Image Demoiréing in RAW and sRGB Domains

Shuning Xu¹, Binbin Song¹, Xiangyu Chen^{1,2}, Xina Liu^{1,2}, and Jiantao Zhou^{1*}

¹ State Key Laboratory of Internet of Things for Smart City,

Department of Computer and Information Science, University of Macau ² Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences {yc07425, yb97426, jtzhou}@um.edu.mo, chxy95@gmail.com, xn.liu95@siat.ac.cn

Abstract. Moiré patterns frequently appear when capturing screens with smartphones or cameras, potentially compromising image quality. Previous studies suggest that moiré pattern elimination in the RAW domain offers greater effectiveness compared to demoiréing in the sRGB domain. Nevertheless, relying solely on RAW data for image demoiréing is insufficient in mitigating the color cast due to the absence of essential information required for the color correction by the image signal processor (ISP). In this paper, we propose to jointly utilize both RAW and sRGB data for image demoiréing (RRID), which are readily accessible in modern smartphones and DSLR cameras. We develop Skip-Connectionbased Demoiréing Module (SCDM) with Gated Feedback Module (GFM) and Frequency Selection Module (FSM) embedded in skip-connections for the efficient and effective demoiréing of RAW and sRGB features, respectively. Subsequently, we design a RGB Guided ISP (RGISP) to learn a device-dependent ISP, assisting the process of color recovery. Extensive experiments demonstrate that our RRID outperforms state-of-the-art approaches, in terms of the performance in moiré pattern removal and color cast correction by 0.62dB in PSNR and 0.003 in SSIM. Code is available at https://github.com/rebeccaeexu/RRID.

Keywords: Image demoiré
ing \cdot RAW domain \cdot Image restoration

1 Introduction

With the increasing prevalence of smartphone cameras, it is common to record on-screen content using smartphone cameras for convenience. However, this type of photography often results in the presence of moiré patterns in the captured images. The appearance of moiré patterns is caused by the spatial frequency aliasing between the camera's color filter array (CFA) and the screen's LCD subpixel [22], resulting in an unsatisfactory visual experience.

Moiré patterns present challenges to be eliminated due to their different scales, indistinct shapes, diverse colors, and varying frequencies. Nowadays, many learning-based image demoiréing approaches [1,5,6,14,15,18,21–23,27,32] have

^{*} Corresponding Author.

2 S. Xu et al.



(c) Demoiréing in RAW and sRGB domains

Fig. 1: Comparison of image demoiréing in different domains. (a) Image demoiréing in sRGB domain. (b) Image demoiréing in RAW domain. (c) Our proposed RRID, which performs image demoiréing in both RAW domain and sRGB domain

been proposed. Among the various methods employed, a commonly used approach is utilizing a multi-scale architecture to remove various sizes of moiré patterns [1,5,18,21,27]. Besides, some works advocate the leverage of frequency components and delicately design frequency-based demoiréing networks [6,14,22,32].

The aforementioned methods are developed to demoiré in the sRGB domain, as depicted in Fig. 1(a). While this scheme seems to be the most direct approach, its effectiveness is limited in removing moiré patterns from complex scenes. The reason for this limitation probably lies in the fact that the nonlinear operations in ISP such as demosaicing deteriorate the moiré patterns originally in the RAW domain. Many schemes consequently suggest that eliminating moiré patterns in the RAW domain is more effective than performing demoiréing in the sRGB domain [28,29]. Due to the easy accessibility of RAW data in modern smartphones and DSLR cameras, image demoiréing in RAW domain can be both feasible and advantageous. RDNet [29] is the pioneering study that investigates image demoiréing in the RAW domain. They explore the demoiréing of RAW images using an encoder-decoder and class-specific learning approach, followed by the utilization of a pre-trained ISP for the RAW-to-sRGB conversion, as depicted in Fig. 1(b). Nevertheless, relying exclusively on RAW data can lead to color cast due to the uncertainty during RAW-to-sRGB conversion.

In order to correct the color cast and make the best use of RAW and sRGB information for the moiré pattern removal, we propose to jointly utilize both RAW and sRGB data for image demoiréing (RRID), as demonstrated in Fig. 1(c). We advocate that introducing paired RAW-sRGB data is advantageous to the moiré pattern removal due to the following reasons: (1) RAW pixels provide more information compared to sRGB pixels since they are typically 12 or 14 bits. (2) Moiré patterns in the RAW domain are less apparent because they are not further affected by nonlinear operations in the ISP. (3) The paired RAW and sRGB data allows the model to learn a device-dependent ISP to aid the process of color recovery. In RRID, we adopt a multi-scale architecture in the Skip-Connectionbased Demoiréing Module (SCDM), to effectively eliminate moiré patterns of different sizes. In order to effectively perform demoiréing on RAW and sRGB features, the Gated Feedback Module (GFM) and Frequency Selective Module (FSM) are integrated into skip-connections within SCDM, as opposed to brutally injecting them into the multi-scale layers. Specifically, for RAW features demoiréing, we introduce the GFM into the skip connection, enabling adaptive differentiation between texture details and moiré patterns through feature gating. Moreover, we design the FSM for the moiré sRGB features, leveraging a learnable band stop filter to mitigate moiré patterns in the frequency domain. Subsequent to the pre-demoiréing operations, we develop the RGB Guided Image Signal Processor (RGISP) to integrate color information from the demoiréd sRGB features with the demoiréd RAW features and learn a device-dependent ISP, facilitating the color recovery process. Ultimately, for the image reconstruction, we utilize multiple Residual Swin Transformer Blocks (RSTBs) [12] and convolutions to accomplish global tone mapping and detail refinement.

In summary, our contributions are listed as follows:

- We propose to exploit image demoiréing in both RAW and sRGB domains. The incorporation of RAW data enhances the moiré pattern removal process, while the inclusion of sRGB data facilitates the RAW-to-sRGB conversion and aids in color restoration of images.
- We develop SCDM with specifically designed modules, GFM and FSM, embeded in the skip-connections to perform efficient and effective demoiréing on the RAW and sRGB features.
- RGISP is designed for the RAW-to-sRGB conversion, incorporating color information from the demoiréd sRGB features with demoiréd RAW features and learn a device-dependent ISP, aiding the color deviation correction.
- Our RRID surpasses state-of-the-arts methods in both qualitative and quantitative evaluations.

2 Related Works

2.1 Image and Video Demoiréing

The purpose of image demoiréing is to restore the original clean image by removing moiré patterns and correcting color deviations from an contaminated image. Many learning-based image demoiréing [1, 6, 15, 18, 22, 23, 32] and video demoiréing methods [2,19,26,28] have been introduced to mitigate the moiré patterns. DMCNN [21] constructs the first real-world dataset for image demoiréing and proposes a multi-resolution neural network to handle moiré patterns across various scales. MopNet [5] devises a multi-scale aggregated network that leverages the edge guidance and pattern attributes for moiré pattern removal. WD-Net [14] employs wavelet transformation to decompose images with moiré patterns into separate frequency bands and establishes a dual-branch network to restore both close-range and far-range information. FHDMi [6] formulates a twostage approach for simultaneous removal of substantial moiré patterns while preserving image details. ESDNet [27] explores a lightweight model aimed at ultrahigh-definition image demoiréing. VDMoiré [2] collects the first video demoiréing dataset and presents a baseline video demoiréing model with a relation-based temporal consistency loss. The aforementioned methods all operate in the sRGB domain. Since it has been observed that moiré patterns are less prominent in the RAW domain, some schemes advocate performing image or video demoiréing in the RAW domain. RDNet [29] introduces the first image demoiréing dataset with RAW data and conducts the demoiréing process in the RAW domain. RawVDemoiré [28] proposes a temporal alignment method for RAW video demoiréing. To remove moiré patterns in the RAW domain, the model have to simultaneously possess the capabilities of demoiréing and and RAW-to-sRGB conversion. However, previous methods encounter challenges in color recovery due to the uncertainty of RAW-to-sRGB conversion. Consequently, we propose an image demoiréing network that utilizes paired RAW-sRGB data, facilitating the color recovery process in image demoiréing.

2.2 Learning-Based RAW Image Processing

RAW pixels offer the potential for leveraging additional information due to their broader bit depth, inherently containing a great wealth of data. Early research in RAW image processing primarily centers on integrating demosaicing and denoising techniques [4,9,10,13]. RTF [9] devices a machine learning approach to the demosaicing and denoising on RAW inputs. SGNet [13] introduces a self-guidance network that leverages the inherent density-map guidance and green-channel guidance for demosaicing and denoising on the RAW images. Recently, some low-level vision tasks endorse the utilization of RAW data, including low-light enhancement [3, 7, 8], super-resolution [16, 25, 30], and reflection removal [11, 20]. The majority of the aforementioned RAW image processing methods only take RAW images as input and generate restored sRGB images as output, indicating that the entire model is responsible for both the restoration and ISP tasks. CR3Net [20] presents a cascaded network that harnesses both the RGB data and their corresponding RAW versions for image reflection removal. Within CR3Net, a module is formulated to emulate point-wise mappings within the ISP, converting the features from RAW to sRGB domain. To further mitigate the color casting problem, our proposed method jointly utilize RAW and sRGB data to learn a device-dependent ISP for image demoiréing.

3 Methodology

Originally caused by frequency interference between the camera's CFA and the screen's LCD subpixel, moiré patterns are further deteriorated by the nonlinear processes in ISP. Therefore, introducing RAW data for image demoiréing may offer greater effectiveness. However, relying solely on RAW data for image demoiréing is insufficient in mitigating color cast due to the uncertainty during the RAW-to-sRGB conversion. Consequently, in this work, we propose to utilize paired RAW and sRGB data, which are readily accessible in modern smartphones and DSLR cameras (*e.g.*, iPhone 15 Pro, HUAWEI P60 Pro). The RAW-sRGB pairs essentially allows the model to learn a device-dependent ISP, aiding the color recovery process. The potential applications in real-world scenarios can be described as follows. When demoiréing function is enabled, the in-device ISP generated sRGB image will not be directly outputted. Instead, together with its corresponding RAW version, will be processed by our RRID to generate a new sRGB image with much suppressed moiré patterns. Afterwards, the RAW image could be discarded, depending on settings of the camera.



Fig. 2: The overview and detailed structures of our proposed RRID.

More specifically, Fig. 2 presents an overview of our proposed RRID for image demoiréing with the paired RAW and sRGB data. As for the RGB branch, the input is an sRGB image $\mathbf{I}_{rgb} \in \mathbb{R}^{H \times W \times 3}$, where $H \times W$ represents the spatial

resolution. Given an input RAW image, it is first packed into the 4-channel RGGB format $\mathbf{I}_{raw} \in \mathbb{R}^{\frac{H}{2} \times \frac{W}{2} \times 4}$, which serves as the input for the RAW branch. RRID is designed to generate the final output sRGB image $(\hat{\mathbf{Y}}_{rgb})$.

Concretely, shallow features \mathbf{F}_{raw} and \mathbf{F}_{rgb} are respectively extracted from the input data \mathbf{I}_{raw} and \mathbf{I}_{rgb} using a convolutional layer and the Dilated Channel Attention Block (DCAB). It is worth noting that a down-sampling layer is applied to the sRGB data to ensure the uniformity of feature shapes. Then, we apply SCDM to acquire pre-demoiréd features \mathbf{D}_{raw} and \mathbf{D}_{rgb} in RAW branch and RGB branch, respectively. GFM is introduced in the skip-connection of SCDM for RAW features, enabling adaptive differentiation between texture details and moiré patterns through feature gating. Similarly, FSM is utilized to to mitigate moiré patterns in the frequency domain. More details will be given in Sec. 3.1. Subsequently, RGISP is designed to realize a RAW-to-sRGB conversion, generating \mathbf{D}_{out} . RGISP incorporates color information from coarsely demoiréd sRGB features \mathbf{D}_{rgb} in a cross-attention mechanism, which will be introduced in Sec. 3.2. For the final image reconstruction, 4 RSTBs [12] are utilized to generate sRGB output $\hat{\mathbf{Y}}_{rgb}$ for their advantages in building long-range dependencies for global tone mapping and detail refinement.

3.1 Design of SCDM

We propose SCDM, a multi-scale architecture designed to remove moiré patterns from shallow RAW features F_{raw} and shallow sRGB features F_{rgb} , as shown in Fig. 2(a). Unlike the vanilla U-Net, SCDM is constructed upon DCAB, facilitating the encoding and decoding of informative features within the respective layers. The features are downsampled twice in the spatial dimension during the encoding process and then upsampled to the original resolution of the input features during decoding. To emphasize the contextual information, we enhance the learned features from the corresponding layers of the encoder with additional modules at skip connections, that is GFM for RAW and FSM for sRGB. The introduction of GFM and FSM in skip-connections is also intended to enhance model efficiency. Then, we can obtain the pre-demoiréd features \mathbf{D}_{raw} and \mathbf{D}_{rab} .

DCAB. DCAB is the essential building block of encoder-decoders for performing RAW demoiréing and the RGB demoiréing, as shown in Fig. 2(b). DCAB is designed using a series of dilated convolution layers followed by ReLU activation and the channel attention mechanism. The receptive field at each block is expanded with the dilated convolution kernels, facilitating the demoiréing operations. Furthermore, by incorporating the channel attention mechanism, DCAB adaptively rescales features according to interdependencies among channels. Also, shortcuts are introduced to establish connections between the input and output of DCAB, enabling the block to learn the residual features.

GFM. Embedded in the skip-connection of SCDM for RAW features, GFM is introduced to enable adaptive differentiation between the texture and moiré patterns through feature gating. The detailed structure of the GFM is shown in Fig. 2(c). For efficiency, we employ point-wise and depth-wise convolutions to aggregate channel and local content information, respectively. Then, the features

are chunked along the channel dimension to generate \mathbf{F}_{gate} and $\mathbf{F}_{content}$. A pointwise multiplication is applied on $\mathbf{F}_{content}$ and \mathbf{F}_{gate} after a GELU activation. During feature gating, we expect that the original image content is adaptively selected and merged along both spatial and channel dimensions. The operations can be formulated as:

$$\{\mathbf{F}_{gate}, \mathbf{F}_{content}\} = \mathrm{DConv}(\mathrm{PConv}(\mathbf{F}_{GFM \ in})), \tag{1}$$

$$\mathbf{F}_{GFM \quad out} = \operatorname{PConv}(\mathbf{F}_{content} \odot \operatorname{GELU}(\mathbf{F}_{gate}) + \mathbf{F}_{GFM \quad in}), \tag{2}$$

where DConv and PConv represent depth-wise convolution and point-wise convolution, respectively. Here, \odot denotes the element-wise multiplication.

FSM. For sRGB demoiréing, we leverage a learnable band stop filter [32] to mitigate moiré patterns in the frequency domain. Considering the formation of moiré patterns is caused by the spatial frequency aliasing between the camera CFA and screen's LCD subpixel, demoiréing in the frequency domain becomes a favorable approach. However, the frequency spectrum of the moiré pattern is often mixed with that of the original contents, thereby making it challenging to disentanglement. Therefore, we propose to perform the disentanglement within small patches, where the moiré patterns tend to be more pronounced in a specific range of frequency bands, facilitating the removal of moiré patterns. Inspired by Zheng et al. [32], we utilize a band stop filter to amplify certain frequencies and suppress others using Block DCT. However, obtaining the frequency prior to separate moiré patterns from normal image textures is challenging and timeconsuming. To address this issue, a set of convolution layers are employed as the learnable band stop filter $B(\cdot)$ to attenuate the specific frequencies of moiré patterns while preserving the original image contents. The process of the FSM can be represented as:

$$\mathbf{F}_{FSM \quad out} = \text{IDCT}(B(\text{DCT}(\mathbf{F}_{FSM \quad in}))), \tag{3}$$

where DCT and IDCT denote the process of Block DCT and the corresponding inverse process. The block size is set to 8×8 for DCT and IDCT.

3.2 Design of RGISP

RGISP is proposed to transform pre-demoiréd RAW features \mathbf{D}_{raw} into sRGB domain \mathbf{D}_{out} . With the assistance of pre-demoiréd sRGB features \mathbf{D}_{rgb} , a device-dependent ISP can be learned in RGISP. Matrix transformation is commonly employed in the traditional ISP pipelines [17]. It enables the enhancement or conversion of image colors to another color space through a channel-wise matrix transformation, which is facilitated by globally shared settings such as environmental illumination and color space specifications [8]. Building upon this principle, we introduce matrix transformation as a means to perform color transformation, as illustrated in Fig. 3. The design of this matrix transformation is motivated by recent advancements in transposed cross-attention. Given $\mathbf{D}_{raw} \in \mathbb{R}^{C \times \frac{H}{2} \times \frac{W}{2}}$, the vectors of query $\mathbf{Q} \in \mathbb{R}^{C \times \frac{H}{2} \frac{W}{2}}$, key $\mathbf{K} \in \mathbb{R}^{\frac{H}{2} \frac{W}{2} \times C}$ are generated through the

8 S. Xu et al.



Fig. 3: The structure of RGISP.

projection with a 1 × 1 convolutional layer Conv and a flatten operation Flatten, where C is the intermediate channel number. Similarly, value $\mathbf{V} \in \mathbb{R}^{C \times \frac{H}{2} \frac{W}{2}}$ can be produced by operations on $\mathbf{D}_{rgb} \in \mathbb{R}^{C \times \frac{H}{2} \times \frac{W}{2}}$ as:

$$\{\mathbf{Q}, \mathbf{K}\} = \text{Flattern}(\text{Conv}(\mathbf{D}_{raw})), \tag{4}$$

$$\mathbf{V} = \text{Flatten}(\text{Conv}(\mathbf{D}_{rgb})). \tag{5}$$

Then, the transformation matrix $\mathbf{M} \in \mathbb{R}^{C \times C}$ is obtained by matrix multiplication. The procedure can be formulated as:

$$\mathbf{M} = \text{Softmax}(\mathbf{Q} \cdot \mathbf{K}^T / \lambda), \tag{6}$$

where a scaling coefficient λ is applied for numerical stability. Subsequently, the vector **V** is transformed by the matrix **M**, performing color space conversion in feature-level. The output feature after color transformation can be obtained by $\mathbf{D}_{out} = \mathbf{M} \cdot \mathbf{V}$. As a complement to the global matrix transformation, we use a depth-wise convolution and a point-wise convolution to refine the local details, generating the output in sRGB domain \mathbf{D}_{out} .

3.3 Loss Function

To effectively remove the moiré patterns and adjust the color deviations, we introduce supervision in both RAW and sRGB domains. The overall training objective can be expressed as:

$$\mathcal{L} = \alpha \| \hat{\mathbf{Y}}_{raw} - \mathbf{Y}_{raw} \|_1 + \| \hat{\mathbf{Y}}_{rgb} - \mathbf{Y}_{rgb} \|_1,$$
(7)

where \mathbf{Y}_{raw} and \mathbf{Y}_{rgb} are the ground-truth RAW and sRGB images, respectively. We empirically set $\alpha = 0.5$.

4 Experiments

4.1 Experimental Setup

Dataset. The experiments are conducted on the RAW image demoiréing dataset published in TMM22 [29], which contains recaptured scenes of natural images,

documents, and webpages. In TMM22, the dataset consists of 540 RAW and sRGB image pairs with ground-truth for training, and 408 pairs for testing. To facilitate the training and comparison process, patches of size 256×256 and 512×512 are cropped respectively for the training and testing sets. The moiré images are captured using four different smartphone cameras and three display screens of varying sizes. In addition, we utilize an image demoiréing dataset FHIMi [6] to verify the generalization ability of our method. FHDMi contains 9981 sRGB image pairs for training and 2019 for testing, featuring diverse and intricate moiré patterns. Notably, images in FHDMi have a higher resolution of 1920×1080 , compared to TMM22.

Training Details. We train RRID using the AdamW optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. A multistep learning rate schedule is employed, with the learning rate initialized at 2×10^{-4} . The RRID model is trained for 500 epochs using a batch size of 80 on 4 NVIDIA RTX 3090 GPUs.

4.2 Quantitative Results

We compare our approach with several image demoiréing methods in sRGB domain, including DMCNN [21], WDNet [14], ESDNet [27], and a RAW image demoiréing method RDNet [29]. We further add comparison with a low-light enhancement method DNF [8] with RAW input, a video demoiréing method RawVDmoiré [28] with RAW input, and a de-reflection method CR3Net [20] with paired RAW-sRGB data. To evaluate the performance of our proposed RRID, we adopt the following three standard metrics to assess pixel-wise accuracy and the perceptual quality: PSNR, SSIM [24], and LPIPS [31]. Additionally, the number of parameters, FLOPs, and inference time are introduced to measure the model complexity. To ensure a fair comparison, we modify the inputs of the sRGB-based models to align with the dataset requirements. Subsequently, all models are fine-tuned using the default settings as provided in their papers on TMM22 dataset. We select the better results from the pretrained and retrained models for comparison, which provides an advantage over the competing methods.

Table. 1 presents the quantitative and the computational complexity comparisons. RRID outperforms the second-best method, RawVDmoiré, by 0.62dB in PSNR and 0.003 in SSIM. As for LPIPS, our methods accomplishes the secondbest result at 0.079, surpassing most of the comparison methods. Even though RRID utilizes both RAW and sRGB data as inputs, our approach can still achieve a satisfactory demoiréing effect with a inference time of merely 0.089s. Compared to previous methods, our RRID demonstrates superiority with acceptable parameters and FLOPs, further demonstrating the superiority of our approach.

To verify the generalization performance of our method, we conduct additional experiments on an image demoiréing dataset FHIMi [6]. In order to adapt to the FHDMi dataset with sRGB data as input, the RAW branch and RGISP are removed. We select methods tailored for image demoiréing: DMCNN [21], Mop-Net [5], MBCNN [32], FHDe²Net [6], and ESDNet [27] for comparison, as shown in Table 2. Although RRID is initially designed for demoiréing on the RAWsRGB image pairs, it still achieves the second-best performance by 24.39dB in

10 S. Xu et al.

Table 1: Quantitative comparison with the state-of-the-art demoiréing approaches and RAW image restoration methods on TMM22 dataset [29] in terms of average PSNR, SSIM, LPIPS and computational complexity. The best results are highlighted with **bold**. The second-best results are highlighted with underline.

					Mat	boda		
Index	Methods							
	DMCNN	WDNet	ESDNet	RDNet	DNF	RawVDmoiré	CR3Net	RRID
	[21]	[14]	[27]	[29]	[8]	[28]	[20]	(Ours)
# Input type	$\mathbf{s}\mathbf{R}\mathbf{G}\mathbf{B}$	$\mathbf{s}\mathbf{R}\mathbf{G}\mathbf{B}$	$\mathbf{s}\mathbf{R}\mathbf{G}\mathbf{B}$	RAW	RAW	RAW	sRGB+RAW	sRGB+RAW
$\mathrm{PSNR}\uparrow$	23.54	22.33	26.77	26.16	23.55	27.26	23.75	27.88
$SSIM\uparrow$	0.885	0.802	0.927	0.921	0.895	0.935	0.922	0.938
$LPIPS\downarrow$	0.154	0.166	0.089	0.091	0.162	0.075	0.102	0.079
Params (M)	<u>1.55</u>	3.36	5.93	6.04	1.25	5.33	2.68	2.38
TFLOPs	0.102	0.055	0.141	0.161	0.013	0.022	0.883	0.093
Inference time (s)	0.052	0.284	0.115	1.094	0.070	0.182	0.058	0.089

Table 2: Quantitative comparison on FHDMi dataset [6].

Method	$\mathrm{PSNR}\uparrow$	$\mathrm{SSIM}\uparrow$
DMCNN [21]	21.54	0.773
MopNet [5]	22.76	0.796
MBCNN [32]	22.31	0.810
FHDe ² Net [6]	22.93	0.789
ESDNet [27]	24.50	0.835
Ours	24.39	<u>0.830</u>

PSNR and 0.830 in SSIM. Compared to ESDNet, an image demoiréing network designed for ultra-high-definition sRGB images with more than twice of the parameters, RRID also yields competitive results. The qualitative results on the FHDMi dataset demonstrate that our method can not only perform well in demoiréing on RAW-sRGB pairs but also solve sRGB image demoiréing, exhibiting a certain degree of generalization.

4.3 Qualitative Results

We present visual comparisons among our approach and the existing methods in Fig. 4. The results demonstrate the advantage of RRID in effectively removing moiré artifacts and correcting color deviations. For the first scene demonstrating a recaptured flower with moiré patterns, previous methods tend to exhibit residual moiré patterns in their restored images. While ESDNet successfully eliminates a significant portion of these patterns, it struggles to accurately restore the original colors of the image. Apart from natural scenes, we also present a case of webpages contaminated with moiré patterns. Through the utilization of both RAW and sRGB data, our approach eliminates moiré patterns while preserving fine image details, even in scenarios characterized by severe color deviations. This ensures a more accurate restoration of the original colors and results in

(b) DMCNN (c) WDNet (d) ESDNet (e) RDNet (a) Input (g) RawVDmoire (f) DNF (h) CR3Net (j) GT (i) Ours & tall men's hing brands c azon azon azon azon azon at to shop (b) DMCNN (d) ESDNet (a) Input (c) WDNet (e) RDNet & tall men's hing brands c azon azon azon azon azon hat to shop (f) DNF (g) RawVDmoire (h) CR3Net (i) Ours (j) GT

enhanced visual effects. Additional qualitative comparisons can be found in the supplementary file.

Fig. 4: Qualitative comparison on RAW image demoiréing TMM22 dataset [29].

4.4 Ablation Study

1) Ablation study on the model architecture and inputs. In Table 3, we investigate the architecture and inputs of RRID and validate the importance of different individual components in the whole RRID. We conduct this ablation study by comparing the proposed RRID and the following variants of RRID:

- B₁. Both the RAW input and the RAW demoiréing branch are removed. Due to the lack of RAW features, the entire RGISP is also deleted.
- B₂. Both the sRGB input and sRGB demoiréing branch are removed.
- B_3 . The sRGB input is substituted with RAW input. The input channel number is adjusted, while the rest of the architecture is kept unchanged.

- 12 S. Xu et al.
 - *B*₄. We replace RGISP with a two convolutional layers after a concatenation of the RAW and sRGB features.
 - B_5 . Reconstruction module is replaced with two convolutional layers.
 - B_6 . The full model of RRID with both RAW and sRGB images as input.

Table 3: Comparison of baseline models for the evaluation of architecture and inputs.

Models	$\mathrm{PSNR}\uparrow$	$\mathrm{SSIM}\uparrow$
B_1 (w/o RAW in and RAW branch)	25.79	0.915
B_2 (w/o sRGB in and sRGB branch) B_3 (w/o sRGB in)	27.24 26.51	$0.929 \\ 0.923$
B ₄ (w/o RGISP)	27.38	0.932
B_5 (w/o Reconstruction)	26.93	0.926
B_6 (full version)	27.88	0.938

Our proposed RRID leverages paired RAW-sRGB data as input. Removing the RAW input and its corresponding RAW demoiréing module (B_1) results in a drastic decrease in PSNR of approximately 2dB. Likewise, eliminating the sRGB input and its corresponding RGB demoiréing module (B_2) leads to a PSNR of only 27.24dB. In addition to the decline in numerical metrics, we can also assess



Fig. 5: Visualization of the ablation study for architecture and inputs comparison.

the visual performance from Fig. 5. When relying only on sRGB data and its corresponding demoiréing modules, moiré patterns are less effectively removed, as depicted in Fig.5(b). Conversely, Fig.5(c) demonstrates that using only RAW data as input generates images with more pronounced color deviations. To further validate the significance of RAW input for image demoiréing, we compare B_3 to B_1 by solely removing the RAW input data while retaining the remaining

13

model structures. Although B_3 demonstrates some improvements in numerical metrics compared to B_1 , its performance remains unsatisfactory due to the lack of information provided by the RAW input and the fact that the original RAW demoiréing module is specifically designed for RAW data. Additionally, we conducted further comparisons by removing the originally designed RGISP (B_4) and Reconstruction modules (B_5), resulting in inferior numerical metrics and performance in Fig. 5 compared to the complete RRID model (B_6).

2) Ablation study on SCDM. SCDM is designed to perform effective demoiréing on both RAW and sRGB features. More specifically, we introduce GFM for RAW features demoiréing and FSM for sRGB features demoiréing.

Table 4: Ablation study on SCDM.

Models	$\mathrm{PSNR}\uparrow$	$\rm SSIM\uparrow$
S_1 (w/o GFM & FSM)	27.00	0.926
$S_2 \ (w/o \ GFM)$	27.17	0.927
$S_3 \ (w/o \ FSM)$	27.39	0.930
S_4 (all GFM)	27.50	<u>0.933</u>
S_5 (all FSM)	27.32	0.930
$S_6 \ (w/o \ DCAB)$	26.55	0.923
S_7 (SCDM)	27.88	0.938

Table 4 demonstrates the effectiveness and indispensability of the meticulously designed GFM and FSM integrated within the skip-connections. Removing the demoiréing modules on the skip-connections (S_1-S_3) individually result in varying degrees of performance decline of the model. For instance, removing both GFM and FSM leads to only 27.00 dB in PSNR. Furthermore, Fig.6(b) demonstrates that the model's ability to remove moiré patterns decreases upon removing the GFM and FSM modules. Furthermore, to assess the specificity of GFM and FSM, we substitute the original modules in S_4 and S_5 , respectively. Despite no reduction in the number of parameters, the PSNR decreases by approximately 0.3-0.5dB. Since DCAB is a crucial component module in SCDM, we substitute DCAB with two convolutional layers to confirm its necessity. Fig. 6(c) showcases more residual moiré patterns and significant color deviation compared to our complete SCDM presented in Fig.6(d).



Fig. 6: Visualization of the ablation study for SCDM.



Fig. 7: Visualization of the ablation study for RGISP.

3) Ablation study on RGISP. RGISP is designed to integrate color information from the coarsely demoiréd sRGB features during the ISP stage, facilitating the color recovery process. The process of the RGISP can be recognized as a cross-attention mechanism. To assess the efficacy of the RGISP design, we compare the following structures in Table 5. For R_1 , we concatenate the pre-demoiréd RAW features with the sRGB features and subsequently apply a self-attention mechanism. In R_2 , the pre-demoiréd RAW features undergo a self-attention mechanism and are then concatenated with the pre-demoiréd sRGB features. Table 5 shows that our proposed RGISP surpasses the self-attention mechanisms (R_1 , R_2), as well as other RAW-to-sRGB conversion methods RRM [20] (R_3). In addition, Fig. 7 depicts a webpage with globally distributed moiré pat-

Table 5: Ablation study on RGISP.

Models	$\mathrm{PSNR}\uparrow$	$\mathrm{SSIM}\uparrow$
$ \frac{R_1 \text{ (self-atten([raw, rgb]))}}{R_2 \text{ ([self-atten(raw), rgb])}} \\ \frac{R_3 \text{ (RRM)}}{R_4 \text{ (RGISP)}} $	27.54 27.42 27.21 27.88	0.933 0.931 0.929 0.938

terns. R_1 - R_4 are all capable of effectively removing the moiré patterns, but R_4 exhibits the best performance in correcting color deviation. With the introduction of sRGB pre-demoiréd features, RGISP is able to learn a device-dependent ISP and restore the original colors of the image effectively.

5 Conclusion

In this paper, we propose to jointly utilize RAW and sRGB data for image demoiréing (RRID). Firstly, we introduce SCDM with GFM and FSM embedded in skip-connections to handle the demoiréing of RAW and sRGB features, respectively. Additionally, we present RGISP to learn a device-dependent ISP, aiding the color recovery process. Extensive experiments demonstrate the superior performance of our RRID method in both moiré pattern removal and color cast correction. Our approach outperforms state-of-the-art methods by 0.62dB in PSNR and 0.003 in SSIM.

Acknowledgements

This work was supported in part by Macau Science and Technology Development Fund under SKLIOTSC-2021-2023, 0072/2020/AMJ, 0022/2022/A, and 0014/2022/AFJ; in part by Research Committee at University of Macau under MYRG-GRG2023-00058-FST-UMDF and MYRG2022-00152-FST; in part by Natural Science Foundation of Guangdong Province of China under EF2023-00116-FST.

References

- Cheng, X., Fu, Z., Yang, J.: Multi-scale dynamic feature encoding network for image demoiréing. In: 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). pp. 3486–3493. IEEE (2019)
- Dai, P., Yu, X., Ma, L., Zhang, B., Li, J., Li, W., Shen, J., Qi, X.: Video demoireing with relation-based temporal consistency. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 17622–17631 (2022)
- Dong, X., Xu, W., Miao, Z., Ma, L., Zhang, C., Yang, J., Jin, Z., Teoh, A.B.J., Shen, J.: Abandoning the bayer-filter to see in the dark. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 17431– 17440 (2022)
- Gharbi, M., Chaurasia, G., Paris, S., Durand, F.: Deep joint demosaicking and denoising. ACM Transactions on Graphics (ToG) 35(6), 1–12 (2016)
- He, B., Wang, C., Shi, B., Duan, L.Y.: Mop moire patterns using mopnet. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 2424–2432 (2019)
- He, B., Wang, C., Shi, B., Duan, L.Y.: Fhde 2 net: Full high definition demoireing network. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXII 16. pp. 713–729. Springer (2020)
- Huang, H., Yang, W., Hu, Y., Liu, J., Duan, L.Y.: Towards low light enhancement with raw images. IEEE Transactions on Image Processing 31, 1391–1405 (2022)
- Jin, X., Han, L.H., Li, Z., Guo, C.L., Chai, Z., Li, C.: Dnf: Decouple and feedback network for seeing in the dark. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 18135–18144 (2023)
- Khashabi, D., Nowozin, S., Jancsary, J., Fitzgibbon, A.W.: Joint demosaicing and denoising via learned nonparametric random fields. IEEE Transactions on Image Processing 23(12), 4968–4981 (2014)
- Kokkinos, F., Lefkimmiatis, S.: Deep image demosaicking using a cascade of convolutional residual denoising networks. In: Proceedings of the European conference on computer vision (ECCV). pp. 303–319 (2018)
- Lei, C., Chen, Q.: Robust reflection removal with reflection-free flash-only cues. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14811–14820 (2021)
- Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., Timofte, R.: Swinir: Image restoration using swin transformer. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 1833–1844 (2021)
- Liu, L., Jia, X., Liu, J., Tian, Q.: Joint demosaicing and denoising with self guidance. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2240–2249 (2020)

- 16 S. Xu et al.
- Liu, L., Liu, J., Yuan, S., Slabaugh, G., Leonardis, A., Zhou, W., Tian, Q.: Waveletbased dual-branch network for image demoiréing. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIII 16. pp. 86–102. Springer (2020)
- Liu, S., Li, C., Nan, N., Zong, Z., Song, R.: Mmdm: Multi-frame and multi-scale for image demoiréing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 434–435 (2020)
- Luo, F., Wu, X., Guo, Y.: And: Adversarial neural degradation for learning blind image super-resolution. Advances in Neural Information Processing Systems 36 (2024)
- 17. Nakamura, J.: Image sensors and signal processing for digital still cameras. CRC press (2017)
- Niu, Y., Lin, Z., Liu, W., Guo, W.: Progressive moire removal and texture complementation for image demoireing. IEEE Transactions on Circuits and Systems for Video Technology (2023)
- Oh, G., Gu, H., Kim, S., Kim, J.: Fpanet: Frequency-based video demoiring using frame-level post alignment. arXiv preprint arXiv:2301.07330 (2023)
- Song, B., Zhou, J., Chen, X., Zhang, S.: Real-scene reflection removal with rawrgb image pairs. IEEE Transactions on Circuits and Systems for Video Technology (2023)
- Sun, Y., Yu, Y., Wang, W.: Moiré photo restoration using multiresolution convolutional neural networks. IEEE Transactions on Image Processing 27(8), 4160–4172 (2018)
- 22. Wang, C., He, B., Wu, S., Wan, R., Shi, B., Duan, L.Y.: Coarse-to-fine disentangling demoiréing framework for recaptured screen images. IEEE Transactions on Pattern Analysis and Machine Intelligence (2023)
- Wang, H., Tian, Q., Li, L., Guo, X.: Image demoiréing with a dual-domain distilling network. In: 2021 IEEE International Conference on Multimedia and Expo (ICME). pp. 1–6. IEEE (2021)
- Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing 13(4), 600–612 (2004)
- Xing, W., Egiazarian, K.: End-to-end learning for joint image demosaicing, denoising and super-resolution. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 3507–3516 (2021)
- Xu, S., Song, B., Chen, X., Zhou, J.: Direction-aware video demoiring with temporal-guided bilateral learning. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 6360–6368 (2024)
- Yu, X., Dai, P., Li, W., Ma, L., Shen, J., Li, J., Qi, X.: Towards efficient and scale-robust ultra-high-definition image demoiréing. In: European Conference on Computer Vision. pp. 646–662. Springer (2022)
- Yue, H., Cheng, Y., Liu, X., Yang, J.: Recaptured raw screen image and video demoireing via channel and spatial modulations. arXiv preprint arXiv:2310.20332 (2023)
- 29. Yue, H., Cheng, Y., Mao, Y., Cao, C., Yang, J.: Recaptured screen image demoiréing in raw domain. IEEE Transactions on Multimedia (2022)
- Yue, H., Zhang, Z., Yang, J.: Real-rawvsr: Real-world raw video super-resolution with a benchmark dataset. In: European Conference on Computer Vision. pp. 608– 624. Springer (2022)

17

- 31. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 586–595 (2018)
- Zheng, B., Yuan, S., Slabaugh, G., Leonardis, A.: Image demoireing with learnable bandpass filters. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3636–3645 (2020)