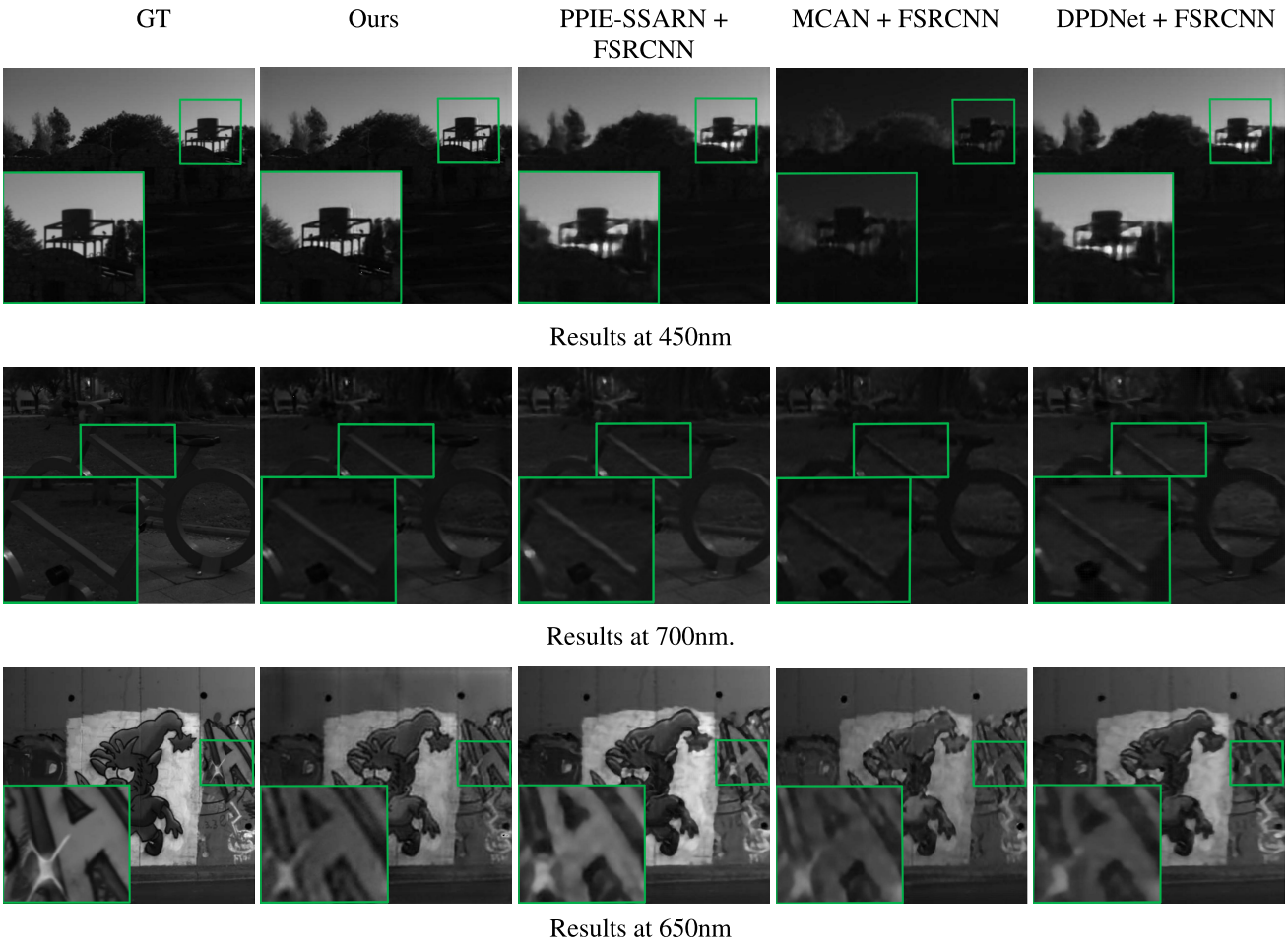


FIGURE 9. Comparison of visual demosaicking and super resolution (×4 upscale) results using the ARAD-1K dataset between our method and competitors (PPIE-SSARN, MCAN, DPD-Net), each coupled with FSRCNN. The initial column displays the ground truth (GT), supplemented by enlarged views of specific regions. Finer details in the results from competitors appear blurry, indicating limitations in preserving high-frequency textures and spatial accuracy during the super resolution process. In contrast, our method achieves sharper details and maintains spectral fidelity. For finer details, please zoom in accordingly.

The Vimeo-90K Dataset [51] is a substantial repository of high-quality videos employed for diverse video processing tasks. Captured using professional-grade cameras, the dataset spans a wide array of subjects. Each frame undergoes independent compression to maintain accuracy and prevent artifacts arising from video codecs. The resolution of all videos is standardized to 448 × 256 pixels. Furthermore, subsets of this dataset serve as benchmarks for various enhancement tasks such as super resolution among others. Additionally, employing the spectral reconstruction model MST++, we reconstructed spectral images from RGB, spanning wavelengths from 400 to 700 nm across 16 bands, with a step of 20 nm. Raw mosaic images are generated from GT using the same 4 × 4 SFA pattern with no dominant band

2) EVALUATION METRICS

We utilize a set of quantitative evaluation metrics to assess the performance of our method. Specifically, we employ the peak signal-to-noise ratio (PSNR) [56], structural similarity index (SSIM) [57], and spectral angle mapper (SAM) [58]. PSNR is employed to quantify the pixel-by-pixel disparity between the demosaicked and super-resolved spectral images and the ground truth (GT). SSIM evaluates the degradation in image quality between the generated images and the GT. Additionally, SAM is utilized to measure the spectral similarity between the generated images and the GT. It is important to note that while PSNR and SSIM tend to increase with higher image quality, SAM's performance decreases as image quality improves.

3) TRAINING DETAILS

Our framework is implemented based on PyTorch, and trained our model using a NVIDIA RTX 2080 GPU equipped with 20GB of VRAM and CUDA 11.8 for a total of 2000 epochs. We employed the Adam optimizer to update model parameters, specifying $\beta_1 = 0.9$ and $\beta_2 = 0.999$. Data augmentation techniques were applied to enhance the robustness of the training process, encompassing random

cropping of spectral images from both the ARAD-1K and Vimeo90k datasets to 128×128 patches, as well as rotations by 90°, 180°, and 270°. Moreover, the batch size was set to 16. To facilitate convergence, we initialized the learning rate to 10^{-4} and adopted a halving strategy every 250 epochs. Moreover, All testing experiments are implemented using the same machine: Intel Core i9-11900k CPU 2.40 GHz, NVIDIA GPU RTX 3070 and 16 Gb of RAM.

B. COMPARISONS WITH STATE-OF-THE-ART

In this subsection, we conduct a comparative analysis of our joint demosaicking-super resolution model against several state-of-the-art demosaicking methods, including MCAN [29], PPIE-SSARN [31], DPD-Net [46], and a conventional demosaicking approach known as WB [17] (weighted bilinear). Each of these methods is paired with super resolution models SRCNN [59], FSRCNN [47] and ESRGAN [60]. Testing is performed on the ARAD-1K dataset, where we implement and fine-tune existing demosaicking models over this latter. Additionally, evaluation is conducted on the SIDQ dataset, which serves as an unseen dataset for all networks, including our own, enabling assessment of our model generalization capabilities.

Furthermore, leveraging the versatility of our upscaling module (see Figure 4), we train our DRDmSR model to perform $2\times$ and $4\times$ upscaling, in addition to $1\times$ upscaling (solely demosaicking). This comprehensive approach allows us to thoroughly evaluate the performance and scalability of our model across various upscaling factors and showcasing the difference between a joint solution and a sequential one.

1) DEMOSAICKING EVALUATION

Table 1 highlights the performance of our model along with state of the arts models for ARAD-1K dataset. Moreover, even though our focus lies on joint demosaicking and super resolution, our model's standalone demosaicking capabilities are notable, demonstrated by its second-place ranking in PSNR, closely following the leading model, PPIE-SSARN. However, our model excels particularly in additional quality metrics such as SAM and SSIM, providing a more comprehensive evaluation of image fidelity and perceptual quality. Our model outperforms others in these metrics, demonstrating superior spatial accuracy and structural similarity. This highlights the effectiveness of our approach in preserving finer details and textures during the demosaicking process, which is crucial for downstream tasks such as image analysis and reconstruction.

Additionally, Figure 7 illustrates visual comparisons between state of the art methods. Conventional WB performs unsatisfactorily in most cases as the estimated images are oversmoothed and blurred. In contrast, deep learning models demonstrate superior performance and are closely related. MCAN exhibits checkerboard distortion caused by the periodic feature extraction operation and has the lowest spectral fidelity among the deep learning models (see first row of Figure 7). DPDNet achieves impressive demosaicking results, yet some undesired streak artifacts persist at the edges, as observed in the third row. PPIE-SARN exhibits the best visual quality alongside ours, although our method demonstrates slightly better spectral fidelity (see Figure 7 third row). Overall, our demosaicking results are performing well, even considering that our primary objective is joint demosaicking and super resolution.

TABLE 1. Average demosaicking results on the ARAD-1K dataset.

Methods	PSNR ↑	SAM ↓	SSIM ↑
WB	36.58	12.27	0.9212
DDM-net	44.81	4.786	0.9869
MCAN	46.46	3.970	0.9903
PPIE-SSARN	**48.46**	<u>3.175</u>	<u>0.9936</u>
Ours	<u>48.02</u>	**2.355**	**0.9954**

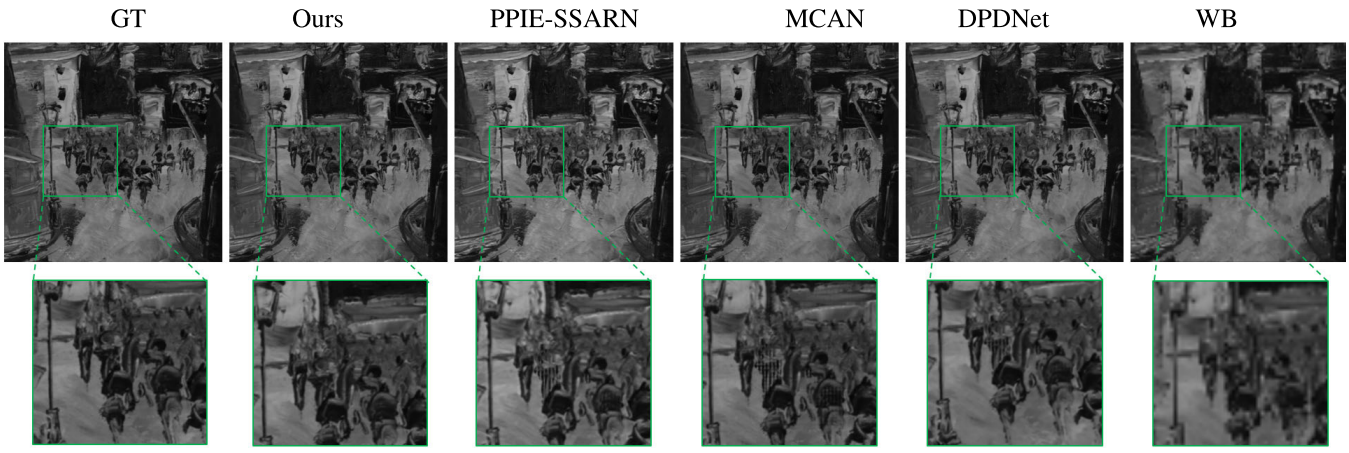
2) DEMOSAICKING AND SUPER RESOLUTION EVALUATION

The findings presented in Table 2 encapsulate the performance metrics of demosaicking methods (PPIE-SSARN, MCAN, DPD-net and WB) integrated with super resolution deep learning based models (SRCNN, FSRCNN, ESRGAN), delineating a sequential approach alongside our proposed joint solution.

TABLE 2. Average demosaicking + super resolution ×2 upscale results on the ARAD-1K dataset.

Method	PSNR ↑	SAM ↓	SSIM ↑
WB + SRCNN	25.53	13.11	0.750
WB + FSRCNN	24.47	11.32	0.9212
WB + ESRGAN	26.19	10.37	0.9245
DPD-Net + SRCNN	29.72	9.95	0.9361
DPD-Net + FSRCNN	30.01	9.14	0.9365
DPD-Net + ESRGAN	32.03	8.50	0.9404
MCAN + SRCNN	30.53	10.60	0.922
MCAN + FSRCNN	31.17	10.74	0.921
MCAN + ESRGAN	32.45	9.97	0.934
PPIE-SSARN + SRCNN	34.06	6.14	<u>0.967</u>
PPIE-SSARN + FSRCNN	34.88	6.19	0.963
PPIE-SSARN + ESRGAN	<u>35.13</u>	<u>6.109</u>	0.9381
Ours	**35.26**	**5.691**	**0.985**

Notably, our joint solution consistently surpasses other sequential methods across all metrics, demonstrating superior image quality and spectral fidelity. Visual comparisons, as depicted in Figure 8, further elucidate the efficacy



(a) Demosaicking results at 400nm

FIGURE 10. Comparison of visual demosaicking results using the SIDQ dataset between our method, PPIE-SSARN [31], MCAN [29], and DPD-Net [46]. The initial column displays the ground truth (GT), supplemented by enlarged views of specific regions. All competitors exhibit visible checkerboard artifacts, while our method eliminates such distortions, producing smoother and more accurate reconstructions. For finer details, please zoom in accordingly.

of our approach compared to sequential methodologies. This effect becomes exacerbated when upscaling, leading to the manifestation of undesirable artifacts. Conversely, MCAN exhibits amplified checkerboard patterns and blur, exacerbating unwanted artifacts that were already present in the demosaicking process, further accentuating them post super resolution. DPD-net, while not displaying obvious artifacts, experiences blurring and yields the lowest metrics among deep learning-based sequential approaches.

In addition to the demosaicking models, FSRCNN surpasses SRCNN in most cases, showcasing better performance in terms of image quality and detail retention. However, ESRGAN outperforms both FSRCNN and SRCNN, at the cost of a high number of parameters and GFLOPs, resulting in significantly longer inference times (see Table 4). In contrast, our joint solution proposes better results at a much lower computational cost and reduced inference time. This highlights the efficiency and effectiveness of our approach, which not only enhances image quality but also maintains computational feasibility.

In contrast, our joint demosaicking super resolution solution delivers the most best results (refer to Figure 8), preserving intricate details such as those observed in the depiction of the bird (see Figure 8 third row), devoid of artifacts and maintained spectral fidelity.

Transitioning to $\times 4$ upscaling, a discernible gap between our joint solution and sequential methodologies becomes apparent. Table 3 presents metrics-based performance evaluations, revealing an amplified gap between PPIE-SARN + ESRGAN and our joint solution. Furthermore, Figure 9 provides visual comparisons, showcasing significantly enhanced detail representation in our joint solution compared to sequential approaches (refer to Figure 9, third row).

In conclusion, our observations indicate that unwanted artifacts originating from the demosaicking step become more pronounced when subjected to upscaling, despite the utilization of state-of-the-art super resolution models such as SRCNN, FSRCNN and ESRGAN. Additionally, it is noteworthy that the gap between our solution and state-of-the-art methods is amplified when augmenting the upscaling factor. This underscores the importance of opting for a joint solution, as demonstrated by our approach, as the optimal strategy for mitigating artifacts in the context of SFA demosaicking and super resolution.

TABLE 3. Average demosaicking + super resolution $\times 4$ upscale results on the ARAD-1K dataset.

Method	PSNR ↑	SAM ↓	SSIM ↑
WB + SRCNN	20.87	21.21	0.610
WB + FSRCNN	20.83	21.226	0.583
WB + ESRGAN	20.88	20.89	0.609
DPD-Net + SRCNN	22.56	15.74	0.693
DPD-Net + FSRCNN	22.60	15.51	0.669
DPD-Net + ESRGAN	23.25	15.21	0.681
MCAN + SRCNN	22.047	16.38	0.674
MCAN + FSRCNN	22.063	16.13	0.649
MCAN + ESRGAN	23.097	15.89	0.667
PPIE-SSARN + SRCNN	25.24	10.38	0.952
PPIE-SSARN + FSRCNN	25.39	9.82	0.954
PPIE-SSARN + ESRGAN	<u>26.18</u>	<u>8.57</u>	<u>0.959</u>
Ours	**26.29**	**8.12**	**0.961**

C. UNSEEN DATASET COMPARISON

In this sub-section, we undertake the evaluation of our proposed framework's performance using unseen spectral

TABLE 4. Efficiency and parameter comparison with different demosaicking + SR methods.

Method	Scale	Params (M)	GFLOPs (G)	Running Time (ms)
DPD-net +SRCNN	×2	1.07	174.27	2.50
DPD-net +FSRCNN	×2	1.09	171.13	1.51
DPD-net +ESRGAN	×2	16.7	1331.3	122.51
MCAN +SRCNN	×2	1.94	124.46	2.19
MCAN +FSRCNN	×2	1.98	120.85	1.20
MCAN +ESRGAN	×2	17	1279.85	123.20
PPIE-SSARN +SRCNN	×2	1.49	275.7	3.14
PPIE-SSARN +FSRCNN	×2	1.6	273.1	2.15
PPIE-SSARN +ESRGAN	×2	16.9	1430.1	124.15
Ours	×2	1.32	181.34	1.45

images. Notably, neither our network nor other state-of-the-art models have encountered the chosen dataset during their training phase. To this end, we employed the SIDQ dataset [61], a meticulously assembled spectral image database comprising nine scenes. These scenes were thoughtfully chosen to represent a diverse variety of materials, including textile, wood, and skin, and were captured utilizing a hyperspectral system (HySpex VNIR-1600). The spectral range spans from 410 to 1000 nm, encompassing 160 spectral bands, with 85 bands falling within the visible light spectrum. Notably, we selectively extracted 16 bands from this dataset, ensuring their alignment with the wavelength range of the ARAD-1K dataset for comprehensive comparison and analysis. Our data processing pipeline follows a consistent methodology, as depicted in Figure 6, where the high-resolution, fully defined spectral image serves as our ground truth. This image is initially downscaled using bicubic degradation and then simulated into a mosaic image, which is subsequently processed by our framework.

The subsequent Table 5 presents a summary of results obtained from our framework and other state-of-the-art models for demosaicking, × 2 upscaling, and × 4 upscaling.

We observe a significant improvement in performance when transitioning from the ARAD-1K dataset to the SIDQ dataset, with our model consistently outperforming all others, even in the demosaicking process for the latter. This notable enhancement can be attributed to our model's additional training on the spectral-reconstructed Vimeo dataset [51], leveraging the state-of-the-art reconstruction model from RGB [53]. This augmented dataset introduces a wide range of diverse imaging conditions, including variations in illumination, textures, and color distributions, which significantly enhance our model's ability to generalize and ensures robust adaptation to unseen datasets.

Although synthetic data cannot fully replicate real-world conditions, it expands the training set, helping to address the limited amount of spectral dataset compared to RGB dataset [52]. The observed performance gain—particularly in terms of PSNR, SSIM, and SAM on SIDQ—demonstrates our model's ability to generalize across diverse imaging conditions effectively. While PPIE-SSARN demonstrates great performance in demosaicking, the sequential models of PPIE-SSARN + ESRGAN fail to outperform other sequential models in terms of metrics. Interestingly, DPD-net + ESRGAN exhibits superior performance among sequential models, particularly evident when upscaling by a factor of ×4.

Figure 10 presents visual comparisons among various demosaicking models, including our own, with a focus on specific areas of interest. Artifacts are distinctly visible in different demosaicking solutions, indicating variations in performance. MCAN exhibits prominent checkerboard artifacts (Figure 10, third column), which are notably more pronounced compared to those observed in the ARAD-1K dataset. Conversely, DPD-net also displays similar artifacts but to a lesser extent; however, it produces the blurriest results among deep learning approaches. PPIE-SSARN showcase excellent performance but still exhibits artifacts in certain areas. Our model stands out as the best performer, showcasing a complete absence of such artifacts (refer to Figure 10).

D. ABLATION STUDY

In this part, we conducted studies to look at the structures and effects of the RIR structure and the effect of the pre-demosaicking used method. To investigate the impacts of every element in the proposed model, we conducted studies to look at the basic blocks features extraction module, where we compare two types of RIR blocks: RCAB [44] and RRDB [60]. In order to provide an equitable comparison, we adjusted the total number of the two fundamental blocks to maintain comparable parameters across all networks (see Table 6) and trained both using the same hyperparameters as described in section V-A3. The network with RCAB blocks performs better than the RRDB at the same model size. Moreover, the choice of the pre-demosaicking method is

TABLE 5. Average results over the SIDQ dataset.

Method	Scale	PSNR ↑	SAM ↓	SSIM ↑
WB	-	25.99	8.08	0.905
DPD-net	-	37.01	4.23	0.912
MCAN	-	38.82	4.14	0.905
PPIE-SSARN	-	41.77	4.27	0.927
Ours	-	**42.14**	**3.14**	**0.969**
WB +SRCNN	×2	25.47	13.32	0.756
WB +FSRCNN	×2	24.36	13.64	0.755
WB + ESRGAN	×2	26.45	13.33	0.749
DPD-net +SRCNN	×2	29.34	9.85	0.895
DPD-net +FSRCNN	×2	29.19	10.16	0.894
DPD-Net + ESRGAN	×2	29.16	9.97	0.722
MCAN +SRCNN	×2	28.35	10.90	0.888
MCAN +FSRCNN	×2	28.22	11.27	0.884
MCAN +ESRGAN	×2	28.37	10.81	0.898
PPIE-SSARN +SRCNN	×2	30.28	10.77	0.923
PPIE-SSARN +FSRCNN	×2	30.18	10.08	0.922
PPIE-SSARN +ESRGAN	×2	30.45	9.23	0.931
Ours	×2	**31.28**	**8.47**	**0.959**
WB +SRCNN	×4	17.37	15.76	0.507
WB +FSRCNN	×4	20.92	16.17	0.487
WB +ESRGAN	×4	22.88	14.89	0.609
DPD-net +SRCNN	×4	27.33	11.96	0.892
DPD-net +FSRCNN	×4	27.37	12.33	0.882
DPD-net + ESRGAN	×4	27.75	10.51	0.861
MCAN +SRCNN	×4	26.50	12.98	0.846
MCAN +FSRCNN	×4	26.49	13.62	0.826
MCAN + ESRGAN	×4	27.09	11.99	0.887
PPIE-SSARN +SRCNN	×4	25.18	12.76	0.852
PPIE-SSARN +FSRCNN	×4	25.24	13.45	0.842
PPIE-SSARN +ESRGAN	×4	26.68	10.57	0.872
Ours	×4	**28.20**	**9.67**	**0.935**

TABLE 6. Performance comparison of different basic block over the ARAD-1K dataset for 2× upscale.

Basic Block	Amounts	Total Parameters	PSNR ↑
RRDB	3	1,649,515	35.11
RCAB	20	1,752,427	35.26

crucial for the final output. To this end we tested our model performance using WB [17], Bicubic Interpolation (BicI) and Bilinear interpolation (BilI). Table 7 shows that using BicI performs better.

TABLE 7. Performance comparison of different basic block over the ARAD-1K dataset for 2× upscale.

Interpolation	PSNR ↑
Bilinear	35.07
WB	35.20
Bicubic	35.26

VI. CONCLUSION

In this article, we introduced a joint demosaicking and super resolution network specifically tailored for SFA images. At the core of our contribution lies the deep residual demosaicking and super resolution module, which seamlessly integrates demosaicking and super resolution tasks within a unified framework. Our model exhibited great performance in demosaicking task, having comparable performance in PSNR and better spectral fidelity compared to state-of-the-art SFA demosaicking models for the ARAD-1K dataset. Furthermore, our investigation revealed an expanding disparity between our joint solution and sequential approaches as the upscaling factor increased, further emphasizing the superiority of the joint solution over sequential methods for SFA images by generating more accurate results and better spectral fidelity. Importantly, leveraging spectral reconstruction from RGB datasets for additional training data enriches the generalization capabilities of our network. This augmentation yields notable performance enhancements, as observed in our evaluation on previously unseen datasets, where our model consistently outperforms existing state-of-the-art methods both quantitatively and qualitatively in demosaicking task and demosaicking + super resolution (×2 and ×4 upscale).

An appealing direction for future research would be to investigate the integration of joint demosaicking, denoising and super resolution in an end-to-end framework. Hence, substantial improvements in the spectral images visual quality can be made by combining the advantages of the three methods.

REFERENCES

[1] W. Su and D. Sun, "Multispectral imaging for plant food quality analysis and visualization," *Comprehensive Rev. Food Sci. Food Saf.*, vol. 17, no. 1, pp. 220–239, Jan. 2018.
[2] U. G. Mangai, S. Samanta, S. Das, P. R. Chowdhury, K. Varghese, and M. Kalra, "A hierarchical multi-classifier framework for landform segmentation using multi-spectral satellite images—A case study over the Indian subcontinent," in *Proc. 4th Pacific-Rim Symp. Image Video Technol.*, Nov. 2010, pp. 306–313.
[3] J. Dias Junior, A. Backes, and M. Escarpinati, "Detection of control points for UAV-multispectral sensed data registration through the combining of feature descriptors," in *Proc. 14th Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl.*, 2019, pp. 444–451.
[4] P.-J. Lapray, X. Wang, J.-B. Thomas, and P. Gouton, "Multispectral filter arrays: Recent advances and practical implementation," *Sensors*, vol. 14, no. 11, pp. 21626–21659, Nov. 2014.

[5] H. K. Aggarwal and A. Majumdar, "Single-sensor multi-spectral image demosaicing algorithm using learned interpolation weights," in *Proc. IEEE Geosci. Remote Sens. Symp.*, Jul. 2014, pp. 2011–2014.

[6] X. Xu, Y. Ye, and X. Li, "Joint demosaicing and super-resolution (JDSR): Network design and perceptual optimization," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 968–980, 2020.

[7] R. Zhou, R. Achanta, and S. Süsstrunk, "Deep residual network for joint demosaicing and super-resolution," in *Proc. Color Imag. Conf.*, vol. 26, Nov. 2018, pp. 75–80.

[8] M. Shoeiby, M. A. Armin, S. Aliakbarian, S. Anwar, and L. Petersson, "Mosaic super-resolution via sequential feature pyramid networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 378–387.

[9] K. Chang, H. Li, Y. Tan, P. L. K. Ding, and B. Li, "A two-stage convolutional neural network for joint demosaicking and super-resolution," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 7, pp. 4238–4254, Jul. 2022.

[10] C.-Y. Tsai and K.-T. Song, "A new edge-adaptive demosaicing algorithm for color filter arrays," *Image Vis. Comput.*, vol. 25, no. 9, pp. 1495–1508, Sep. 2007.

[11] F.-L. He, Y. F. Wang, and K.-L. Hua, "Self-learning approach to color demosaicking via support vector regression," in *Proc. 19th IEEE Int. Conf. Image Process.*, Sep. 2012, pp. 2765–2768.

[12] J. Sun and M. F. Tappen, "Separable Markov random field model and its applications in low level vision," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 402–407, Jan. 2013.

[13] X. Wang, J.-B. Thomas, J. Y. Hardeberg, and P. Gouton, "Discrete wavelet transform based multispectral filter array demosaicking," in *Proc. Colour Vis. Comput. Symp. (CVCS)*, Sep. 2013, pp. 1–6.

[14] G. J. Verhoeven, P. F. Smet, D. Poelman, and F. Vermeulen, "Spectral characterization of a digital still camera's NIR modification to enhance archaeological observation," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 10, pp. 3456–3468, Oct. 2009.

[15] G. Tsagkatakis, M. Bloemen, B. Geelen, M. Jayapala, and P. Tsakalides, "Graph and rank regularized matrix recovery for snapshot spectral image demosaicing," *IEEE Trans. Comput. Imag.*, vol. 5, no. 2, pp. 301–316, Jun. 2019.

[16] L. Zhuang, M. K. Ng, X. Fu, and J. M. Bioucas-Dias, "Hy-demosaicing: Hyperspectral blind reconstruction from spectral subsampling," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5515815.

[17] J. Brauers and T. Aach, "A color filter array based multispectral camera," in *Proc. Workshop Farbbildverarbeitung*, Ilmenau, Germany, 2006, pp. 5–6.

[18] M. Chini, A. Chiancone, and S. Stramondo, "Scale object selection (SOS) through a hierarchical segmentation by a multi-spectral per-pixel classification," *Pattern Recognit. Lett.*, vol. 49, pp. 214–223, Nov. 2014.

[19] S. Mihoubi, O. Losson, B. Mathon, and L. Macaire, "Multispectral demosaicing using intensity-based spectral correlation," in *Proc. Int. Conf. Image Process. Theory, Tools Appl. (IPTA)*, Nov. 2015, pp. 461–466.

[20] S. Mihoubi, O. Losson, B. Mathon, and L. Macaire, "Multispectral demosaicing using pseudo-panchromatic image," *IEEE Trans. Comput. Imag.*, vol. 3, no. 4, pp. 982–995, Dec. 2017.

[21] T. A. Habtegebrial, G. Reis, and D. Stricker, "Deep convolutional networks for snapshot hypercpectral demosaicking," in *Proc. 10th Workshop Hyperspectral Imag. Signal Process., Evol. Remote Sens. (WHISPERS)*, Sep. 2019, pp. 1–5.

[22] Z. Pan, B. Li, H. Cheng, and Y. Bao, "Joint demosaicking and denoising for CFA and MSFA images using a mosaic-adaptive dense residual network," in *Proc. Comput. Vis. Workshops (ECCV)*, Glasgow, U.K. Cham, Switzerland: Springer, Jan. 2020, pp. 647–664.

[23] K. Li, D. Dai, and L. Van Gool, "Jointly learning band selection and filter array design for hyperspectral imaging," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2023, pp. 6373–6383.

[24] P. Amba, J. B. Thomas, and D. Alleysson, "N-LMMSE demosaicing for spectral filter arrays," in *Proc. Color Imag. Conf.*, vol. 61, Jul. 2017, pp. 130–140.

[25] F. Xiong, J. Zhou, and Y. Qian, "Material based object tracking in hyperspectral videos," *IEEE Trans. Image Process.*, vol. 29, pp. 3719–3733, 2020.

[26] Z. Pan, B. Li, H. Cheng, and Y. Bao, "Deep residual network for MSFA raw image denoising," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 2413–2417.

[27] Y. Monno, S. Kikuchi, M. Tanaka, and M. Okutomi, "A practical one-shot multispectral imaging system using a single image sensor," *IEEE Trans. Image Process.*, vol. 24, no. 10, pp. 3048–3059, Oct. 2015.

[28] K. Shinoda, S. Yoshiba, and M. Hasegawa, "Deep demosaicking for multispectral filter arrays," 2018, *arXiv:1808.08021*.

[29] K. Feng, Y. Zhao, J. C. Chan, S. G. Kong, X. Zhang, and B. Wang, "Mosaic convolution-attention network for demosaicing multispectral filter array images," *IEEE Trans. Comput. Imag.*, vol. 7, pp. 864–878, 2021.

[30] Z. Pan, B. Li, Y. Bao, and H. Cheng, "Deep panchromatic image guided residual interpolation for multispectral image demosaicking," in *Proc. 10th Workshop Hyperspectral Imag. Signal Process., Evol. Remote Sens. (WHISPERS)*, Sep. 2019, pp. 1–5.

[31] B. Zhao, J. Zheng, Y. Dong, N. Shen, J. Yang, Y. Cao, and Y. Cao, "PPI edge infused spatial–spectral adaptive residual network for multispectral filter array image demosaicing," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5405214.

[32] M. Bevilacqua, A. Roumy, C. Guillemot, and M.-L.-A. Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," in *Proc. Brit. Mach. Vis. Conf.*, 2012, pp. 135.1–135.10.

[33] R. Timofte, V. De, and L. V. Gool, "Anchored neighborhood regression for fast example-based super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1920–1927.

[34] A. N. Fsian, J.-B. Thomas, J. Y. Hardeberg, and P. Gouton, "Bayesian multispectral videos super resolution," in *Proc. 11th Eur. Workshop Vis. Inf. Process. (EUVIP)*, Sep. 2023, pp. 1–6.

[35] Y. Yoon, H.-G. Jeon, D. Yoo, J.-Y. Lee, and I. S. Kweon, "Learning a deep convolutional network for light-field image super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis. Workshop (ICCVW)*. Cham, Switzerland: Springer, Dec. 2015, pp. 57–65.

[36] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[37] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1646–1654.

[38] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1132–1140.

[39] V. Singh, K. Ramnath, S. Arunachalam, and A. Mittal, "Going much wider with deep networks for image super-resolution," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2020, pp. 2332–2343.

[40] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep Laplacian pyramid networks for fast and accurate super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5835–5843.

[41] O. Sidorov and J. Y. Hardeberg, "Deep hyperspectral prior: Single-image denoising, inpainting, super-resolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 3844–3851.

[42] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2472–2481.

[43] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.

[44] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Jan. 2018, pp. 286–301.

[45] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[46] S. Liu, Y. Zhang, J. Chen, K. P. Lim, and S. Rahardja, "A deep joint network for multispectral demosaicking based on pseudo-panchromatic images," *IEEE J. Sel. Topics Signal Process.*, vol. 16, no. 4, pp. 622–635, Jun. 2022.

[47] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *Proc. 14th Eur. Conf. Comput. Vis. (ECCV)*, Amsterdam, The Netherlands. Cham, Switzerland: Springer, Jan. 2016, pp. 391–407.

[48] V. Dumoulin, J. Shlens, and M. Kudlur, "A learned representation for artistic style," 2016, *arXiv:1610.07629*.

[49] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1874–1883.
[50] B. Arad et al., "NTIRE 2022 spectral demosaicing challenge and data set," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2022, pp. 881–895.
[51] T. Xue, B. Chen, J. Wu, D. Wei, and W. T. Freeman, "Video enhancement with task-oriented flow," *Int. J. Comput. Vis.*, vol. 127, no. 8, pp. 1106–1125, Aug. 2019.
[52] A. Fsian, J.-B. Thomas, J. Hardeberg, and P. Gouton, "Spectral reconstruction from RGB imagery: A potential option for infinite spectral data?" *Sensors*, vol. 24, no. 11, p. 3666, Jun. 2024.
[53] Y. Cai, J. Lin, Z. Lin, H. Wang, Y. Zhang, H. Pfister, R. Timofte, and L. V. Gool, "MST++: Multi-stage spectral-wise transformer for efficient spectral reconstruction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2022, pp. 744–754.
[54] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar, "Generalized assorted pixel camera: Postcapture control of resolution, dynamic range, and spectrum," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2241–2253, Sep. 2010.
[55] Y. Monno, H. Teranaka, K. Yoshizaki, M. Tanaka, and M. Okutomi, "Single-sensor RGB-NIR imaging: High-quality system design and prototype implementation," *IEEE Sensors J.*, vol. 19, no. 2, pp. 497–507, Jan. 2019.
[56] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," *Electron. Lett.*, vol. 44, no. 13, pp. 800–801, Jun. 2008.
[57] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[58] F. A. Kruse, A. B. Lefkoff, J. W. Boardman, K. B. Heidebrecht, A. T. Shapiro, P. J. Barloon, and A. F. H. Goetz, "The spectral image processing system (SIPS)-interactive visualization and analysis of imaging spectrometer data," *Remote Sens. Environ.*, vol. 283, pp. 192–201, Jan. 1993.
[59] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.
[60] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. C. Loy, "ESRGAN: Enhanced super-resolution generative adversarial networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV) Workshops*, Jan. 2019, pp. 63–79.
[61] S. L. Moan, S. George, M. Pedersen, J. Blahová, and J. Y. Hardeberg, "A database for spectral image quality," *Proc. SPIE*, vol. 9396, pp. 225–232, Feb. 2015.

ABDELHAMID N. FSIAN (Student Member, IEEE) received the master's degree in automation and control theory, in 2020, and the master's degree in image processing from the University of Burgundy, in 2022, where he is currently pursuing the Ph.D. degree, with a focus on multispectral imaging, video processing, and computer vision.

JEAN-BAPTISTE THOMAS received the bachelor's degree in applied physics and the master's degree in optics, image, and vision from Université Jean Monnet, France, in 2004 and 2006, respectively, and the Ph.D. degree from Université de Bourgogne, in 2009.

Since 2010, he has been an Associate Professor with the Université de Bourgogne. From 2015 to 2016, he was delegated to CNRS and a Guest Researcher with EPFL. From 2016 and 2021, he was on a sabbatical at NTNU as a Researcher and then an Associate Professor. He has worked extensively on the development of spectral imaging systems through the Spectral Filter Arrays technology for which he has contributed strongly to the definition of the imaging pipeline to enable spectral imaging for computer vision, outside of the laboratories. Since 2016, he has been working on understanding material appearance and its measure by using imaging systems. For more information please visit http://jbthomas.org/.

JON Y. HARDEBERG (Senior Member, IEEE) received the sivilingeniør (M.Sc.) degree in signal processing from Norwegian Institute of Technology, Trondheim, Norway, in 1995, and the Ph.D. degree from the Ecole Nationale Supérieure des Télécommunications, Paris, France, in 1999.

After a short but extremely valuable industry career near Seattle, WA, USA, where he designed, implemented, and evaluated color imaging system solutions for multifunction peripherals and other imaging devices and systems, he returned to academia and Norway, in 2001. He is currently a Professor of color imaging with the Department of Computer Science, Norwegian University of Science and Technology (NTNU). He is a member with the Colourlab, where he teaches, supervises M.Sc. and Ph.D. students, manages international study programs and research projects, and researches in the field of color imaging. He has led several research projects funded by the Research Council of Norway, been NTNU's representative in two Erasmus Mundus Joint Master Degrees (CIMET and COSI), and the coordinator of three Marie Curie ITN projects (CP7.0, ApPEARS, CHANGE). His current research interests include multispectral color imaging, print and image quality, colorimetric device characterization, material appearance, medical imaging, and cultural heritage imaging. He has co-authored more than 300 publications within the field. More information at https://www.ntnu.edu/employees/jon.hardeberg.

PIERRE GOUTON joined the Department of Image Processing, Laboratory of Electronics, Data-Processing and Images, in 1993. Since 2004, he has been the Head of the Computer, Electronic, and Mechanical Department, University of Science and Technology, Dijon, France. He is currently a Professor with the University of Burgundy, France. He is a member of ISIS (a Research Group in Signal and Image Processing of the French National Scientific Research Committee) and also a member of the French Imaging Color Group. His main research interests include the segmentation of images by linear or non-linear methods (morphology, classification), color science, and multispectral images.