MIPI 2024 Challenge on Demosaic for Hybridevs Camera: Methods and Results

Challenge and Workshop Organizers

Yaqi Wu[1] Zhihao Fan[1] Xiaofeng Chu[1] Jimmy S. Ren[1] Xiaoming Li[2] Zongsheng Yue[2]
Chongyi Li[3] Shangcheng Zhou[2] Ruicheng Feng[2] Yuekun Dai[2] Peiqing Yang[2] Chen Change Loy[2]

Challenge Participants

Senyan Xu[4] Zhijing Sun[4] Jiaying Zhu[4] Yurui Zhu[4] Xueyang Fu[4] Zheng-Jun Zha[4]
Jun Cao[5] Cheng Li[5] Shu Chen[5] Liang Ma[5] Shiyang Zhou[6] Haijin Zeng[7] Kai Feng[8]
Yongyong Chen[6] Jingyong Su[6] Xianyu Guan[9] Hongyuan Yu[9] Cheng Wan[10] Jiamin Lin[9]
Binnan Han[9] Yajun Zou[9] Zhuoyuan Wu[9] Yuan Huang[9] Yongsheng Yu[11] Daoan Zhang[11]
Jizhe Li[9] Xuanwu Yin[9] Kunlong Zuo[9] Yunfan Lu[12] Yijie Xu[12] Wenzong Ma[12]
Weiyu Guo[12] Hui Xiong[12] Wei Yu[13] Bingchun Luo[13] Sabari Nathan[14] Priya Kansal[14]

Abstract

The rising demand for computational photography on mobile devices drives development of advanced image sensors and algorithms for camera systems. But the lack of opportunities for in-depth exchange between industry and academia are constraining the development of Mobile Intelligent Photography and Imaging (MIPI). Building on the successes of the prior MIPI Workshops at ECCV 2022 and CVPR 2023, we are pleased to introduce our third MIPI challenge, which includes three tracks focusing on novel image sensors and imaging algorithms. In this paper, we summarize and review the Demosaic for Hybridevs Camera track on MIPI 2024. A total of 110 participants from both industrial and academic backgrounds contributed many valuable solutions to address the difficulty of the restoration of HybridEVS's raw data, thus raising the reconstructed performance to a new height. This paper gives a comprehensive description and analysis of all solutions developed during this challenge. More detailed information about this challenge is available at https://mipi-challenge.org/MIPI2024.

1. Introduction

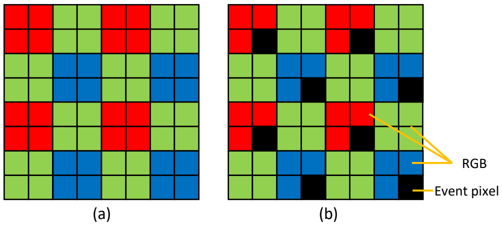


Figure 1. (a) Quad Bayer pattern, (b) HybridEVS pattern.

Quad Bayer Color Filter Array (CFA) (Fig. 1(a)) is a popular CFA pattern widely used in smartphone cameras such as the Galaxy S20 FE and Redmi Note8 Pro. Quad Bayer CFA differs from the traditional Bayer CFA by using 2x2 cells of identical color filters. By utilizing demosaic technics, it can acquire high-resolution image with good image quality. Also, this design ensures exceptional low-light performance through a 2x2 binning operation. But in existing ISP related research, exploration of Quad Bayer CFA is very limited, with most pipelines concentrating on Bayer CFA [6, 8, 23]. Event Vision Sensors (EVS) determine, at pixel level, whether a temporal contrast change beyond a predefined threshold is detected [5, 16]. Compared to CMOS image sensors (CIS), this new modality inherently

[1] SenseTime Research
[2] S-Lab, Nanyang Technological University
[3] Nankai University
[4] University of Science and Technology of China
[5] Xiaomi Inc., China
[6] Harbin Institute of Technology (Shenzhen)
[7] IMEC-UGent
[8] Northwestern Polytechnical University
[9] Multimedia Department, Xiaomi Inc.
[10] Georgia Institute of Technology
[11] University of Rochester
[12] AI Thrust, The Hong Kong University of Science and Technology (Guangzhou)
[13] Harbin Institute of Technology
[14] Couger Inc, Japan

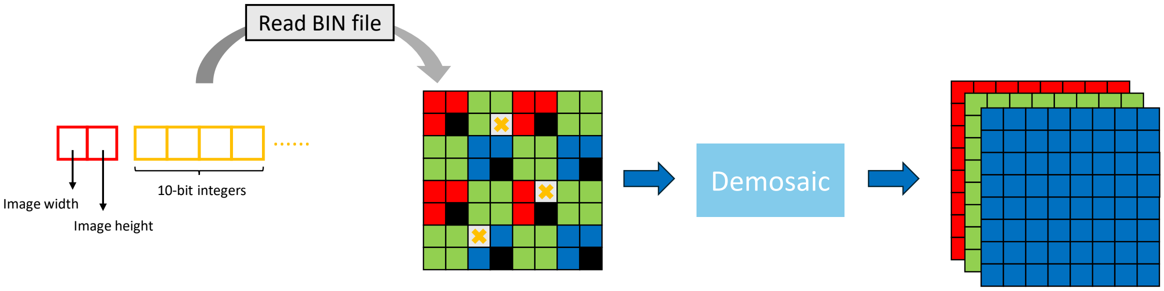


Figure 2. The Demosaic for the Hybridevs Camera aims to reconstruct HybridEVS data into a high-quality RGB result with the same resolution. This process involves passing the data through a demosaic module, which corrects defects and event pixels while reconstructing a three-channel RGB image of matching resolution.

provides data-compression functionality and hence, enables high-speed, low-latency data capture while operating at low power. EVS has tremendous application potential in object tracking, 3D detection, or slow-motion.

Hybrid Event-based Vision Sensor (HybridEVS) [9] is a novel hybrid sensor formed by combining Quad Bayer CFA with Event-based Vision technics. As shown in (Fig. 1(b)), within the 4x4 block, two event pixels are used to capture event signals, while the remaining pixel are utilized to obtain color information. Due to the inability of event pixels to capture color and texture information, demosaicing tasks become more challenging for HybridEVS. Moreover, due to pixel flaws caused by the sensor's manufacturing process, defect pixels may occasionally arise, characterized by significantly divergent pixel values from those of unaffected pixels.

Due to the existing of event pixels and defect pixels, the Demosaic for Hybridevs Camera has become increasingly challenging, with very limited related academic research available. Therefore, we are organizing this competition with the vision of discovering innovative solutions to elevate related research level of this task to a new height.

We hold this challenge in conjunction with the third MIPI Challenge which will be held on CVPR 2024. Similar to the previous MIPI challenge [4, 17, 18, 22, 27], we are seeking an efficient and high-performance image restoration algorithm to handle the Hybridevs camera demosaic task. MIPI 2024 consists of three competition tracks:

- **Few-shot RAW Image Denoising** is geared towards training neural networks for raw image denoising in scenarios where paired data is limited.
- **Demosaic for HybridEVS Camera** is to reconstruct HybridEVS's raw data which contains event pixels and defect pixels into RGB images.
- **Nighttime Flare Removal** is to improve nighttime image quality by removing lens flare effects.

2. MIPI 2024 Demosaic for Hybridevs Camera

To facilitate the development of efficient and high-performance demosaic solutions, we provide a high-quality dataset to be used for training and testing and a set of evaluation metrics that can measure the performance of developed solutions. This challenge aims to advance research on demosaic for HybridEVS camera.

2.1. Problem Definition

As illustrated in Figure 2, the Demosaic for Hybridevs Camera is dedicated to reconstructing HybridEVS input data to a promising RGB result. Due to manufacturing defects, HybridEVS data may contain defect pixels whose actual values deviate significantly from the ideal values. Additionally, the presence of event pixels, essential for capturing motion information, poses challenges to the reconstruction task. Given a HybridEVS input $\mathbf{I}_{\text{in}} \in \mathbb{R}^{H\times W}$, a demosaic method $\mathbf{F}$ reconstructs $\mathbf{I}_{\text{in}}$ into an RGB result $\mathbf{I}_{\text{out}} \in \mathbb{R}^{H\times W\times 3}$. We define the reconstruction task using the following formula:

$$\mathbf{I}_{\text{out}} = \mathbf{F}\left(\mathbf{I}_{\text{in}}\right) \tag{1}$$

2.2. Datasets

As shown in Figure 3, the training dataset consists of 800 pairs of Hybridevs's input data and label result with a resolution of 2K. Both the input and label have the same spatial resolution. The input is of 10 bits in the .bin format, while the output is of 8 bits in the .png format. The validation and testing set has 50 scenes each.

During the testing phase, in order to achieve a more accurate and comprehensive evaluation, the challenge employed a test dataset comprising both simulated and real-world scenarios. For real-world scenarios, we utilize commercially available smartphones equipped with imaging sensors such as Samsung's GN2 sensor to capture images. As illustrated

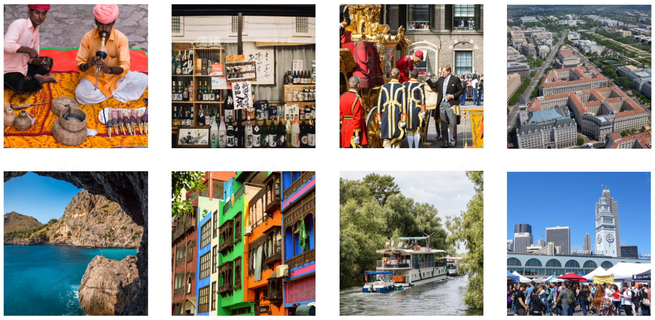


Figure 3. Some sample images from training dataset, the scenes include natural landscapes, architectural views, and other such scenes.

Figure 4. Some sample images from test dataset. Test dataset includes various scenarios like indoor scenes and outdoor scenes

in Figure 4, the scenes we captured encompass a variety of settings, including indoor image quality testing scenes, outdoor architectural scenes, and outdoor strong lighting scenes, etc.

2.3. Evaluation

In this competition, we compare the recovered images with the ground truth images. We utilize the widely adopted Peak Signal-to-Noise Ratio (PSNR) and the complementary Structural Similarity (SSIM) [20] index to evaluate the quality of recovered images. Participants can view these metrics of their submission to optimize the model's performance.

2.4. Challenge Phase

The challenge consisted of the following phases:

1. Development: The registered participants get access to the data and baseline code, and are able to train the models and evaluate their running time locally.
2. Validation: The participants can upload their models to the remote server to check the fidelity scores on the validation dataset, and to compare their results on the validation leaderboard.
3. Testing: The participants submit their final results, code, models, and factsheets.

3. Challenge Results

Among 108 registered participants, 7 teams successfully submitted their results, code, and factsheets in the final test phase. In order to ensure fairness in the competition, we have decided to exclude public datasets such as div2k from the final testing rankings and instead utilize remaining non-public datasets for the final ranking. Table 1 reports the final test results and rankings of the teams.

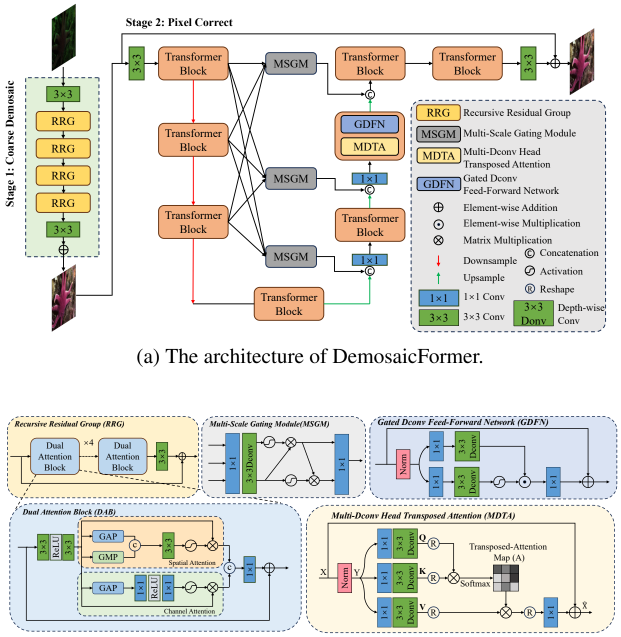
Table 1. Results of MIPI 2024 challenge on the Demosaic for Hybridevs Camera. PSNR and SSIM are computed between the test results and ground truth. To ensure fairness in the competition, publicly available datasets such as DIV2K are excluded. The running time of input of 1080 × 1920 was measured. The measurement was taken on an NVIDIA Geforce GTX 1660Ti.

rank	team	PSNR	SSIM	Time (s)
1	USTC604	**44.8464**$_{(1)}$	**0.9854**$_{(1)}$	51.315$_{(6)}$
2	lolers	44.6234$_{(2)}$	0.9847$_{(2)}$	18.231$_{(2)}$
3	Lumos_Demosaicker	44.4951$_{(3)}$	0.9845$_{(3)}$	26.284$_{(4)}$
4	High_speed_Machines	43.9564$_{(4)}$	0.9838$_{(4)}$	101.768$_{(7)}$
5	Yunfan	42.6508$_{(5)}$	0.9810$_{(5)}$	37.508$_{(5)}$
6	HIT-CVLAB	41.3280$_{(6)}$	0.9780$_{(6)}$	25.421$_{(3)}$
7	CougerAI	41.0736$_{(7)}$	0.9752$_{(7)}$	**6.331**$_{(1)}$

Finally, the USTC604 team is the first place winner of this challenge, while lolers team win the second place and Lumos Demosaicker team is the third place, respectively. The overall performance of all participating teams' solutions consistently exceeds 40dB in test dataset, indicating that all participating teams can achieve relatively good reconstruction results. The PSNR of the first-place model reached 44.8464 dB, leading the second-place by 0.223 dB. The third-place model has a significant advantage over the fourth-place, with a margin of 0.5386 dB, which indicates that the top three contestants exhibit considerable superiority.

4. Methods

USTC604 This team proposes a coarse-to-fine framework named DemosaicFormer which comprises a coarse demosaicing network and a pixel correction network (see Figure 5). For the coarse demosaicing stage, to produce a preliminary high-quality estimate of the RGB image from the HybridEVS raw data, this team introduce Recursive Residual Group (RRG) [24] which employs multiple Dual Attention Blocks (DABs) to refine the feature representation progressively. For pixel correction stage, aiming to enhance the performance of image restoration and mitigate the impact of defective pixels, this team introduces the Transformer Block[25] which consists of Multi-Dconv Head Transposed Attention (MDTA) and Gated-Dconv Feed-Forward Network (GDFN). The key innovation is the design of a novel Multi-Scale Gating Module (MSGM) applying the integration of cross-scale features inspired by [1], which allows feature information to flow between different scales. Due to



(a) The architecture of DemosaicFormer.

(b) The structures of sub-modules in the main architecture.

Figure 5. The architecture of DemosaicFormer proposed by team USTC604 to demosaic the raw data captured by HybridEVS cameras.

the inability to accurately model defective pixels, inspired by [24], this team extract the defect pixels map from the training data of the challenge to generate more diverse and realistic inputs for data augmentation. At the training phase, this team randomly rotate and flip ground-truth images of training split, then sample them according to HybridEVS pattern, and randomly cover the sampled images with defect pixels map. The augmentation technology is applied at the initial training of the proposed approach for improving the model's generalization and robustness. The training phase of the proposed method could be divided into two stages:

(1) **Initial training of DemosaicFormer**. This team use a progressive training strategy at first. Start training with patch size 80×80 and batch size 84 for 58K iterations. The patch size and batch size pairs are updated to $[(128^2, 30), (160^2, 18), (192^2, 12)]$ at iterations [36K, 24K, 24K]. The initial learning rate is 5×10^{-4} and remains unchanged when patch size is 80. Later the learning rate changes with Cosine Annealing scheme to 1×10^{-7}. The best model at this stage is used as the initialization of the second stage.

(2) **Fine-tuning DemosaicFormer**.This team start training with patch size 192×192 and batch size 12. The initial learning rate is 1×10^{-4} and changes with Cosine Annealing scheme to 1×10^{-7}, including 20K iterations in total. Note that use the entire training data from the challenge without any data augmentation technologies at this stage. Exponential Moving Average (EMA) is applied for the dynamic adjustment of model parameters.

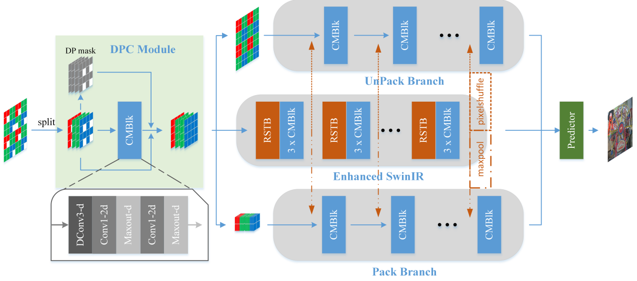


Figure 6. The multi-branch network architecture proposed by lolers team

lolers This team proposed a SwinIR-based multi-branch network for hybridevs demosaicing (see Figure 6). The backbone uses an enhanced SwinIR [11] to extract rich features with long-range dependencies and pass them to the UnPack branch and Pack branch, which are used to indicate the spatial position of full resolution and provide uniform sampling, respectively. In addition, the authors propose a lightweight Cascaded Maxout Block (CMBlk), which consists of a depthwise convolution and multiple consecutive Maxouts, to give the model powerful representation with a small number of parameters. The UnPack and Pack branches are composed of the same number of CMBlk.

For a defected QCFA raw, the authors first separate them into 4 single-channel inputs I, and use a defect pixel mask (DP mask) to perform defect pixel correction (DPC).

$$DP\_mask = (I == 0)$$

$$I = CMBlk(I) * DP\_mask + I$$

After passing the DPC module, the 4-channel input I is packed and unpacked respectively, restored to QCFA raw and 16-channel sub-images, and then input to the backbone and two branches. For the backbone network, they use a SwinIR with a depth of 6, and connect 3 CMBlk after each RSTB to improve its feature extraction capability. Accordingly, the features of each layer are passed separately into two lightweight branches for feature enhancement. Finally, the three branches are fused to inference the final demosaicing result.

In training stage, the authors use the Adam optimizer, and the initial learning rate is set to 2e-4. First, they trained the DPC module and backbone by L1 loss. When the network is fitted, they fix backbone's parameters and add two branches to continue training. After 50w iterations, the parameters of all three branches are updated in an end-to-end manner.

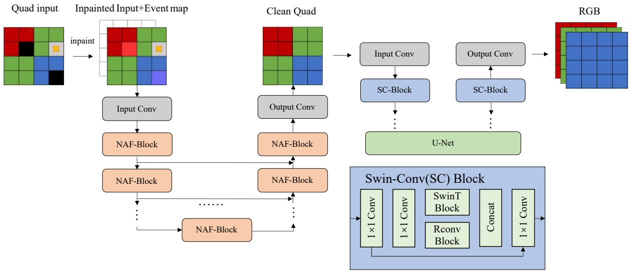


Figure 7. The network architecture of the proposed TSIDN

Lumos Demosaicker This team introduces a novel two-stage network, termed the Two-Stage joint Inpainting and Demosaicing Network (TSIDN), as depicted in Figure 7. Initially, the network addresses the influence of event points by employing an inpainting process, which replaces them with the average values of neighboring pixels. Subsequently, the primary task is segmented into two stages, facilitating independent and joint training for each sub-network. The first stage features a Quad-to-Quad (Q2Q) network, which takes inpainted Quad Bayer data and event pixel maps as input. It utilizes a NAFNet [1] to effectively restore both event and defect pixels, integrating positional information to enhance restoration accuracy. Building upon this foundation, the second stage employs a Quad-to-RGB (Q2R) network based on SCUNet [26] to focus on demosaicing. This network benefits from Swin-Conv and U-Net structures, ensuring efficient demosaicing performance. During the training phases, strategies such as phase-based training and progressive learning are incorporated to enhance network performance. In the joint training stage, a progressive learning strategy is employed, starting with a patch size of 256 pixels, which progressively increases to 382 and then to 500 pixels. l_1 loss is utilized during the pretraining stage, while PSNR loss is applied during joint training. The initial learning rate is set at 1×10^{-4} and is gradually decreased to 1×10^{-7}, contributing to the robustness and efficiency of the proposed network.

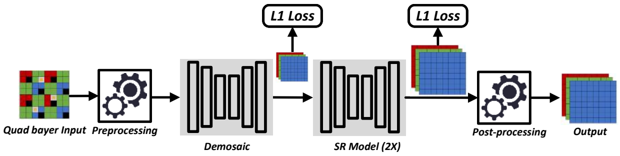


Figure 8. Step-by-step Demosaic model for Hybrid Evs Camera(SBSDM).

High-speed Machines The Demosaic for Hybrid Evs Camera requires a 4x scale expansion, which is quite challenging for the model (see Figure 8). Therefore, a large number of parameters are needed for the model to learn, which in turn requires a substantial amount of data for training. However, the amount of training data provided by the competition is quite low. To address this difficulty, this team proposes the Step-by-step Demosaic model for Hybrid Evs Camera (SBSDM).

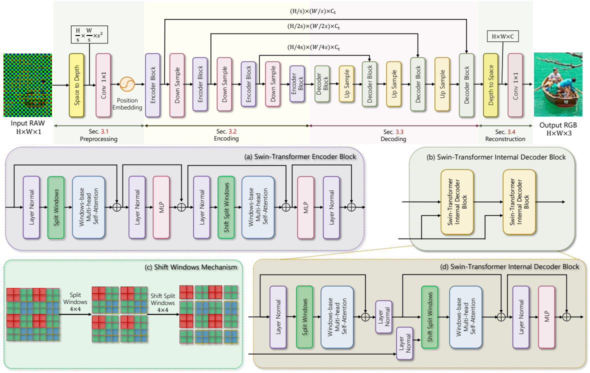


Figure 9. Overview of the proposed demosaicing method for event cameras: The process begins with preprocessing the RAW image, followed by feature extraction via an encoder using Swin Transformer blocks and a Shifted Window mechanism. The decoder, mirroring the encoder and including skip connections, reconstructs spatial details. The final stage is image reconstruction to produce the RGB output. Illustrated components include (a) the encoder architecture, (b) the Shifted Window mechanism for enhanced interaction, and (c) the decoder architecture.

This team's model, SBSDM, is a two-stage model. The first stage involves transforming the Quad Bayer input into a 2x RGB output image. In the data preprocessing part, since the PD point locations do not contain valid pixel information, this team has removed this information from the input and decomposed the original image into 14 channels. In the model, this team adopted the SPAN [19] model and expanded the channels to 96. The second stage involves a 2x super-resolution. Since there are many public models for super-resolution tasks, this team can conveniently use models pre-trained on large datasets as the second-stage model. Here, this team used EDSR [12], HAT [2], and SwinIR [11] to construct different models. Finally, this team trained the entire model end-to-end using the L1 loss function.

Yunfan This team has devised a comprehensive method for demosaicing images from event cameras by leveraging a sophisticated framework that combines the Swin Transformer [13] and U-Net [15] architectures, as shown in Fig. 9. They have methodically outlined a multi-faceted approach consisting of preprocessing, encoding, decoding, reconstruction, and a novel loss function, each contributing uniquely to the image reconstruction process.

In the preprocessing phase, the team effectively transforms the input RAW image and reduces computational

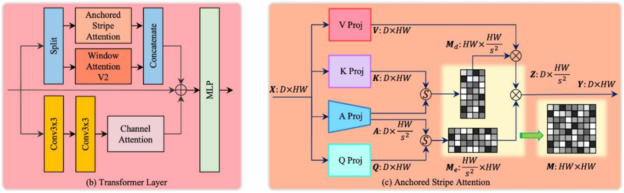


Figure 10. (b) The architecture of the adopted GRL layer. (c) The key anchored stripe attention mechanism.

complexity using space-to-depth operations [7] and 1 × 1 convolutions. The encoding stage, utilizing Swin Transformer blocks, extracts multi-scale features and captures long-range dependencies, while the decoding phase mirrors the encoding structure, progressively recovering the image's spatial resolution. The team's reconstruction module is adept at generating the final RGB image from upsampled features by reversing the preprocessing operations and refining the high-dimensional features into a standard color space.

Notably, the team's strategic innovation lies in their two-stage training methodology that employs a Charbonnier loss for initial training and a Pixel Focus Loss for fine-tuning. They have meticulously engineered the Pixel Focus Loss to address long-tail distribution issues in training loss, focusing on edge detail enhancement. This loss function is instrumental in the model's ability to distinguish between high-frequency edge information and low-frequency color block differences, enhancing fine details' restoration.

Through this systematic approach, the team has strengthened the network's capability to learn global colour distributions and local edge details, culminating in high-quality RGB image reconstruction. Their experiments demonstrate the efficacy of their method.

HIT-CVLAB This team designs a UNet structure network based on the GRL [10] layer (see Figure 10), which explicitly models image hierarchies at global, regional, and local scales through anchored strip self-attention, window self-attention, and channel attention-augmented convolutions. Training stratgies: For training, the authors adopt mini-batch stochastic gradient descent (SGD) with a batch size of 64 for 600 epochs. The initial value of the learning rate is 3e-4 and gradually decreases to 1e-6 with a CosineAnnealing schedule. This learning rate decreasing strategy helps the model adjust parameters more carefully when it is close to convergence, thereby improving the generalization performance of the model and being able to better cope with Noise and uncertainty in training data. Training loss functions include L1 loss, L2 loss, and Sobel loss.

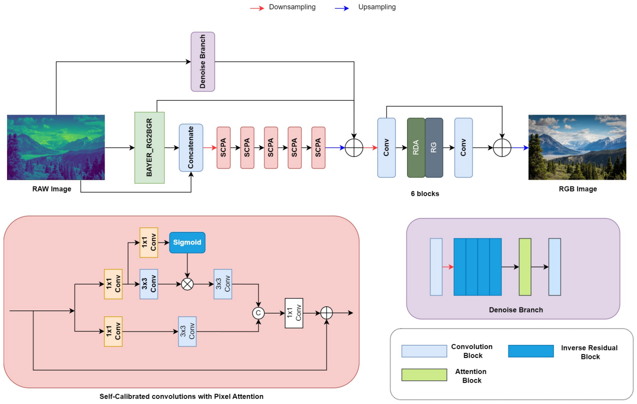


Figure 11. Multi-Stage Fusion Demosaicing architecture proposed by CougerAI team

CougerAI Our demosaicing model begins with the initial step of converting the input image from its raw Bayer pattern (RGB) to the standard RGB color space using a BayerRG2BGR method (see Figure 11). The resulting RGB image is then fused with the raw input image and jointly processed through a novel architecture inspired by recent advancements in computer vision.

This joint processing involves passing the concatenated images through a Self-Calibrated convolution with a pixel attention block (SCPA)[14, 28], which enhances the spatial and spectral coherence of the input. Simultaneously, the raw input image is fed into a sophisticated denoising block[3] to produce a three-channel denoised image. The denoising block is composed of four inverse convolutional layers followed by an attention mechanism that effectively suppresses noise while preserving important image details.

After denoising, the denoised image and the output of the SCPA block are combined with the converted RGB image and fed through a downsampling layer to reduce computational complexity and enhance feature extraction. The downsampled features are then input to a residual learning block, which consists of a series of residual group blocks[21] followed by a residual channel dense attention block. This architecture allows the model to effectively capture both local and global contextual information, improving its ability to reconstruct the image.

Finally, the output of the residual learning block is added back to the input image and upsampled to produce the final demosaiced image.

5. Conclusions

In this report, we review and summarize the methods and results of MIPI 2024 challenge on the Demosaic for Hybridevs Camera. The participants have made significant contributions to this challenging track, and we express our gratitude for the dedication of each participant.

6. Teams and Affiliations

USTC604

Title: DemosaicFormer: Coarse-to-Fine Demosaicing Network for HybridEVS Camera
Members:
Senyan Xu[1] (syxu@mail.ustc.edu.cn)
Zhijing Sun[1] Jiaying Zhu[1] Yurui Zhu[1] Xueyang Fu [1] Zheng-Jun Zha[1]
Affiliations:
[1] University of Science and Technology of China

lolers

Title: Multi-Resolution SwinMaxIR for QCFA Raw Demosaic
Members:
Jun Cao[1] (caojun6@xiaomi.com)
Cheng Li[1] Shu Chen[1] Liang Ma [1]
Affiliations:
[1] Xiaomi Inc., China

Lumos Demosaicker

Title: Two-Stage joint Inpainting and Demosaicing Network
Members:
Shiyang Zhou[1] (shiyangzhou2023@163.com)
Haijin Zeng[2] Kai Feng[3] Yongyong Chen[1] Jingyong Su[1]
Affiliations:
[1] Harbin Institute of Technology (Shenzhen)
[2] IMEC-UGent
[3] Northwestern Polytechnical University

High-speed Machines

Title: Step-by-step Demosaic model for Hybrid Evs Camera
Members:
Xianyu Guan[1] (guanxianyu@xiaomi.com)
Hongyuan Yu[1] Cheng Wan[3] Jiamin Lin[1] Binnan Han[1] Yajun Zou[1] Zhuoyuan Wu[1] Yuan Huang[1] Yongsheng Yu[2] Daoan Zhang[2] Jizhe Li[1] Xuanwu Yin[1] Kunlong Zuo[1]
Affiliations:
[1] Multimedia Department, Xiaomi Inc.
[2] University of Rochester
[3] Georgia Institute of Technology

Yunfan

Title: Event Camera Demosaicing via Swin Transformer and Pixel-focus Loss
Members:
Yunfan LU[1] (ylu066@connect.hkust-gz.edu.cn)
Yijie XU[1] Wenzong MA[1] Weiyu GUO[1] Hui XIONG[1]
Affiliations:
[1] AI Thrust, The Hong Kong University of Science and Technology (Guangzhou)

HIT-CVLAB

Title: Efficient and Explicit Hierarchies Modelling Network for HybridEVS Camera Demosaic
Members:
Wei Yu[1] (20b903014@stu.hit.edu.cn)
Bingchun Luo[1]
Affiliations:
[1] Harbin Institute of Technology

CougerAI

Title: Multi-Stage Fusion Demosaicing with Integrated Pixel Attention and Residual Learning
Members:
Sabari Nathan[1] (sabari@couger.co.jp)
Priya Kansal[1]
Affiliations:
[1] Couger Inc, Japan

References

[1] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *European Conference on Computer Vision*, pages 17–33. Springer, 2022. 3, 5

[2] Xiangyu Chen, Xintao Wang, Jiantao Zhou, Yu Qiao, and Chao Dong. Activating more pixels in image super-resolution transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 22367–22377, 2023. 5

[3] Vasluianu F. Nathan S. Timofte R. Conde, M.V. Real-time under-display cameras image restoration and hdr on mobile devices. In *Computer Vision – ECCV 2022 Workshops*. Springer, 2022. 6

[4] Yuekun Dai, Chongyi Li, Shangchen Zhou, Ruicheng Feng, Qingpeng Zhu, Qianhui Sun, Wenxiu Sun, Chen Change Loy, Jinwei Gu, Shuai Liu, et al. Mipi 2023 challenge on nighttime flare removal: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2852–2862, 2023. 2

[5] Guillermo Gallego, Tobi Delbrück, Garrick Orchard, Chiara Bartolozzi, Brian Taba, Andrea Censi, Stefan Leutenegger, Andrew J Davison, Jörg Conradt, Kostas Daniilidis, et al. Event-based vision: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 44(1):154–180, 2020. 1

[6] Biay-Cheng Hseih, Hasib Siddiqui, Jiafu Luo, Todor Georgiev, Kalin Atanassov, Sergio Goma, HY Cheng, JJ Sze, RJ Lin, KY Chou, et al. New color filter patterns and demosaic for sub-micron pixel arrays. In *Proceedings of the International Image Sensor Workshop, Vaals, The Netherlands*, pages 8–11, 2015. 1

[7] Pei-Hsiang Hsu, Pei-Jun Lee, Trong-An Bui, and Yi-Shau Chou. Yolo-spd: Tiny objects localization on remote sensing

based on you only look once and space-to-depth convolution. In *2024 IEEE International Conference on Consumer Electronics (ICCE)*, pages 1–3. IEEE, 2024. 6

[8] Yongnam Kim and Yunkyung Kim. High-sensitivity pixels with a quad-wrgb color filter and spatial deep-trench isolation. *Sensors*, 19(21):4653, 2019. 1

[9] Kazutoshi Kodama, Yusuke Sato, Yuhi Yorikado, Raphael Berner, Kyoji Mizoguchi, Takahiro Miyazaki, Masahiro Tsukamoto, Yoshihisa Matoba, Hirotaka Shinozaki, Atsumi Niwa, et al. 1.22 μm 35.6 mpixel rgb hybrid event-based vision sensor with 4.88 μm-pitch event pixels and up to 10k event frame rate by adaptive control on event sparsity. In *2023 IEEE International Solid-State Circuits Conference (ISSCC)*, pages 92–94. IEEE, 2023. 2

[10] Yawei Li, Yuchen Fan, Xiaoyu Xiang, Denis Demandolx, Rakesh Ranjan, Radu Timofte, and Luc Van Gool. Efficient and explicit modelling of image hierarchies for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18278–18289, 2023. 6

[11] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. *arXiv preprint arXiv:2108.10257*, 2021. 4, 5

[12] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2017. 5

[13] Ze Liu, Jia Ning, Yue Cao, Yixuan Wei, Zheng Zhang, Stephen Lin, and Han Hu. Video swin transformer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3202–3211, 2022. 5

[14] P. Nathan, S.; Kansal. End-to-end depth-guided relighting using lightweight deep learning-based method. pages 9, 175. 6

[15] Nahian Siddique, Sidike Paheding, Colin P Elkin, and Vijay Devabhaktuni. U-net and its variants for medical image segmentation: A review of theory and applications. *Ieee Access*, 9:82031–82057, 2021. 5

[16] B Son, Y Suh, S Kim, H Jung, JS Kim, C Shin, K Park, K Lee, J Park, J Woo, et al. A 640× 480 dynamic vision sensor with a 9 um pixel and 300 meps address-event representation. In *Proceedings of the IEEE International Conference on Solid-State Circuits, San Francisco, CA, USA*, pages 5–9, 2017. 1

[17] Qianhui Sun, Qingyu Yang, Chongyi Li, Shangchen Zhou, Ruicheng Feng, Yuekun Dai, Wenxiu Sun, Qingpeng Zhu, Chen Change Loy, Jinwei Gu, et al. Mipi 2023 challenge on rgbw remosaic: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2877–2884, 2023. 2

[18] Qianhui Sun, Qingyu Yang, Chongyi Li, Shangchen Zhou, Ruicheng Feng, Yuekun Dai, Wenxiu Sun, Qingpeng Zhu, Chen Change Loy, Jinwei Gu, et al. Mipi 2023 challenge on rgbw fusion: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2870–2876, 2023. 2

[19] Cheng Wan, Hongyuan Yu, Zhiqi Li, Yihang Chen, Yajun Zou, Yuqing Liu, Xuanwu Yin, and Kunlong Zuo. Swift parameter-free attention network for efficient super-resolution. *arXiv preprint arXiv:2311.12770*, 2023. 5

[20] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 3

[21] Egiazarian K. Xing, W. End-to-end learning for joint image demosaicing, denoising and super-resolution. In *Proceedings of the IEEE/CVF international conference on computer vision*, 2021. 6

[22] 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 *European Conference on Computer Vision*, pages 21–35. Springer, 2022. 2

[23] K Yonemoto. Principles and applications of ccd/cmos image sensors, 2003. 1

[24] 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 *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2696–2705, 2020. 3, 4

[25] 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 *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5728–5739, 2022. 3

[26] Kai Zhang, Yawei Li, Jingyun Liang, Jiezhang Cao, Yulun Zhang, Hao Tang, Deng-Ping Fan, Radu Timofte, and Luc Van Gool. Practical blind image denoising via swin-conv-unet and data synthesis. *Machine Intelligence Research*, 20(6):822–836, 2023. 5

[27] Qingpeng Zhu, Wenxiu Sun, Yuekun Dai, Chongyi Li, Shangchen Zhou, Ruicheng Feng, Qianhui Sun, Chen Change Loy, Jinwei Gu, Yi Yu, et al. Mipi 2023 challenge on rgb+ tof depth completion: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2863–2869, 2023. 2

[28] Ye T. Zheng W. Zhang Y. Chen L. Wu Y. Zou, W. Self-calibrated efficient transformer for lightweight super-resolution. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3507–3516, 2021. 6