

# Efficient Unified Demosaicing for Bayer and Non-Bayer Patterned Image Sensors

Haechang Lee<sup>1,4,\*</sup>, Dongwon Park<sup>2,3,\*</sup>, Wongi Jeong<sup>1,\*</sup>,  
Kijeong Kim<sup>4</sup>, Hyunwoo Je<sup>4</sup>, Dongil Ryu<sup>4</sup> and Se Young Chun<sup>1,2,3,†</sup>  
<sup>1</sup>Dept. of ECE, <sup>2</sup>INMC, <sup>3</sup>IPAI, Seoul National University, Republic of Korea,  
<sup>4</sup>SK hynix, Republic of Korea

{harrylee, donglpark, wg7139, sychun}@snu.ac.kr {kijeong1.kim, hyunwoo.je, dongil.ryu}@sk.com

## Abstract

As the physical size of recent CMOS image sensors (CIS) gets smaller, the latest mobile cameras adopt unique non-Bayer color filter array (CFA) patterns (e.g., Quad, Nona,  $Q \times Q$ ), which consist of homogeneous color units with adjacent pixels. These non-Bayer CFAs are superior to conventional Bayer CFA thanks to their changeable pixel-bin sizes for different light conditions, but may introduce visual artifacts during demosaicing due to their inherent pixel pattern structures and sensor hardware characteristics. Previous demosaicing methods have primarily focused on Bayer CFA, necessitating distinct reconstruction methods for non-Bayer CIS with various CFA modes under different lighting conditions. In this work, we propose an efficient unified demosaicing method that can be applied to both conventional Bayer RAW and various non-Bayer CFAs' RAW data in different operation modes. Our Knowledge Learning-based demosaicing model for Adaptive Patterns, namely KLAP, utilizes CFA-adaptive filters for only 1% key filters in the network for each CFA, but still manages to effectively demosaic all the CFAs, yielding comparable performance to the large-scale models. Furthermore, by employing meta-learning during inference (KLAP-M), our model is able to eliminate unknown sensor-generic artifacts in real RAW data, effectively bridging the gap between synthetic images and real sensor RAW. Our KLAP and KLAP-M methods achieved state-of-the-art demosaicing performance in both synthetic and real RAW data of Bayer and non-Bayer CFAs.

## 1. Introduction

Demosaicing (DM) is the process of interpolating single-channel input images into RGB output images within an embedded Image Signal Processor (ISP). With the growing demand for high-quality mobile camera images, CMOS image sensor (CIS) resolution has increased dramatically, even reaching 200 million pixels in the latest smartphones.

However, as image sensors cannot infinitely increase in size, pixel size has been reduced to enhance resolution. Smaller CISs are more vulnerable to noise and degradation in image restoration capabilities because they are more sensitive to variations in light reception, especially in low-light condition [13, 23, 38, 39]. As a result, modern high-end smartphones have started using image sensors that group adjacent homogeneous pixels, resulting in non-Bayer Quad, Nona, and Quad-by-Quad ( $Q \times Q$ ) sensors [19, 38, 41], while still retaining some of the properties of the standard Bayer CFA [5] pattern. Quad, Nona, and  $Q \times Q$  sensors combine the same color pixel arrays of  $2 \times 2$ ,  $3 \times 3$ , and  $4 \times 4$  respectively, resulting in homogeneous pixel units (i.e., Gr, R, B, and Gb) for each sensor, as shown in Fig. 1(a).

Demosaicing for modern non-Bayer CFAs is more complex and computationally demanding than for standard Bayer CFAs. This is because as the number of pixel arrays within each unit increases, the distance between the units becomes greater, requiring interpolation with inaccurate pixel values from distant locations. Therefore, there is growing interest in using deep learning for demosaicing methods, leading to active research on both Bayer pattern demosaicing [64, 42, 12, 7, 56, 1, 34, 62, 45, 18, 26, 17] and non-Bayer pattern demosaicing [28, 27, 20, 3, 46, 10].

However, the aforementioned methods focus on a single CFA DM task and do not cover DM tasks for multiple CFA patterns. Modern mobile phones with non-Bayer patterned CIS adapt their CFA modes dynamically based on lighting conditions, controlled by the CIS's ISP. Using independent models (IMs) for each pattern, tailored to different CFA modes, would demand loading and operating multiple models within the limited circuit space of the CIS. This would result in excessive memory and power consumption if the models were kept standby on the mobile application processor (AP) and switched accordingly. Moreover, the task of tuning models for each CFA would be laborious.

Currently, no existing method can handle dynamically changing CFA modes in a non-Bayer patterned CIS as a unified model (UM). Several recent studies have explored the

\* Equal contribution, † Corresponding author.

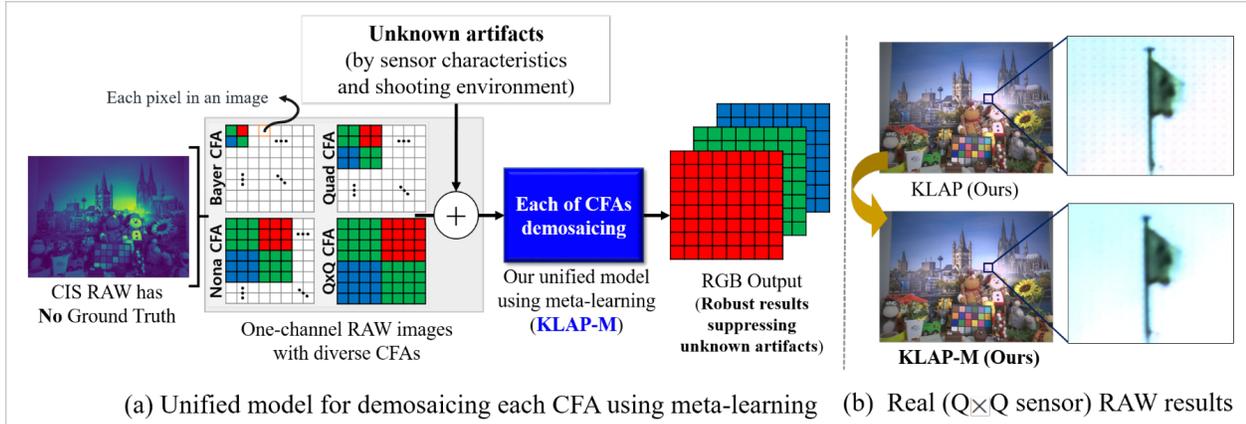


Figure 1: (a) Overview of our unified model (UM) for demosaicing all the Bayer and non-Bayer CFAs, called the **K**nowledge **L**earning-based demosaicing model for **A**daptive **P**atterns using **M**eta-test learning (KLAP-M), even when ground truth is unavailable and unknown artifacts are present. (b) Comparing CIS RAW demosaicing results of KLAP (KLAP-M without meta-test learning) and KLAP-M (KLAP with meta-test learning).

concept of all-in-one image restoration, which deals with multiple types of unknown degradation [31, 9, 30]. However, these existing methods do not fully account for real ‘unknown’ artifacts in real ‘CIS RAW’ restoration process. To address this limitation, where ground truth (GT) may be missing or largely unavailable, we will conduct a comprehensive investigation of these methods. Since such unknown artifacts may fail to yield high-quality photos, we are motivated to propose a UM with robust meta-learning-based DM methods that can handle these obstacles.

In this work, we propose efficient unified demosaicing methods that bridge the gap between synthetic and real CIS RAW images, enabling various non-Bayer CISs through a new pipeline. Our proposed **K**nowledge **L**earning-based demosaicing model for **A**daptive **P**atterns (KLAP) is capable of simultaneously handling multiple CFAs’ demosaicing, which consists of two following steps. Firstly, we train a baseline UM using the two-stage knowledge learning (TKL) [9], making it more efficient to find Adaptive Discriminative filters for each specific CFA Pattern (ADP). Secondly, after TKL, we further fine-tune the UM model using ADP, which is a method to identify a small set of discriminative filters in CNN filters, serving as independent key parameters for each specific CFA demosaicing task. Lastly, we propose KLAP-M, which combines KLAP (TKL+ADP) with Meta-test learning, integrating self-supervised learning to handle domain gaps between synthetic RAW and real CIS RAW caused by unknown artifacts in real-life scenarios. Our proposed meta-test learning consists of pixel binning loss based on CIS domain knowledge and self-supervised denoising techniques. Fig. 1(a) provides an overview of our KLAP-M approach, which handles both Bayer and Non-Bayer patterns. Additionally, Fig. 1(b) shows the results of our meta-test learning technique, addressing the domain gap in real RAW images.

Our contributions are summarized as follows: (1) Our efficient unified network, KLAP, effectively performs demosaicing for multiple CFAs, (2) KLAP-M, a version of KLAP that incorporates a meta-learning approach, effectively reduces unknown visual artifacts in genuine CIS RAW images that are caused by diverse sensor characteristics and shooting environments, (3) KLAP and KLAP-M achieve state-of-the-art performance on the synthetic benchmark dataset and real CIS RAW samples captured by CIS chips.

## 2. Related Works

### 2.1. Deep Learning-based Demosaicing

**IMs for DM only.** Traditional demosaicing without applying deep learning techniques either apply a fixed DM filter to each pixel without considering other parameters as features or utilize spectral and spatial features available in neighboring pixels to interpolate the unknown pixel as closely as possible to the original [36, 15]. Due to the complexity of various CIS CFAs, traditional methods are cumbersome, leading to an increasing interest in deep learning-based demosaicing models. Stojkovic *et al.* [47] suggested IMs of each Bayer and Quad demosaicing based on CDM-Net [11]. Kim *et al.* [28, 27] applied the duplex pyramid network structure to Quad CFA and Nona CFA, respectively. Sharif *et al.* [46] proposed a GAN-based spatial-asymmetric attention for Nona CFA reconstruction. For QxQ CFA, Cho *et al.* [10] proposed an efficient pyramidal network using progressive distillation based on PyNet [18].

**Multi-tasks joint with DM.** There have been proposals to combine DM methods with other closely related ISP tasks, such as denoising (DN) and super-resolution (SR). Some [12, 54, 7, 34, 22, 29] proposed convolutional neural networks approach for joint DM and DN to improve the quality of the restored image. Ma *et al.* [35] and Xu

*et al.* [58] proposed models for simultaneous DM and SR. Xing *et al.* [56] introduced a multi-task learning approach to jointly address three tasks: DM, DN, and SR. Previous studies mainly concentrate on multi-task approaches for single CFA demosaicing and known noise sources. In contrast, our proposed method introduces a unified model that handles both Bayer and non-Bayer CFAs, incorporating meta-learning to ensure robust performance even in the presence of unknown noise.

## 2.2. Image Restoration for Multi-tasks

**IMs for multi-tasks.** Beyond DM tasks, recent papers [63, 37, 61, 53, 52, 43, 8] have introduced various approaches that share a common framework capable of multiple image restoration tasks, including denoising, deblurring, and deraining. While the mentioned IM excels in individual tasks, it necessitates multiple network parameters as multiple networks are needed to handle all the required tasks.

**Unified model (UM) for multi-tasks.** To overcome the drawbacks of IMs, Chen *et al* [9] proposed a single UM for two-stage knowledge learning mechanism based on multi-teacher and single student approach for multiple degradations on images that contains rain, haze, and snow. Li *et al.* [30] proposed a single UM using a contrastive-based degraded encoder, called the degradation-guided restoration network (DGRN), which adaptively works with three degradations: rain, noise, and blur. Park *et al.* [40] introduced a single UM equipped with dedicated filters for degradation, achieving remarkable results in rain-noise-blur and rain-snow-haze tasks. To the best of our knowledge, there is currently no reported method that can handle all Bayer and non-Bayer demosaicing tasks using a single unified model.

## 2.3. Meta-learning-based Image Restoration

For image reconstruction, a large number of samples are usually necessary, but it may not be feasible in many real-world situations. Meta-learning, also known as learn-to-learn, provides a promising solution to the problem of adapting models quickly to new data. This learning method empowers models to achieve efficient task performance even with limited additional incoming data. Finn *et al.* [14] proposed an algorithm for model-agnostic meta-learning that achieved state-of-the-art performance in few-shot learning tasks. Meta-SR [16] enables super-resolution for arbitrary scale factors by applying the Meta-Upscale Module. We propose the use of meta-learning to achieve robust results, even in the presence of unknown artifacts in CIS RAW images.

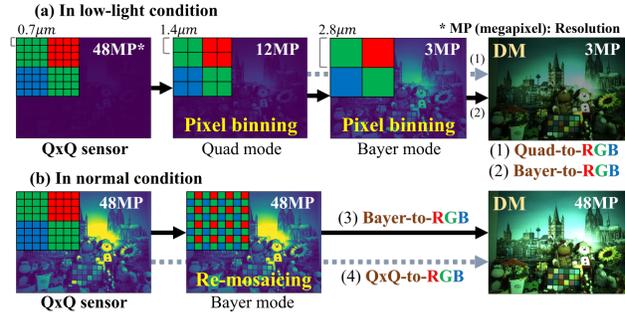


Figure 2: DM scenario in real CIS. For example, in the case of  $Q \times Q$  CIS: (a) In low-light conditions, the  $Q \times Q$  sensor converts its pattern to either the (1) Quad or (2) Bayer mode (pixel-binning), sacrificing resolution, and then performs DM. (b) In normal conditions, the  $Q \times Q$  sensor can either re-mosaic the pattern to the Bayer mode and then perform DM or directly perform  $Q \times Q$  DM, with full resolution.

## 3. Deep Demosaicing for Each Non-Bayer CFA

### 3.1. Operating Principles of Non-Bayer Sensors

With the decreasing size of camera sensors, the physical area of light captured by a pixel has been reduced. Consequently, the introduction of non-Bayer sensors allows for capturing more light. In case of  $Q \times Q$  as an example, when there is sufficient light, as scenario (3) and (4) in Fig. 2,  $Q \times Q$  sensors can handle the entire resolution with Bayer DM (after ‘re-mosaicing’) and direct  $Q \times Q$  DM. On the other hand, especially in low-light conditions,  $Q \times Q$  CIS pixels have the advantage of using ‘pixel binning’ to enhance their light sensitivity and reduce the noise [65, 60], sacrificing their resolution (but still acceptable), resulting in clear image quality with reduced noise (shown as scenario (1) and (2) in Fig. 2). Pixel binning is the merging of neighboring pixels in an image through summation or averaging in ISP, typically done by the ISP after pixel-readout. Quad DM or Bayer DM methods are specifically required in such cases. In recent non-Bayer CIS, supporting diverse CFA pattern modes is crucial; however, using separate DM networks for each pattern increases network parameters, leading to a larger CIS chip area, and multiple DM models require frequent switching, consuming more memory and power in mobile environments. Our proposed unified DM model handles all non-Bayer sensor patterns, including standard Bayer sensors, providing effective solutions for this issue. It offers flexibility for different CIS product lines and CFA pattern modes, reducing product development time with minimal fine-tuning required for specific product characteristics.

### 3.2. Data Synthesis for Demosaicing All CFAs

To train input images resembling real CIS RAW, we propose a data synthesis pipeline that generates realistic RAW-

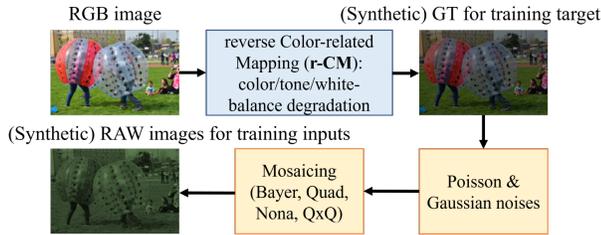


Figure 3: Overview of our realistic RAW image synthesis pipeline for Bayer and Non-Bayer demosaicing. The r-CM (reverse Color-related Mapping functions) towards RAW-like synthesis consists of invertible linear operations that relate RGB color spaces.

like images. Using a high-quality sRGB dataset, we follow the front-end of Fig. 3 to generate synthetic RAW-like images. This involves applying four reverse color-related mapping functions (r-CM) from the ISP chain, including color tone degradation, inverse gamma correction, inverse color correction, and inverse auto white balance correction functions. We analyzed and adjusted the previous ISP chains, resulting in a pipeline structure similar to previous methods. [51, 50, 44, 6, 57]. Using this method, we generate RGB synthetic GT labels for demosaicing training. Furthermore, we add Gaussian and Poisson noise to simulate various types of real noise [44, 13, 6, 57]. Each image is then converted into a mosaic pattern for Bayer, Quad, Nona, and  $Q \times Q$  CFA, as depicted in the bottom row of Fig. 3. This process generates the training inputs. The reverse color mapping (r-CM) consists of linear operations and can be easily "re-reversed" to obtain the original color mapping (CM). CM generates final output images that resemble realistic images as perceived by humans. Our proposed data synthesis pipeline considers demosaicing for both Bayer and Non-Bayer patterns and incorporates a realistic noise that combines Gaussian and Poisson noise. More detailed information is Sec. S.1 in the supplementary material.

### 3.3. Domain Gap in Synthetic and Real CIS RAW

Synthetic data-trained models often struggle with real data due to the domain gap issue, a persistent problem in image restoration tasks [44, 6, 21]. The domain gap arises from variations in sensor hardware characteristics due to differences in circuit structure, manufacturing processes, and component variations across CIS brands and product lines. The upper image in Fig. 1(b) shows visual artifacts in real CIS RAW, mainly caused by crosstalk effects [24, 25, 32] between inner and outer pixels (details in Sec. S.2 in the supplementary). Moreover, unknown artifacts can emerge in different shooting environments and vary across CIS types. To address this, we propose a meta-learning method to minimize the domain gap, enabling the effective handling of unexpected unknown artifacts.

## 4. Unified Deep Demosaicing for Multiple Bayer and Non-Bayer CFAs

Fig. 4 displays the proposed single unified DM method for all Bayer and non-Bayer sensor patterns (KLAP) and its additional meta-learning during inference framework for robustness (KLAP-M). In Step 1 as Fig. 4(a), our approach augments the network capacity of the integrated model using the Two-stage Knowledge Learning [9] (TKL). This maximizes the effectiveness of the Adaptive Discriminative filters for each specific CFA Pattern (ADP) discovered in the subsequent step. In Step 2 as Fig. 4(b), we further enhance the UM using a small number of specialized network kernels for each DM task. Lastly, as Fig. 4(c), we introduce a meta-test framework that ensures robust DM output in the presence of unknown artifacts.

### 4.1. Step 1: Two-stage Knowledge Learning

This step aims to train the unified DM model for all CFAs using the two-stage knowledge learning [9] (TKL), with independent DM models for each CFA (IMs). The IMs, with the same network architecture, have independent network parameters  $(\{\theta_i\}_{i=1}^k)$  dependent on each CFA-specific DM task ( $i$ ), where  $k$  represents the total number of tasks, which is 4 in our case.  $\theta_{um}$  represents the network architecture and its network parameters of UM. IM achieves high performance as a specialized model for each task, but requires a model  $k$  times larger than UM ( $\theta_{um}$ ).

First, we pre-train each individual IM based on NAFNet [8] since its IMs outperform existing DM methods, as shown in IMs' comparison in Tab 1. Then, in the knowledge collection (KC) stage, set the IMs specialized for each CFA DM task as the teacher network and UM as the student network to learn and collect knowledge from the teacher. In the knowledge examination (KE) stage after KC, train only using the student network and GT labels without guidance from the teacher network. We applied TKL method to increase the model's capacity after feature-level guidance for each CFA pattern, in order to maximize the effect of top filter detection in the next step (see 'TKL' results in Tab. 1).

### 4.2. Step 2: Adaptive Discriminative Filters for each specific CFA Pattern

Xie *et al.* [55] proposed FAIG, which can detect discriminative filters of specific degradation. FAIG measures integrated gradient (IG) [48, 49] between baseline and target models. Inspired by FAIG and its application in another domain [40], we applied CNN for Adaptive Discriminant filters for each specific CFA Pattern (ADP) using the leveraged FAIG method. FAIG score is as follows:  $FS_j = FAIG_j(\theta_{um}^j, \theta_i^j, x_i)$ , for multiple CFA filters  $i = 1, \dots, k$  and kernel index  $j$ . Once the FAIG scores are computed, they are then ranked in descending order. The top  $q\%$  of

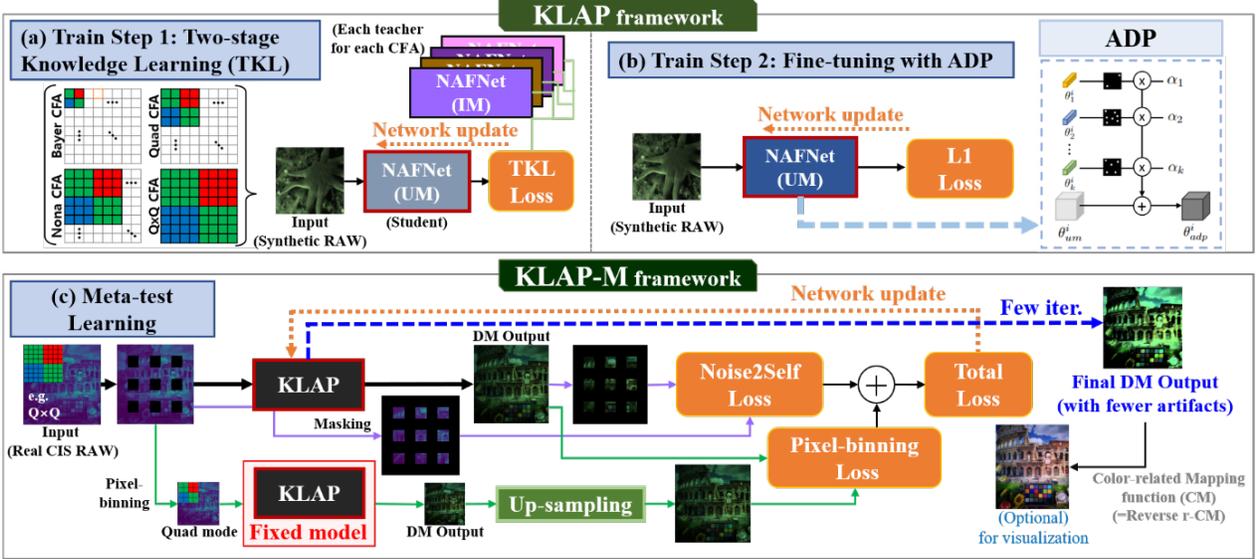


Figure 4: The overview of our proposed unified DM model, Knowledge Learning-based demosaicing model for Adaptive Pattern (KLAP) and KLAP with Meta-test learning (KLAP-M). KLAP consists of 2 steps: (a) two-stage knowledge learning (TKL) for training baselines, (b) fine-tuning using Adaptive Discriminant filters for each specific CFA Pattern (ADP). (c) KLAP-M employs meta-learning to reduce unknown artifacts in real RAW images during inference.

kernels are selected for each demosaicing process, with  $q$  representing a fixed value between 0.5 and 5.

We propose ADP, implemented by the masks  $M_c$  that are selected kernels using FAIG as illustrated in Fig. 4 and defined as follows:

$$\theta_{adp}^j = \theta_{um}^j + \sum_{i=1}^k \alpha_i \theta_i^j * M_i^j \quad (1)$$

where  $j$  is kernel index,  $*$  is point-wise multiplication,  $\theta_{um}^j$  refers to the pre-trained integrated model in Step 1, and  $\alpha_i$  is a coefficient for a specific CFA pattern and is set up either as 1 or 0. Note that in a real non-Bayer CIS on a mobile device, the pattern mode  $\alpha_i$  is determined by the mobile AP after detecting the lighting conditions. Also  $\theta_i^j$  is an additional kernel for specific CFA pattern. The ratio  $q$  in the mask is determined empirically to be 1%. For example, the ratio of 1% in the mask is 1% for 4 demosaicing types, our proposed method uses an additional 4% of the entire network parameters as compared to the baseline UM. More detailed information is in the supplementary Sec. S.3. Our proposed KLAP, a combination of TKL and ADP, achieves state-of-the-art performance in various CFA DM tasks by replacing only relevant CNN kernels in UM from TKL.

### 4.3. Meta-learning during Inference

As shown in Fig. 4, we propose meta-learning during inference (meta-test learning) to mitigate unknown artifacts caused by sensor characteristics or shooting environments.

By performing a few network updates during inference, this approach produces robust results. Our proposed meta-learning during inference consists of pixel binning loss and Noise2Self [4] (N2S) loss, one of the self-supervised denoising techniques. As mentioned in Sec. 3.1, pixel binning compensates for resolution loss by increasing the light sensitivity, thus reducing noise. Based on CIS domain knowledge, we propose a self-supervised denoising method using a pixel binning loss to remove unknown artifacts.

$$\mathcal{L}_{pix} = |G(x_{J^c}, \theta_{adp}) - U(G(m(x_{J^c}), \theta_{adp'}))|, \quad (2)$$

where  $x$  and  $J^c$  denote the CIS RAW data and mask used by N2S,  $m$  and  $U$  represent average-based pixel-binning operation and up-sampling operation, respectively.  $G$  is a unified network structure and  $\theta_{adp}$  is network parameters of ADP.  $\theta_{adp'}$  is the initial network parameters that are not updated.

Additionally, we apply modified N2S loss to maintain robustness against noise (Poisson and Gaussian noise) that may occur depending on the shooting environment and to prevent blur caused by pixel binning loss:

$$\mathcal{L}_{N2S} = |G(x_{J^c}, \theta_{adp})_J - x_J| \quad (3)$$

where  $x_J$  and  $x_J^c$  are represent independent images using the mask scheme. Additional information about the pixel binning loss and N2S loss can be found in the supplementary material.

The total loss for meta-learning during inference is as follows:

$$\mathcal{L}_{\text{total}} = \lambda_{\text{pix}} \mathcal{L}_{\text{pix}} + \lambda_{N2S} \mathcal{L}_{N2S} \quad (4)$$

where  $\lambda_{\text{pix}}$  and  $\lambda_{N2S}$  are used to balance different loss conditions and is experimentally found through visualization.

## 5. Experimental Results

As stated in Sec. 3.2, we generate synthetic DF2K Bayer and Non-Bayer CIS (DF2K-CIS) dataset, a combination of two open source datasets, DIV2K [2] and Flickr2K [33]. The training set comprises 2,500 images, with a validation set of 50 images and a test set of 1,000 images. Furthermore, we propose to use the DF2K-CIS test dataset with strong noise to evaluate the effectiveness of our proposed meta-test learning in generating robust results. The DF2K-CIS strong noise test dataset comprises 200 images with noise parameters four times larger than those used in training. Then, we evaluate our proposed meta-learning method, KLAP-M, using 7 Q×Q CIS RAW images (48MP) with a resolution of 8000×6000, Quad CIS RAW images [59] with a resolution of 1200×1600, and 3 Bayer CIS RAW images (50MP) with a resolution of 8192×6144, all of which are 10-bit images captured directly by each type of CIS chip. In the meta-test learning, KLAP-M is trained using the loss function in Eq.(4) with  $\lambda_{\text{pix}} = 1$  and  $\lambda_{N2S} = 0.02$ . Note that Meta-test (KLAP-M) does not utilize IMs but instead employs a unified model, and we conducted KLAP-M evaluations on each new full image for each sensor type. More implementation details and demosaicing RAW results can be found in the supplementary materials Sec. S.5 and S.6.

### 5.1. Results on Synthetic RAW Dataset

#### 5.1.1 Comparison of Ablation Studies and KLAP with Other Methods

**Ablation study for KLAP.** We perform ablation studies on the proposed KLAP approach based on NAFNet [8], including TKL and ADP, as shown in Fig. 4 (a) and (b), using the DF2K-CIS test dataset. Tab. 1 summarized the performance of PSNR (dB) and the number of parameters (Million). Baseline UM is a simple integrated model trained on all tasks, while IMs require 4 times more network parameters than Baseline UM. Baseline UM-Large (Baseline UM-L) refers to a modified version of NAFNet [8] with increased network blocks. In TKL and ADP in the table, each step is independently applied to the baseline UM. TKL-to-IM refers to the fine-tuned IMs after applying TKL.

Using TKL and ADP independently leads to only a marginal improvement of 0.05 dB and 0.11 dB, respectively, compared to Baseline UM. Our proposed KLAP (TKL+ADP) further improved performance by 0.4 dB with a slightly increased number of network parameters compared to Baseline UM. Notably, Our KLAP achieved signif-

Table 1: Ablation studies and quantitative performance comparison (PyNetQ×Q [10], Kim [27], Uformer [53], Chen [9] and Li [30]) for KLAP (TKL+ADP) on DF2K-CIS test dataset. PyNetQ×Q [10], Kim [27] and Uformer [53] are independent models, while Chen [9] and Li [30] are unified models. The methods highlighted in purple are based on NAFNet [8]. Baseline-UM is a simple unified model. TKL refers to the TKL-applied Baseline UM, and ADP refers to the ADP-applied Baseline UM independently. TKL-to-IM refers to the fine-tuned IM using TKL. Avg. denotes mean of all CFA’s PSNR (dB), and Par. denotes the required number of parameters (Million).

| Method             | Ba.          | Qu.          | No.          | QxQ          | Avg.         | Par.        |
|--------------------|--------------|--------------|--------------|--------------|--------------|-------------|
| PyNetQxQ [10]      | 37.08        | 37.38        | 36.65        | 36.44        | 36.89        | 4.2         |
| Kim [27]           | 41.33        | 40.81        | 39.85        | 37.02        | 39.75        | 13.8        |
| Uformer [53]       | 41.89        | 41.19        | 40.60        | 40.74        | 41.11        | <b>83.0</b> |
| IM                 | 42.18        | 41.80        | 41.14        | 41.42        | 41.64        | <b>68.4</b> |
| <b>TKL-to-IM</b>   | <b>42.36</b> | <b>41.89</b> | <b>41.58</b> | <b>41.60</b> | <b>41.86</b> | <b>68.4</b> |
| Baseline UM        | 41.90        | 41.40        | 41.03        | 41.09        | 41.35        | 17.1        |
| Baseline UM-L      | 41.95        | 41.44        | 41.08        | 41.13        | 41.40        | 19.4        |
| Chen [9]           | 41.43        | 40.89        | 40.54        | 40.49        | 40.84        | 28.7        |
| Li [30]            | 38.28        | 38.08        | 38.23        | 36.94        | 37.88        | 7.6         |
| TKL                | 41.89        | 41.44        | 41.11        | 41.15        | 41.40        | 17.1        |
| ADP                | 42.06        | 41.50        | 41.14        | 41.16        | 41.46        | 17.8        |
| <b>KLAP (Ours)</b> | <b>42.25</b> | <b>41.75</b> | <b>41.42</b> | <b>41.41</b> | <b>41.71</b> | <b>17.8</b> |

icantly higher performance than Baseline UM-L (41.71dB vs. 41.40dB) with fewer parameters (17.8M vs. 19.4M). In addition, fine-tuning each IM with pre-trained TKL resulted in a notable improvement compared to the original IMs, attributed to the inclusion of contrastive learning loss in TKL. Our proposed KLAP method significantly improves demosaicing performance for all CFAs with small parameters.

**Comparisons among other unifying methods.** We evaluate the performance of our KLAP with NAFNet [8] on a DF2K-CIS test dataset and summarize the results in Tab. 1 in terms of PSNR (dB) and the number of parameters. We use the official codes provided by the authors of Airnet [30] and Chen [9]. The Chen [9] method uses the MSBDN-based TKL method. Despite a slight increase in network parameters by 0.7M (about 4%) in NAFNet, our KLAP yields significantly improved performance by 0.4 dB compared to the IM method. Notably, our KLAP yields the highest PSNR among all-in-one methods [9, 30] while using smaller network parameters compared to existing methods applied to NAFNet networks. Fig. 5 shows DM results on synthetic datasets for visual comparisons. We adjust CM in Sec. 3.2 for better visualization. The images on the 1st to 4th rows are input synthetic RAW images and their DM outputs of UM, Chen [9], Li [30], and our KLAP are on the 2nd, 3rd, 4th, 5th column of Fig. 5, respectively. This shows that our KLAP outperforms other state-of-the-art unifying methods on DF2K-CIS test datasets.

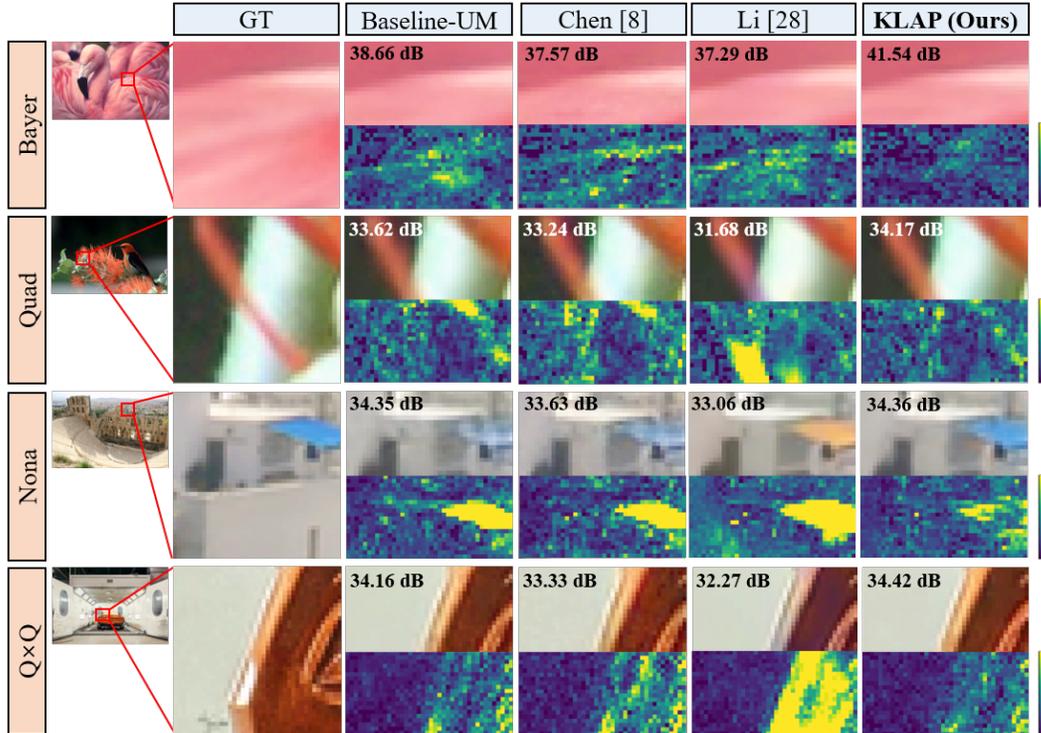


Figure 5: Comparison of demosaiced results (**top**) and their difference maps (**bottom**) on the synthetic RAW (DF2K-CIS) test set produced by different methods. The PSNR (dB) values displayed in the top-left corner are for the entire image. As shown, our proposed KLAP demonstrates the best performance. Note that CM is applied to the DM outputs for visualization.

### 5.1.2 Performance and Selected Filter Locations

To demonstrate the superiority of FAIG [55] over random selection, we evaluate various mask selection strate-

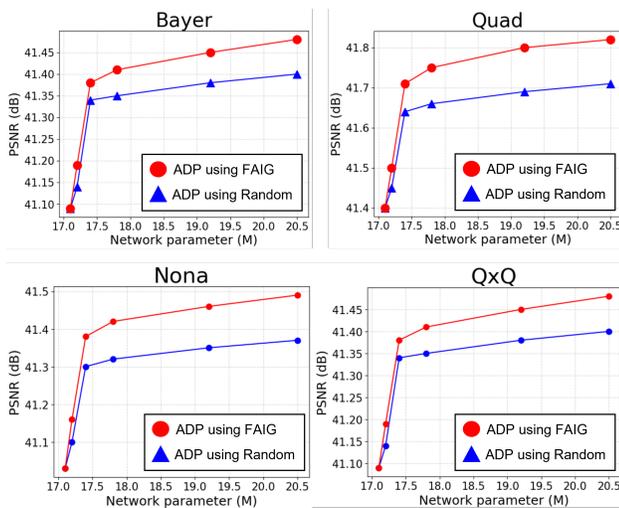


Figure 6: Performance comparisons among different filter location selections (0%, 0.1%, 0.5%, 1%, 3%, and 5%, respectively) for UM with ADP: Random selection method and FAIG adjusting ADP on DF2K CIS test dataset.

gies in our ADP method on synthetic datasets with both Bayer and non-Bayer patterns. The mask selection ratios are set to 0.1%, 0.5%, 1%, 3%, and 5%. After TKL, we add adaptive network kernels in proportion to the  $q$  ratio. Two mask selection methods are investigated: random selection and the FAIG method introduced in Sec. 4.2. Fig. 6 summarizes our results, indicating that our ADP outperforms random filter selection, underscoring the effectiveness of selecting discriminative filters for each CFA DM task. This implies that discriminative filters can be defined as task-specific filters, rather than randomly selected filters.

### 5.1.3 Analysis of Robustness in Strong Noise

To assess KLAP-M’s robustness, we evaluate it on the DF2K-CIS ‘strong noise’ test dataset (4 times larger noise parameters than the training set) and present the results in Tab. 2. KLAP exhibits slightly better performance than existing methods, and with KLAP-M, we observe an average improvement of 1.8 dB in PSNR with just 10 iterations.

## 5.2. Results on Real CIS RAW

We evaluate the performance of our KLAP with meta-learning on a real CIS RAW dataset and present the results in Fig. 7. The number of iterations for meta-learning is fixed

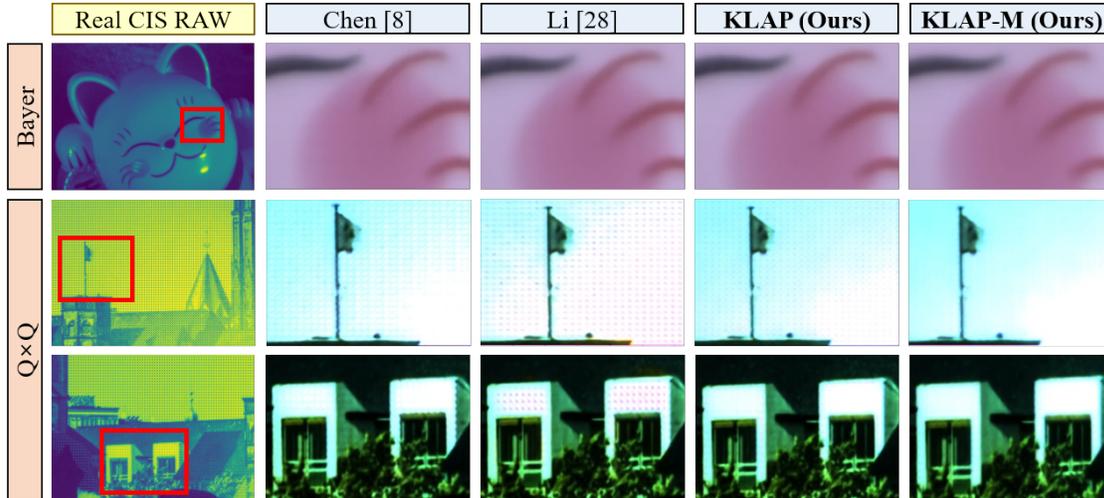


Figure 7: Qualitative DM results on the real CIS RAW. Note that KLAP with meta-test learning (KLAP-M) shows robust performance in real CIS RAW, despite of existence of sensor-generic unknown artifacts. Note that CM is applied to the DM outputs for visualization.

Table 2: Performance comparisons among different methods to assess robustness on the DF2K-CIS test dataset with strong noise. The noise parameters used here are four times larger than those used in the training, and KLAP-M employs a fixed number of 10 meta-test iterations.

| CFA   | Chen [9] | Li [30] | KLAP  | KLAP-M       |
|-------|----------|---------|-------|--------------|
| Bayer | 32.60    | 31.61   | 32.98 | <b>33.32</b> |
| Quad  | 32.48    | 31.58   | 32.93 | <b>35.41</b> |
| Nona  | 32.44    | 31.64   | 32.88 | <b>35.06</b> |
| Q×Q   | 32.45    | 31.38   | 32.86 | <b>35.41</b> |

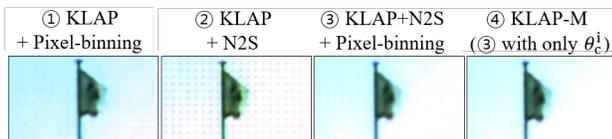


Figure 8: Ablation study of KLAP-M. The comparison shows the effect of each component in meta-test learning.

at 45. Note that CM (as mentioned in Sec. 3.2) is applied to the inference outputs after DM model inference solely for better visualization. In the Bayer case, our method, as well as Chen [9] and Li [30]’s methods, show robust results on real data. However, In the case of demosaicing Q×Q, Chen and Li’s methods are unable to alleviate artifacts, while our method significantly mitigates resulting artifacts during inference by reducing domain gap through meta-learning. Fig. 8 shows the ablation study of KLAP-M and demonstrates superior performance compared to other method combinations. Note that the Bayer output is an image that has been squared by 0.7 from the original outputs for visual comparison purposes. We represent the two

Q×Q output images, with their pixel values (range of 0 to 1) cubed, to compare the artifact mitigation performance with other models. More demosaicing RAW results can be found in the supplementary materials S.6

### 5.3. Limitations

Deep learning-based DM models for CIS can be limited by specialized circuits with embedded AI accelerators. Thorough HW design validation, considering power, performance, and silicon area (PPA), is crucial for on-sensor implementation. Nonetheless, our unified demosaicing model for multiple CFAs is a significant first step towards practical CIS applications.

## 6. Conclusion

Our proposed method employs knowledge distilling and task-specific kernels to demosaic multiple CFAs, integrating a meta-testing framework for efficiency and robustness, resulting in low computational complexity, resilience to unknown artifacts, and high-quality demosaiced images.

## Acknowledgments

This work was supported in part by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) [NO.2021-0-01343, Artificial Intelligence Graduate School Program (Seoul National University)], the National Research Foundation of Korea(NRF) grants funded by the Korea government(MSIT) (NRF-2022R1A4A1030579), Basic Science Research Program through the NRF funded by the

Ministry of Education(NRF-2017R1D1A1B05035810) and Creative-Pioneering Researchers Program through Seoul National University. Also, the authors acknowledged the financial support from the BK21 FOUR program of the Education and Research Program for Future ICT Pioneers, Seoul National University. CIS RAW data and CIS domain knowledge were supported by CIS Development Representative at SK hynix.

## References

- [1] SM A Sharif, Rizwan Ali Naqvi, and Mithun Biswas. Beyond joint demosaicking and denoising: An image processing pipeline for a pixel-bin image sensor. In *CVPR*, 2021.
- [2] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *CVPRW*, 2017.
- [3] Boaz Arad, Radu Timofte, Rony Yahel, et al. NTIRE 2022 spectral demosaicing challenge and data set. In *CVPRW*, 2022.
- [4] Joshua Batson and Loic Royer. Noise2self: Blind denoising by self-supervision. In *ICML*, 2019.
- [5] Bryce E Bayer. Color imaging array. *US Patent 3,971,065*, 1976.
- [6] Tim Brooks, Ben Mildenhall, Tianfan Xue, Jiawen Chen, Dillon Sharlet, and Jonathan T Barron. Unprocessing images for learned raw denoising. In *CVPR*, 2019.
- [7] Jierun Chen, Song Wen, and S-H Gary Chan. Joint demosaicking and denoising in the wild: The case of training under ground truth uncertainty. In *AAAI*, 2021.
- [8] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. *ECCV*, 2022.
- [9] Wei-Ting Chen, Zhi-Kai Huang, Cheng-Che Tsai, Hao-Hsiang Yang, Jian-Jiun Ding, and Sy-Yen Kuo. Learning multiple adverse weather removal via two-stage knowledge learning and multi-contrastive regularization: Toward a unified model. In *CVPR*, 2022.
- [10] Minhyeok Cho, Haechang Lee, Hyunwoo Je, Kijeong Kim, Dongil Ryu, Jinsu Kim, Jonghyun Bae, and Albert No. Pynet-qxq: A distilled pynet for qxq bayer pattern demosaicing in cmos image sensor. *arXiv:2203.04314*, 2022.
- [11] Kai Cui, Zhi Jin, and Eckehard Steinbach. Color image demosaicking using a 3-stage convolutional neural network structure. In *IEEE ICIP*, 2018.
- [12] Valéry Dewil, Adrien Courtois, Mariano Rodríguez, Thibaud Ehret, Nicola Brandonisio, Denis Bujoreanu, Gabriele Facciolo, and Pablo Arias. Video joint denoising and demosaicing with recurrent cnns. In *WACV*, 2023.
- [13] Steven Diamond, Vincent Sitzmann, Frank Julca-Aguilar, Stephen Boyd, Gordon Wetzstein, and Felix Heide. Dirty pixels: Towards end-to-end image processing and perception. *ACM ToG*, 2021.
- [14] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *ICML*, 2017.
- [15] Felix Heide, Markus Steinberger, Yun-Ta Tsai, Mushfiqur Rouf, Dawid Pająk, Dikpal Reddy, Orazio Gallo, Jing Liu, Wolfgang Heidrich, Karen Egiazarian, et al. Flexisp: A flexible camera image processing framework. *ACM ToG*, 2014.
- [16] Xuecai Hu, Haoyuan Mu, Xiangyu Zhang, Zilei Wang, Tieniu Tan, and Jian Sun. Meta-sr: A magnification-arbitrary network for super-resolution. In *CVPR*, 2019.
- [17] Andrey Ignatov, Grigory Malivenko, Radu Timofte, Yu Tseng, Yu-Syuan Xu, Po-Hsiang Yu, Cheng-Ming Chiang, Hsien-Kai Kuo, Min-Hung Chen, Chia-Ming Cheng, et al. Pynet-v2 mobile: Efficient on-device photo processing with neural networks. In *ICPR*, 2022.
- [18] Andrey Ignatov, Luc Van Gool, and Radu Timofte. Replacing mobile camera isp with a single deep learning model. In *CVPRW*, 2020.
- [19] Dongyoung Jang, Donghyuk Park, Seungwon Cha, Heesang Kwon, Mihye Kim, Seungwook Lee, Haewon Lee, Seonok Kim, Nakyung Lee, Jinhwa Han, et al. 0.8  $\mu\text{m}$ -pitch cmos image sensor with dual conversion gain pixel for mobile applications. In *IISW*, 2019.
- [20] Kyeonghoon Jeong, Jonghyun Kim, and Moon Gi Kang. Color demosaicing of rgbw color filter array based on laplacian pyramid. *Sensors*, 2022.
- [21] Xiaozhong Ji, Yun Cao, Ying Tai, Chengjie Wang, Jilin Li, and Feiyue Huang. Real-world super-resolution via kernel estimation and noise injection. In *CVPRW*, 2020.
- [22] Qiyu Jin, Gabriele Facciolo, and Jean-Michel Morel. A review of an old dilemma: Demosaicking first, or denoising first? In *CVPRW*, 2020.
- [23] Jaejin Jung, Sinhwan Lim, Jiyong Kim, Kwisung Yoo, Won-tak Choi, Youngsun Oh, Juhyun Ko, and Kyoungmin Koh. A 1/1.33-inch 108mpixel cmos image sensor with 0.8  $\mu\text{m}$  unit nonacell pixels. In *IEEE ISCAS*, 2022.
- [24] Mehdi Khabir, Hamzeh Alaibakhsh, and Mohammad Azim Karami. Electrical crosstalk analysis in a pinned photodiode cmos image sensor array. *Applied Optics*, 2021.
- [25] Mehdi Khabir and Mohammad Azim Karami. Characterization and analysis of electrical crosstalk in a linear array of cmos image sensors. *Applied Optics*, 2022.
- [26] Byung-Hoon Kim, Joonyoung Song, Jong Chul Ye, and Jae-Hyun Baek. Pynet-ca: enhanced pynet with channel attention for end-to-end mobile image signal processing. In *EC-CVW*, 2020.
- [27] Irina Kim, Dongpan Lim, Youngil Seo, Jeongguk Lee, Yun-seok Choi, and Seongwook Song. On recent results in demosaicing of samsung 108mp cmos sensor using deep learning. In *IEEE TENSYP*, 2021.
- [28] Irina Kim, Seongwook Song, Soonkeun Chang, Sukhwan Lim, and Kai Guo. Deep image demosaicing for submicron image sensors. *JIST*, 2019.
- [29] Filippos Kokkinos and Stamatios Lefkimmiatis. Deep image demosaicking using a cascade of convolutional residual denoising networks. In *ECCV*, 2018.
- [30] Boyun Li, Xiao Liu, Peng Hu, Zhongqin Wu, Jiancheng Lv, and Xi Peng. All-in-one image restoration for unknown corruption. In *CVPR*, 2022.
- [31] Ruoteng Li, Robby T Tan, and Loong-Fah Cheong. All in one bad weather removal using architectural search. In *CVPR*, 2020.

- [32] Yihao Li, Junyu Long, Yun Chen, Yan Huang, and Ni Zhao. Crosstalk-free, high-resolution pressure sensor arrays enabled by high-throughput laser manufacturing. *Advanced Materials*, 2022.
- [33] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *CVPRW*, 2017.
- [34] Lin Liu, Xu Jia, Jianzhuang Liu, and Qi Tian. Joint demosaicing and denoising with self guidance. In *CVPR*, 2020.
- [35] Karima Ma, Michael Gharbi, Andrew Adams, Shoaib Kamil, Tzu-Mao Li, Connelly Barnes, and Jonathan Ragan-Kelley. Searching for fast demosaicking algorithms. *ACM ToG*, 2022.
- [36] Shruti H Mahajan and Varsha K Harpale. Adaptive and non-adaptive image interpolation techniques. In *ICCCA*, 2015.
- [37] Chong Mou, Qian Wang, and Jian Zhang. Deep generalized unfolding networks for image restoration. In *CVPR*, 2022.
- [38] Youngsun Oh, Munhwan Kim, Wonchul Choi, et al. A 0.8  $\mu\text{m}$  nonacell for 108 megapixels cmos image sensor with fd-shared dual conversion gain and 18,000 e-full-well capacitance. In *IEEE IEDM*, 2020.
- [39] Tetuya Okawa, S Ooki, H Yamajo, M Kawada, M Tachi, K Goi, T Yamasaki, H Iwashita, M Nakamizo, T Ogasahara, et al. A 1/2inch 48m all pdaif cmos image sensor using 0.8  $\mu\text{m}$  quad bayer coding  $2 \times 2$ oc1 with 1.0 lux minimum af illuminance level. In *IEEE IEDM*, 2019.
- [40] Dongwon Park, Byung Hyun Lee, and Se Young Chun. All-in-one image restoration for unknown degradations using adaptive discriminative filters for specific degradations. In *CVPR*, 2023.
- [41] Hye Yeon Park, Yunki Lee, Jonghoon Park, et al. Advanced novel optical stack technologies for high snr in cmos image sensor. In *IEEE VLSI Technology and Circuits*, 2022.
- [42] Mara Pistellato, Filippo Bergamasco, Tehreem Fatima, and Andrea Torsello. Deep demosaicing for polarimetric filter array cameras. *IEEE TIP*, 2022.
- [43] Kuldeep Purohit, Maitreya Suin, AN Rajagopalan, and Vishnu Naresh Boddeti. Spatially-adaptive image restoration using distortion-guided networks. In *ICCV*, 2021.
- [44] Jaesung Rim, Geonung Kim, Jungeon Kim, Junyong Lee, Seungyong Lee, and Sunghyun Cho. Realistic blur synthesis for learning image deblurring. *ECCV*, 2022.
- [45] Eli Schwartz, Raja Giryes, and Alex M Bronstein. Deepisp: Toward learning an end-to-end image processing pipeline. *IEEE TIP*, 2018.
- [46] SMA Sharif, Rizwan Ali Naqvi, and Mithun Biswas. SAGAN: Adversarial spatial-asymmetric attention for noisy nona-bayer reconstruction. *arXiv:2110.08619*, 2021.
- [47] Ana Stojkovic, Ivana Shopovska, Hiep Luong, Jan Aelterman, Ljubomir Jovanov, and Wilfried Philips. The effect of the color filter array layout choice on state-of-the-art demosaicing. *Sensors*, 2019.
- [48] Mukund Sundararajan, Ankur Taly, and Qiqi Yan. Gradients of counterfactuals. *arXiv preprint arXiv:1611.02639*, 2016.
- [49] Mukund Sundararajan, Ankur Taly, and Qiqi Yan. Axiomatic attribution for deep networks. In *ICML*, 2017.
- [50] Ethan Tseng, Ali Mosleh, Fahim Mannan, Karl St-Arnaud, Avinash Sharma, Yifan Peng, Alexander Braun, Derek Nowrouzezahrai, Jean-Francois Lalonde, and Felix Heide. Differentiable compound optics and processing pipeline optimization for end-to-end camera design. *ACM TOG*, 2021.
- [51] Ethan Tseng, Yuxuan Zhang, Lars Jebe, Xuaner Zhang, Zhihao Xia, Yifei Fan, Felix Heide, and Jiawen Chen. Neural photo-finishing. *ACM TOG*, 2022.
- [52] Zhengzhong Tu, Hossein Talebi, Han Zhang, Feng Yang, Peyman Milanfar, Alan Bovik, and Yinxiao Li. Maxim: Multi-axis mlp for image processing. In *CVPR*, 2022.
- [53] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *CVPR*, 2022.
- [54] Xun Wu, Zhihao Fan, Jiesi Zheng, Yaqi Wu, and Feng Zhang. Learning to joint remosaic and denoise in quad bayer cfa via universal multi-scale channel attention network. In *ECCVW*, 2023.
- [55] Liangbin Xie, Xintao Wang, Chao Dong, Zhongang Qi, and Ying Shan. Finding discriminative filters for specific degradations in blind super-resolution. *NeurIPS*, 2021.
- [56] Wenzhu Xing and Karen Egiazarian. End-to-end learning for joint image demosaicing, denoising and super-resolution. In *CVPR*, 2021.
- [57] Jun Xu, Yuan Huang, Ming-Ming Cheng, Li Liu, Fan Zhu, Zhou Xu, and Ling Shao. Noisy-as-clean: Learning self-supervised denoising from corrupted image. *IEEE TIP*, 2020.
- [58] Xuan Xu, Yanfang Ye, and Xin Li. Joint demosaicing and super-resolution (jdsr): Network design and perceptual optimization. *IEEE TCI*, 2020.
- [59] Qingyu Yang, Guang Yang, Jun Jiang, Chongyi Li, Ruicheng Feng, Shangchen Zhou, Wenxiu Sun, Qingpeng Zhu, Chen Change Loy, Jinwei Gu, et al. Mipi 2022 challenge on quad-bayer re-mosaic: Dataset and report. In *ECCVW*, 2022.
- [60] Yoonjong Yoo, Jaehyun Im, and Joonki Paik. Low-light image enhancement using adaptive digital pixel binning. *Sensors*, 2015.
- [61] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *CVPR*, 2022.
- [62] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Cycleisp: Real image restoration via improved data synthesis. In *CVPR*, 2020.
- [63] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *CVPR*, 2021.
- [64] Tao Zhang, Ying Fu, and Cheng Li. Deep spatial adaptive network for real image demosaicing. In *AAAI*, 2022.
- [65] Zhimin Zhou, Bedabrata Pain, and Eric R Fossum. Frame-transfer cmos active pixel sensor with pixel binning. *IEEE T-ED*, 1997.