

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2024)06-1646-21

论文引用格式: Yue H J, Yang W H, Li C Y, Yang Y, Liu W Y and Yang J Y. 2024. Advances of low-level vision reconstruction in raw domain. Journal of Image and Graphics, 29(06):1646-1666(岳焕景, 杨文瀚, 李重仪, 杨铀, 刘文予, 杨敬钰. 2024. 像感域(Raw域)底层视觉重建技术进展. 中国图象图形学报, 29(06):1646-1666)[DOI:10.11834/jig.230794]

## 像感域(Raw域)底层视觉重建技术进展

岳焕景<sup>1</sup>, 杨文瀚<sup>2</sup>, 李重仪<sup>3</sup>, 杨铀<sup>4</sup>, 刘文予<sup>4</sup>, 杨敬钰<sup>1\*</sup>

1. 天津大学电气自动化与信息工程学院, 天津 300072; 2. 鹏城实验室战略与交叉前沿研究部, 深圳 518055;
3. 南开大学计算机学院, 天津 300350; 4. 华中科技大学电子信息与通信学院, 武汉 430074

**摘要:** 底层视觉重建技术旨在受限于成像条件下重建高质量图像/视频, 对后续视觉处理与呈现具有重要意义。由于像感域数据(raw data)具有高位宽、与感光量成线性响应等特点, 近年来基于像感域的视觉重建技术在学术界和工业界获得的关注日益提高。本文聚焦于6种代表性视觉重建任务, 包括低光增强与去噪、超分辨率、高动态范围重建、去摩尔纹、多任务联合重建以及数据生成, 重点综述了深度学习驱动的像感域视觉重建领域的进展; 系统地总结了领域代表性方法, 概述各类方法的优势与局限, 分析了不同任务中像感域数据相较于颜色域数据(经降噪、去马赛克、白平衡、色调映射和颜色空间转换(如RGB、sRGB等)等处理之后的数据)的独特属性与优势; 梳理了各个领域的开源数据集, 包括图像数据集、快速连拍数据集以及视频数据集, 总结了数据集的构造方法以及配对数据的空间/时间对齐策略, 为后续研究的数据集创建提供了参考与指引; 总结了现有方法存在的问题与困境, 展望了像感域底层视觉重建的发展趋势。

**关键词:** 像感域(Raw域)图像重建; Raw域图像(视频)低光增强; Raw域图像(视频)去噪; Raw域图像(视频)超分辨率; Raw域图像(视频)高动态范围重建; Raw域图像(视频)去摩尔纹

### Advances of low-level vision reconstruction in raw domain

Yue Huanjing<sup>1</sup>, Yang Wenhan<sup>2</sup>, Li Chongyi<sup>3</sup>, Yang You<sup>4</sup>, Liu Wenyu<sup>4</sup>, Yang Jingyu<sup>1\*</sup>

1. School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China;
2. Department of Strategic and Advanced Interdisciplinary, PengCheng Laboratory, Shenzhen 518055, China;
3. College of Computer Science, Nankai University, Tianjin 300350, China; 4. School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan 430074, China

**Abstract:** The low-level vision reconstruction technology aims to reconstruct high-quality images and videos under limited imaging conditions, which is important for subsequent visual analysis. The images (videos) in the raw domain have two advantageous features: wider bit depth (10, 12, 14 bits) and intensity linear to the irradiance. As a result, raw images contain the most original information and the noise statistics are also simpler than those in standard RGB (sRGB) domain. Therefore, low-level vision reconstruction with raw inputs has achieved an increasing attention from academic and industrial communities in recent years. This review focuses on the low-level vision reconstruction technology in the raw domain and mainly investigates the progress of deep learning-based vision reconstruction. Six representative vision reconstruction

收稿日期: 2024-01-19; 修回日期: 2024-02-29; 预印本日期: 2024-03-07

\* 通信作者: 杨敬钰 yjy@tju.edu.cn

基金项目: 国家自然科学基金项目(62072331, 62231018, 62376102)

Supported by: National Natural Science Foundation of China(62072331, 62231018, 62376102)

tasks in raw domain are selected, namely, low-light enhancement and denoising, super-resolution, high dynamic range (HDR) reconstruction, moiré removal, multi-task joint reconstruction, and raw image generation, for a comprehensive review. Representative methods in the six fields are systematically summarized, the advantages and problems of various methods are outlined, and the advantages and unique attributes of raw images (videos) compared with sRGB images (videos) are highlighted in different tasks. Thereafter, the currently open-source low-level vision reconstruction datasets in raw domain in various fields are summarized, including image, burst image, and video datasets. The dataset construction methods for each task are introduced. Different strategies to solve the key problems in dataset construction, namely, spatial alignment and temporal alignment, are also introduced. We hope these summarization and comparisons can provide references for the followers who construct their own datasets. This review would like to point out that the six tasks not only have unique problems but also have common issues. For example, for denoising and enhancement of videos captured in low light, constructing a supervised dataset with realistic motions and fine-scale textures is still difficult. For multi-frame super-resolution, the key problem is constructing the accurate alignment module. For HDR reconstruction, the deghosting performance still needs to be improved in dark and over-exposed areas. For demoiréing, balancing the performance between color recovery and moiré removal needs to be explored. For multi-task joint reconstruction, improving the adjustability and interpretability of the model is a key problem. Meanwhile, all the six tasks need to recover the correct colors while completing their own tasks. However, they have different optimization directions. Introducing special modules to ensure their similar optimization directions may be a good solution. In addition, achieving accurate alignment between degraded and ground truth images is difficult, and many datasets exhibit misalignment. Then, we review representative industrial applications of raw domain reconstruction, including intelligent image signal processing and night imaging in smartphones, low-light and HDR imaging in security monitoring cameras, and raw domain detection in driverless cars. Finally, based on the existing problems and challenges of raw domain vision reconstruction, we identify four possible development trends for raw domain vision reconstruction. First is exploiting the specific properties of raw images (videos) for a specific task. Current methods usually utilize the advantages of raw data in terms of wider bit depth and linearity to intensity. Only a few works utilize the specialized structures of raw data. For example, the moiré distribution in different channels differs, and the green channel usually has higher intensities than other channels. We expect more works exploring the special properties of raw data in popular denoising and super-resolution tasks. Second is improving the availability of large-scale raw data. Many cameras do not provide the raw outputs due to the large memory cost. Therefore, the current constructed raw domain datasets are usually smaller than those in sRGB domain. A feasible solution is to design the raw image compression method with sRGB image guidance for enabling raw domain decoding with a few meta data. Third is alleviating the data-bias problem. The model trained with the raw data captured with one camera may not work well when dealing with raw images captured with other cameras. Alleviating the data-bias is important for real applications. One feasible solution is to jointly utilize physics- and data-driven models. Fourth is further improving raw reconstruction performance with large models. The scale of data is important to improve the reconstruction quality. One solution is to first train a large model with a large-scale dataset and then distill the large model to a small one. Then, the small model can be deployed in various edge devices. In summary, we expect more works exploring low-level vision reconstruction in raw domain to improve the imaging quality of various vision systems.

**Key words:** raw-domain vision reconstruction; low light image (video) enhancement in raw domain; raw image (video) denoising; raw image (video) super-resolution; raw image (video) high dynamic range (HDR) reconstruction; raw image (video) demoiréing

## 0 引言

在电子成像系统中,图像传感器将外界光信号

转化为电信号,再经模数转换后输出数字图像信号,本文将其称为像感域(Raw域)图像。对于多数成像设备而言,像感域数据经过图像信号处理器(image signal processor, ISP)的一系列处理,转换为通用颜

色域的图像数据以方便压缩、存储、传输和显示等后续处理环节<sup>①</sup>。

图1展示了一种典型的相机ISP处理过程。现有视觉重建方法旨在通过图像处理方法进一步提升图像/视频质量(例如图像降噪与增强(都双丽等,

2023;马龙等,2022)、超分辨率(魏凌云和孙帮勇,2022;江俊君等,2023)、高动态范围重建(唐凌峰等,2022)),大多数采用通用颜色空间的图像/视频作为输入数据。近年来,基于像感域的视觉重建被越来越多的研究者所关注。

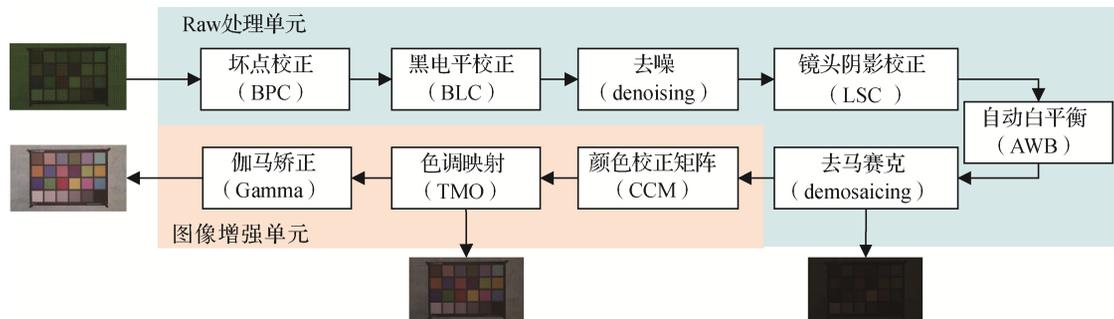


图1 相机ISP过程

Fig. 1 The ISP process in camera imaging

相较于颜色域视觉重建,基于像感域的视觉重建有诸多优势。首先,像感域数据具有高位宽(10、12、14 bits为常见格式),拥有更丰富的原始信息,而颜色域数据为量化后的8 bits低位宽数据,信息丢失较多。其次,像感域数据与感光量成线性响应,成像过程中的线性失真经ISP处理后使得图像退化过程更为复杂。例如去马赛克(demaicing)过程使得空间独立的噪声具有空间相关性,同时也使颜色域摩尔纹相较于像感域摩尔纹更为复杂。最后,像感域信号与物理信号存在直接的对应关系,基于像感域进行视觉重建,结果更加直接明了,便于内嵌至ISP过程中,进而提升多类终端的成像能力。鉴于此,基于像感域的视觉重建技术日益涌现,在学术界和工业界均受到广泛关注。例如,在学术界,2018年发表在旗舰会议与期刊的像感域重建论文仅有10篇,2023年发表的相关论文已攀升至30篇。在工业界,众多手机厂商发布了面向像感域处理的智能ISP芯片,例如OPPO的马里亚纳、小米的澎湃C1等。

然而,像感域视觉重建也面临诸多难点与挑战。作为视觉重建领域的一个子任务,现有的像感域视觉重建方法大多以颜色域重建性能优良的视觉重建模型为基础,加以考虑Raw的特性进行设计。例如在视频超分辨率任务中表现出色的EDVR(video

restoration with enhanced deformable convolutional networks)(Wang等,2019)基础上,引入拜尔模式分离、交织拜尔分支与分离拜尔分支的联合交互,ISP模块联调等措施,以适配像感域去噪与超分辨率任务(Yue等,2020,2022);在Transformer的基础上引入适配于特定CFA(color filter array)的token化方式(Fan等,2023),以更好地利用像感域数据分布特性。该类方法提升了像感域视觉重建性能,但仅针对拜尔模式进行不同的适配,缺少特定任务下针对Raw数据特性的更细粒度的分析与利用。另外,相较于现有颜色域的视觉数据规模,像感域视觉数据规模依然受限。例如,颜色域自然图像超分辨率数据集LSDIR(a large scale dataset for image restoration)(Li等,2023)的数据规模已达8万余幅,而像感域最大的超分辨率数据集ImagePairs(Joze等,2020)仅包含8591幅图像。深度学习模型的性能依赖于高质量的标签数据,如何方便地获取高质量的像感域数据也是亟待突破的方向之一。

为了更好地概述领域进展,为后续研究提供总结性材料,促进领域发展,本文选取6种代表性的像感域视觉重建任务,即去噪与增强、超分辨率、高动态成像(high dynamic range, HDR)、去摩尔纹、多任务联合重建和数据生成,以算法和数据集为研究对

<sup>①</sup>图像传感器直接输出的数据在英文文献中称为Raw data,本文将Raw domain意译为像感域,将Raw data译为像感域数据。此外,与像感域数据对应,本文将经降噪、去马赛克、白平衡、色调映射、颜色空间转换(如RGB、sRGB等)等处理之后的数据称为颜色域数据。

象,全面回顾Raw域视觉重建技术发展历程,并对应用现状和发展趋势进行总结。

## 1 Raw域去噪与增强

如图2所示,成像过程中会受到散粒噪声、读出噪声和暗噪声的影响,在低光环境下尤甚。该类噪声经过非线性ISP变换至颜色域后,变成与空间、通道相关的复杂噪声,因此在像感域进行(ISP前)去噪更具优势。考虑到暗光图像/视频增强的主要任务也是噪声去除与细节恢复,因此本节介绍像感域去噪与增强方向的相关进展。首先介绍相关方法,之后介绍数据集发展状况,包括噪声仿真与真实数据采集。

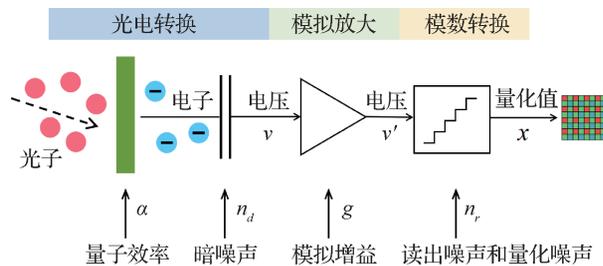


图2 成像过程中的噪声来源

Fig. 2 Noise sources during imaging

### 1.1 Raw域去噪与增强方法

#### 1.1.1 Raw域图像去噪与增强

不同于颜色域图像去噪,像感域去噪需要考虑图像的拜尔模式,且数据集较少。同时,由于图像最终在颜色域观看,去噪过程需联合考虑ISP的影响。基于上述考虑,近年来涌现出诸多出色的像感域去噪与增强工作。Chen等人(2018)创建了首个像感域暗光增强去噪数据集SID(see-in-the-dark),并提出了基于U型网络的端到端映射,直接将像感域低光带噪图像映射为颜色域无噪增强图像,抑制了暗光噪声并恢复了正确颜色。针对像感域数据少且不同相机拜尔模式不同的问题,Liu等人(2019)提出了拜耳模式统一方法(BayerUnify),从而利用异构数据集来训练单个去噪模型,提升了模型对不同数据的适配能力。另外,提出了拜耳保留增广(BayerAug)方法,以进一步提升模型效果。Zhang等人(2022)进一步提出了基于自监督的像感域图像去噪方法,利用噪声较大的数据集上训练的模型来优化噪声图

像以获得下轮迭代需要的标签。由于其只需单个噪声图像和噪声模型即可降噪,对于数据稀缺场景具有较大优势。针对噪声标定复杂费时的问题,Jin等人(2023b)提出了一种无需校准暗光图像去噪方法,通过解耦相机信息和噪声模型,实现了对目标相机的快速适配。即无论数字增益或相机传感器如何,都无需重复训练,只需利用少量镜头配对数据进行微调即可适应目标相机。相较于单图降噪,基于连拍的降噪方法(burst image denoising)可融合多帧信息,取得了更出色的降噪效果。考虑到难以捕获动态场景的带噪一干净数据对,Liang等人(2020)提出了将运动对齐学习与噪声学习相解耦的去噪方法,从而利用动态仿真数据集与静态真实数据集的互补优势,取得了优异的连拍降噪效果。

面向极低光照条件下的图像增强在去除噪声的同时,也需恢复良好的图像色彩。该任务一般包括单阶段低光增强,即直接由像感域映射至颜色域(Chen等,2018;Lamba和Mitra,2021;Zamir等,2021),以及双阶段低光增强,即网络中间输出有对应约束(Xu等,2020a;Dong等,2022)。相较于单阶段增强方法,双阶段增强性能更优,但复杂度更高,容易导致误差累积。Huang等人(2022)针对像感域的额外信息使用(相较于颜色域)进行了详尽数值分析,初步评估了使用ISP过程的中间变量作为引导带来的增强性能增益。Jin等人(2023a)提出了基于分离重建思想的增强策略,有效避免了误差累积问题,取得了更优越的低光增强效果。

#### 1.1.2 Raw域视频去噪与增强

相较于图像去噪,视频去噪需同时考虑时空相关性,且像感域视频降噪数据集更难收集。考虑到难以收集成对的Raw视频增强数据集,Chen等人(2019)创建了一个低光视频数据集DRV(dark raw video),对同一个静态场景捕获多个带噪帧,之后通过约束不同带噪帧去噪结果的一致性来缓解视频降噪的时域抖动问题。所采用的ResUNet网络同时完成了去噪与ISP任务。与之不同,Yue等人(2020)采用定格动画的拍摄方式创建了第1个带噪一干净视频帧配对的动态视频数据集CRVD(captured raw video denoising),提出了像感域视频去噪网络RViDeNet(raw video denoising network)。与颜色域视频处理网络EDVR类似,该网络包括帧间对齐、非局部注意力和时空融合等模块以充分利用空间、通

道和时间相关性。该网络在去噪模块后加入预训练的ISP模块完成像感域到颜色域的转换,增强了模型的灵活性。鉴于RViDeNet复杂度较高,Maggioni等人(2021)提出了一种高效的多级视频去噪算法EMVD(efficient multi-stage video denoising),通过时域融合,预去噪与修正3个模块,在不影响甚至提高去噪性能的同时大幅降低了计算复杂度。考虑到定格动画拍摄数据的方式耗时长且运动模式简单等问题,Fu等人(2023)通过拍摄屏幕视频的方式构建了低光下具有复杂运动 and 高质量干净帧的Raw域视频去噪数据集,并提出了基于Transformer的视频降噪网络,取得了优异的降噪效果。

## 1.2 Raw域去噪与增强数据集

深度学习模型高度依赖训练所用数据的质量与规模。如何制作大规模高质量数据是影响像感域降噪与增强性能的重要因素。考虑到制作大量真实配对的“带噪—干净”数据对非常耗时且昂贵。因此,采用合成噪声数据集进行训练也是一种常见策略。本节介绍像感域噪声合成以及实际像感域成对数据集构建的相关进展。

### 1.2.1 Raw域噪声合成

噪声的精确仿真建模一直是学界与业界重点研究问题,在深度学习时代,该问题变得尤为重要。本节介绍两类典型的像感域噪声合成方法。

1)基于物理模型的Raw域噪声合成。基于物理模型的像感域噪声建模逐步分析噪声产生的物理机制,并据此对成像过程中的每类噪声分别建模。Foi等人(2008)将像感域噪声建模为散粒噪声(shot noise)与读出噪声(read noise),前者符合与信号相关的泊松分布,后者符合与信号无关的高斯分布。Wang等人(2020)在“泊松—高斯”噪声模型的基础上提出了K-Sigma变换,可将不同ISO(International Organization for Standardization)设置下的噪声图像映射到ISO不变噪声空间,提升了小网络处理大范围噪声的性能。

考虑到泊松—高斯噪声模型难以准确建模极低照度下的传感器噪声,Wei等人(2020, 2022)提出了一种基于物理先验的噪声模型,对传感器噪声进行细粒度建模,包括光子散粒噪声、像素电路噪声和量化噪声等,提升了极暗光照下的噪声建模精度。Zhang等人(2021)提出基于采样的噪声建模方法,通过采样拍摄的暗帧来模拟与信号无关的噪声真

值,提出了拜尔模式对齐的图像块采样方法以及高位重建方法有效提升了极暗光照下的噪声建模精度。Feng等人(2024)针对低光Raw图去噪任务中“带噪—干净”数据映射的可学习性问题,提出了增加数据数量、提高数据质量、降低数据复杂度等策略,以提升数据的可学习性,进而显著提升了模型的降噪性能。

2)基于深度学习模型的Raw域噪声合成。与基于物理统计的模型不同,基于深度学习模型的像感域噪声合成旨在利用神经网络生成匹配特定相机设置的带噪样本。Abdelhamed等人(2019)以光照强度、相机类型和增益等关键变量为条件,构建了一个归一化流模型Noise Flow。该模型可以逼近相机噪声水平函数,有能力建模更复杂的噪声形式。Chang等人(2020)提出了基于生成对抗网络的相机感知(camera-aware)噪声样本生成模型,通过引入相机编码网络,使得同一个噪声生成模型可适配多种不同相机。所提方法提高了噪声生成精度,进而提升了去噪效果。Zou和Fu(2022)提出了一种基于对比学习的细粒度噪声估计模型,所提对比学习策略使得模型摆脱了对实际带噪—干净数据对的需求,仅需一幅带噪图像即可估计出其噪声参数,进而仿真生成带噪图像。考虑到基于物理模型以及基于深度学习的模型在噪声建模方面各有优劣,Zhang等人(2023)采用物理模型建模与信号相关的噪声,采用基于学习的方式建模与信号无关的噪声。此外,为了更好地区分合成噪声和真实噪声,其提出了一种有效的多尺度鉴别器,使得合成噪声更加逼近真实噪声。然而,基于学习的噪声建模方式大多依赖“带噪—干净”数据对进行训练,所以模型效果与训练所用数据集高度相关,对未见数据的泛化能力值得继续探讨研究。

综上,基于物理模型的噪声合成依赖于物理模型的准确度,而基于深度学习的噪声合成依赖于数据的完备性。在实际应用中,准确度与完备度均难以实现。例如散粒噪声方差的标定,标定结果存在一定随机性,且受到设备温度的影响。基于深度学习的方式可通过数据拟合估计噪声分布参数,但噪声参数同时受限于所用的数据。因此,噪声估计依然有诸多挑战性问题。未来可通过针对性的数据推广来缓解该问题。

### 1.2.2 Raw域图像去噪与增强数据集

本节梳理介绍7种常见的像感域图像去噪及增强数据集,统计信息见表1。

1)RENOIR。Anaya和Barbu(2018)率先构建了包含噪声图像和干净图像的真实去噪数据集RENOIR(a dataset for real low-light image noise reduction)。RENOIR总共包含120个场景,每个场景包含两幅真值(ground truth)图像与两幅带噪图像,其中真值图像由两幅低ISO图像取平均获得。该数据集中真值图像与噪声图像的对齐效果欠佳,并且低ISO图像仍含有明显噪声。

2)DND。Plotz和Roth(2017)通过调整ISO数值和快门速度制作了DND(darmstadt noise dataset)数据集。将长曝低ISO图像作为干净图像,将高ISO(短曝光时间)图像作为噪声图像。之后经空间对齐、亮度校正、低频修正等过程,形成“带噪(高ISO)—干净(低ISO)”数据对。该数据集包含由4个相机拍摄的共50个室内外场景。该数据集目前用于评测不同去噪模型的性能。

3)SID。Chen等人(2018)制作了暗光下的SID数据集,用于训练测试低光图像增强模型。该数据集包含5094幅Raw域短曝图像,且每个短曝图像都有一个长曝图像作为干净的参考图像。由于同一场

景拍摄了多幅带噪图像,该数据集的干净图共包含424幅,即共有424个场景。在每个场景中,通过调整光圈、ISO和焦距等设置以最大限度地提高参考(长曝)图像的质量。

4)SIDD。Abdelhamed等人(2018)创建了用于去噪的SIDD(smartphone image denoising dataset)数据集。通过调整ISO数值(15种ISO, 50~10000)、光照色温条件(3种不同光照)以及明亮程度(低、中、高),用5个智能手机对10个场景进行了拍摄,每次拍摄150幅带噪图像,共收集了约30000幅带噪图像(10×5×4×150)。对于每幅带噪图像,通过特定处理生成其对应的干净真值图像。该处理过程精细,包含坏点矫正(defective pixel correction)、离群图像剔除(outlier image removal)、强度校正、局部图像配准和均值图估计等。该数据集数量大,质量较高,目前广泛用于去噪模型的性能评测。

5)ELD。为了评估噪声生成模型的有效性,Wei等人(2020)收集了一个极暗光增强数据集ELD(a physics-based noise formation model for extreme low-light raw denoising)。该数据集由4个相机拍摄,涵盖10个室内场景,并记录了每个相机的偏置和平场帧以提供初始的噪声参数。对于每个场景和每个相机,首先拍摄基本ISO下的干净参考图像,然后通

表1 Raw域图像去噪与增强数据集

Table 1 Public datasets for raw domain image denoising and enhancement

数据集	发表期刊/会议	场景数量	图像数量/幅	有无训练集	分辨率/像素	拍摄设备
RENOIR	arXiv2014	120	240	无	3 684 × 2 760 5 202 × 3 465 4 208 × 3 120	Xiaomi Mi3, Canon S90, Canon T3i
DND	CVPR2017	50	1 000(切块后)	无	512 × 512(切块后)	Sony A7R, Olympus E-M10, Sony RX100 IV, Huawei Nexus 6P
SID	CVPR2018	424	5 094	有	4 240 × 2 832 6 000 × 4 000	Sony a7S II, Fujifilm X-T2
SIDD	CVPR2018	10	30 000	有	5 000 × 3 000	Apple iPhone 7, Google Pixel, Samsung Galaxy S6 Edge, Motorola Nexus 6, LG G4
ELD	CVPR2020	10	240	无	5 000 × 3 000	Sony A7S2, Nikon D850, Canon EOS70D, Canon EOS700D
LRID	TPAMI2023	138	5 754	有	4 624 × 3 472	Redmi K30(IMX686)
LRD	ICCV2023	100	1 800	有	4 000 × 3 000	IMX586

注: CVPR: Conference on Computer Vision and Pattern Recognition; ICCV: International Conference on Computer Vision; TPAMI: IEEE Transactions on Pattern Analysis and Machine Intelligence。

过降低曝光时长来拍摄暗光带噪图像。共3种ISO(800, 1 600, 3 200)以及两种曝光时长衰减(100, 200),因此共包含240幅带噪图像( $3 \times 2 \times 10 \times 4$ )。

6)LRID。Feng等人(2024)构建了一个高质量的低光Raw图像去噪增强数据集LRID(likelihood ratio imbalance degree),共包含138个场景(82个室内场景和56个室外场景),总共5 754幅图像。首先在ISO 100下拍摄25幅长曝光图像,然后立即在ISO 6400下拍摄5组短曝光图像,每组10幅。长曝光图像经融合后形成干净帧,之后干净图像与带噪图像经过空间对齐、强度对齐操作后组成“带噪—干净”数据对。

7)LRD。考虑到SID数据集的长曝帧依然包含噪声,不适合用于评测像感域降噪性能。Zhang等人(2023)采集了一个低光像感域去噪数据集LRD(low-light raw denoising dataset)用于训练和去噪性能评测。与SID数据集设置固定曝光时间来捕获长曝光和短曝光图像不同,该数据集通过改变曝光值(exposure value, EV)来捕获长曝光和短曝光图像。对于每个场景,在ISO100下拍摄一幅长曝光图像作为无噪声的参考图像。然后,使用不同的ISO级别和EV值拍摄多幅短曝光图像,并在每幅图拍摄过程中停留1~2 s,以避免由于传感器过热带来的噪声。

### 1.2.3 Raw域视频去噪增强数据集

由于动态视频帧的干净数据难以采集,像感域视频去噪数据较少,本节综述4种由不同拍摄方式收集的动态视频降噪增强数据集,统计信息见表2。

1)分光镜拍摄数据集SMOID(learning to see moving objects in the dark)。由于难以同时采集同一场景的低光帧与正常帧,Jiang和Zheng(2019)提出利用分光镜来同时捕获正常帧与低光帧。环境光通过中继透镜以调整焦距和光束方向,然后通过分光镜将相同的光照视野提供给两个相机。其中一台摄

像机配备了中性密度滤光片(ND filter,透光率1%),因此采集的视频较暗。另一台相机无滤光片,按正常光照条件拍摄。两个摄像头通过信号发生器同步,以视频帧速率拍摄Raw序列。

2)定格动画拍摄数据集CRVD(captured raw video denoising)。考虑到难以对动态场景同时采集带噪帧与干净帧,Yue等人(2020)提出利用定格动画拍摄法进行采集。对于当前场景,首先拍摄其带噪帧与干净帧(低ISO多帧取平均),然后调整移动当前场景中的物体,继续拍摄带噪帧与干净帧。之后将拍摄的干净帧(带噪帧)按时间顺序相连,形成配对的干净视频与带噪视频。由于室外物体场景难以控制,该数据集的成对数据仅包含室内场景。对于室外场景,该数据集仅包含带噪视频。

3)相机移动拍摄数据集SDSD(seeing dynamic scene in the dark)。由于构造动态场景较为困难,Wang等人(2021)通过移动相机进行拍摄动态视频。其使用机电系统来精确控制滑轨相机的运动。为了捕捉一对视频,需将滑轨运行两次。在第1轮中,捕获无噪、明亮的视频,具有良好的对比度和鲜艳的色彩;在第2轮中,在相机镜头上安装中性滤光片并提高相机ISO以拍摄噪点严重的低光视频。

4)屏幕拍摄数据集LRMVD(low-light raw video denoising with a high-quality realistic motion dataset)。考虑到采用分光镜拍摄会降低真值图像的质量,而定格动画拍摄和相机移动拍摄会导致运动较为单一,Fu等人(2023)提出屏幕拍摄法。拍摄对象为公开的70个高质量的4 K视频,将其在DELL U2720QM显示器上播放,然后使用索尼Alpha 7R IV相机进行拍摄。通过调整屏幕和摄像机之间的距离来避免拍摄产生摩尔纹。与先前的数据集相比,此数据集具有更丰富的场景内容。但由于是翻拍屏幕,带噪图像的画质与自然图像存在差异。

表2 Raw域视频去噪与增强数据集

Table 2 Datasets for raw domain video denoising and enhancement

数据集	发表期刊/会议	场景数量	图像数量/幅	噪声等级数量	分辨率/像素	拍摄设备
SMOID	ICCV2019	179	约35 800	5	1 800 × 1 000	FLIR GS3-U3-23S6C
CRVD	CVPR2020	11	3 850	5	1 920 × 1 080	IMX385
SDSD	ICCV2021	150	37 500	1	1 920 × 1 080	Canon EOS 6D Mark II
LRMVD	TMM2022	70	约3 000	6	9 504 × 6 336	Sony Alpha 7R IV

注:TMM:IEEE Transactions on Multimedia。

### 1.3 总结与讨论

综上,像感域去噪与增强可分为图像去噪与增强、视频去噪与增强。在图像去噪与增强方面,根据网络设计的不同,可将其分为单阶段(Raw2Raw)、双阶段(串行的Raw2Raw2RGB)以及解耦网络(并行的Raw2Raw与Raw2RGB)。根据数据利用方式的不同,可分为全监督、半监督与自监督方法。全监督方法需要利用较多成对数据,半监督方法只需要采集特定传感器的少量成对数据即可实现降噪,而自监督方案只需要带噪数据即可实现降噪。在视频去噪与增强方面,根据网络设计的不同,可分为单阶段(Raw2RGB)以及双阶段(Raw2Raw+固定ISP)方法。根据所用数据的不同,可分为基于静态多帧数据的视频去噪、动态定格动画去噪、翻拍屏幕数据去噪以及相机移动数据去噪。图3总结了上述类别的相关

工作。

虽然像感域降噪与增强取得了显著进展,但依然存在诸多挑战性问題待解决。首先,在模型设计方面,对于像感域增强任务,网络需同时完成ISP与降噪功能,但这两个任务优化方向不一致,如何基于单阶段网络完成双任务值得探索;其次,在数据集构建方面,从真实场景中采样的视频降噪数据集依然难以兼顾运动真实性与细粒度纹理(屏幕翻拍导致一定程度纹理失真),如何提升数据集质量值得进一步探索;最后,在模型性能评测方面,目前用于评测的数据规模和种类较少,难以全面测量评估不同降噪算法的优劣,而且单纯峰值信噪比(peak signal-to-noise ratio, PSNR)、结构相似性(structural similarity, SSIM)等像素域评价指标难以反映实际的降噪视觉效果。

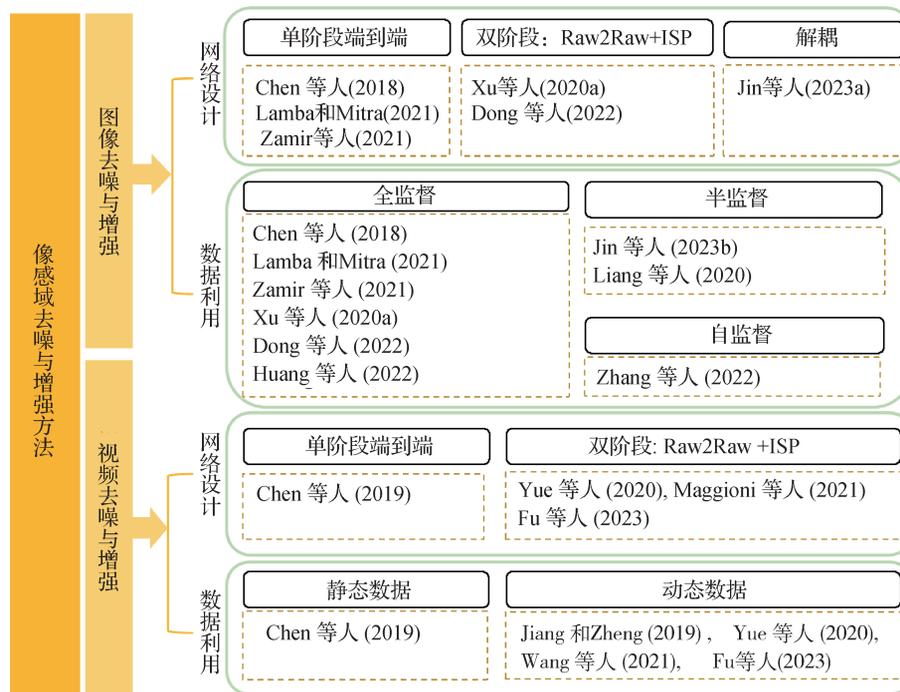


Fig. 3 Raw domain image (video) denoising and enhancement methods

## 2 Raw域超分辨率

超分辨率(super-resolution, SR)任务旨在利用计算成像方法从1幅或多幅低分辨率图像重建质量更高、细节更丰富的高分辨率图像。图像超分辨率可以在相机传感器等硬件条件受限的情况下,实现给定场景的高分辨率观测。随着深度卷

积神经网络的迅猛发展,基于学习的超分辨率方法已经取得了令人瞩目的成就。然而,目前大多数方法面向ISP处理后的颜色域图像。近年来,人们开始着眼于利用信息更丰富、噪声分布更简单的像感域数据进行重建。本节将依次介绍针对于像感域设计的单图像超分辨率方法、Burst图像(快速连拍图像)超分辨率方法以及视频超分辨率方法。



## 2.1 Raw域超分辨率方法

### 2.1.1 Raw域单图像超分辨率

与像感域图像去噪及增强的方法类似,像感域单图超分辨率方法一方面需要考虑与ISP联调以生成可视化的sRGB图;另一方面,为了实现基于学习的像感域图像超分辨率,构建高低分辨率图像成对数据集也至关重要。Zhang等人(2019)通过光学变焦策略,用单反相机拍摄同一场景的“低分辨率—高分辨率”像感域图像对。通过对比同样场景不同数据形式训练得到的超分模型,即“sRGB(输入)—sRGB(输出)”、合成Raw-sRGB以及Raw-sRGB等3种模型,证明了以像感域图像为输入的超分模型性能显著优于另外两种模型,充分说明了像感域超分的优越性。另外,提出了对非严格对齐数据对更友好的CoBi(contextual bilateral loss)损失函数,以缓解高分辨率与低分辨率图像无法准确对齐对训练过程的影响。考虑到难以获得对齐的真实降质高低分辨率图像对,Xu等人(2019)设计了一种新的数据生成方法,模拟数码相机的成像过程,以合成像感域和颜色域的高低分辨率数据对,并提出了一个基于卷积神经网络(convolutional neural networks, CNN)的双分支网络:第1个分支利用像感域数据恢复清晰的结构和精细的细节;第2个分支则使用颜色域的图像作为参考来估计色彩转换矩阵,从而能恢复具有高保真度的颜色信息。通过联合训练这两个分支,较好地实现了像感域到颜色域的跨域映射,显著提升了像感域图像超分辨率性能。考虑到由像感域到颜色域的转换(ISP)过程会带来额外的计算开销,Tang等人(2022)提出通过卷积神经网络建模低维空间的像感域图像降质过程,并与高分辨率重建网络进行联合训练和优化,实现了输入输出均为像感域图像的端到端超分辨率重建。

### 2.1.2 Raw域Burst图像超分辨率

主流智能手机的图像采集设备已经支持在单次快门下快速获取连续多帧图像,即Burst图像。相比于单图超分辨率,多帧图像超分辨率需要考虑多帧之间的时域相关性以及拍摄过程中相机与物体的微小运动,因此该任务的主要挑战在于图像间精准对齐和多帧高效信息融合。Bhat等人(2021a)创建了首个像感域Burst超分辨率数据集BurstSR,提出基于光流的像素级显式对齐方法实现多帧之间的配准,并利用基于CNN的编码器—解码器结构和基于

注意力的融合模块实现自适应地融合每幅图像中的信息。Bhat等人(2021b)提出将最大后验(maximum a posteriori, MAP)目标从传统的图像空间转化到深度特征空间,利用一个解码器网络来对目标高分辨率图像进行重参数化,从而在高维特征空间中融合优化,进一步提升了像感域Burst超分辨率的有效性和鲁棒性。考虑到手持移动设备的抖动会引入未知的运动,Lecout等人(2021)提出了针对像感域Burst图像的联合对齐和超分辨率方法,设计了一个可微的图像配准模块,提升了模型对于不同运动场景的重建效果。针对在其他任务的数据集上预训练的光流网络难以正确建模Burst图像之间的复杂运动的问题,Dudhane等人(2022)提出了BIPNet(burst image restoration network),利用可变形卷积技术在特征域隐式对齐多帧。另外,提出了反投影操作进行边缘的增强和细化以及伪-Burst特征融合策略,提高了超分辨率性能。考虑到直接利用可变形卷积的对齐容易丢失细节信息,Mehta等人(2023)提出了GMTNet(gated multi-resolution transfer network),通过引入注意力引导的对齐策略和基于转置注意力的特征融合模块,更好地建模多尺度非局部上下文信息,进一步提高了对于局部和全局信息的利用效率。为了增强对复杂的不对齐情况的鲁棒性,Dudhane等人(2023)进一步提出Burstormer,提出渐进式的聚合和上采样模块以增强帧与帧之间的信息交互,获得了出色的重建效果。

### 2.1.3 Raw域视频超分辨率

相比像感域的单幅图像和Burst图像,像感域视频信号的多帧之间通常具有更大的物体位移和更复杂的场景变换。因此,像感域视频超分辨率任务更需要探索高效的时空相关性提取和信息融合技术,以恢复缺失的细节并增加图像的清晰度。近年来,已有学者在像感域的视频超分辨率领域进行了相关探索。考虑到像感域高低分辨率的视频数据对难以直接获取,Liu等人(2021)通过模拟像感域信号的降质过程,构建了一个合成的像感域视频超分辨率数据集RawVD,并提出了双分支超分网络RawVSR,分别用于像感域高分辨率重建和颜色矫正。重建分支采用逐次相邻对齐和基于注意力的融合策略实现多帧信息融合,以高效利用时域相关性。颜色矫正分支通过ISP生成颜色域的参考信息,并通过网络生成变换矩阵来指导像感域图像到颜色

域的转换。考虑到在合成数据上训练的模型难以泛化到真实的像感域视频超分辨率任务上, Yue 等人(2022)构建了首个真实降质的像感域视频超分辨率数据集, 通过分束器构造了双相机系统, 给双相机分别设定长短焦距, 分别获取高分辨率和低分辨率的视频信号。相应地, 提出了一个基于像感域视频序列的双分支交互超分辨率网络, 将像感域视频帧以交织 Bayer 模式和分离 Bayer 模式一起输入网络, 通过特征对齐和交互, 有效利用多帧之间的

时空相关性, 获得了优越的 Raw 域视频超分辨率性能。

## 2.2 Raw 域超分辨率数据集

现有的像感域超分辨率数据集大多依赖光学变焦或双相机的技术实现对同一场景的高低分辨率图像的获取, 随之带来了数据对齐、噪声抑制和时间同步等方面的挑战。本节介绍现有的像感域单图像、Burst 图像和视频数据集的构建方法和策略。相关数据集的统计对比见表 3。

表 3 Raw 域超分辨率数据集

Table 3 Public datasets for raw domain super-resolution

数据集	类型	帧数	LR 图像获取方式	拍摄设备	倍率	训练集数量/幅	测试集数量/幅	HR 图像分辨率/像素
SR-RAW	单图像	1	真实拍摄	单反相机	2×~8×	450	50	约 4 000 × 2 500
ImagePairs	单图像	1	真实拍摄	单反相机	2×	8 591	2 830	3 504 × 2 332
BurstSR	快速连拍	14	真实拍摄	智能手机/单反相机	4×	180	20	640 × 640
RealBSR	快速连拍	14	真实拍摄	单反相机	4×	20 842	2 377	640 × 640
RawVD	视频	≥100	合成	单反相机	2×/4×	105	5	1 920 × 1 080
Real-RawVSR	视频	150	真实拍摄	单反相机	2×/3×/4×	130	20	640 × 1 440

注: 验证集(若有)数量被合并计入训练集。

### 2.2.1 Raw 域单图像超分辨率数据集

1) SR-RAW。Zhang 等人(2019)构建了首个像感域单图像超分辨率数据集。该数据集使用 Sony FE 24-240mm 焦段的可变焦镜头进行拍摄, 每个场景均使用 7 个不同焦距拍摄一组图像, 以用于不同放大倍数的超分辨率网络训练。为了解决调整焦距时因相机轻微移动而导致的不对齐, 采用了视场匹配和几何变换的方式进行数据后处理, 将捕获的图像在不同缩放级别下依次进行对齐。

2) ImagePairs。Joze 等人(2020)构建了一个大型的包含像感域与颜色域图像的超分辨率数据集。该数据集使用分束器保证两台不同焦距的相机捕获相同的场景。在获取低分辨率和高分辨率图像对后, 应用 ISP、图像去畸变、成对图像对齐以及边缘裁剪等操作, 进一步减小高低分辨率图像对之间视角的差异。

### 2.2.2 Raw 域 Burst 图像超分辨率数据集

1) BurstSR。Bhat 等人(2021a)构建了首个像感域 Burst 图像超分辨率数据集 BurstSR。该数据集使用三星 Galaxy S8 捕获低分辨率的 Burst 图像序列, 高分辨率的图像序列则由佳能 5D Mark IV 单反相机

收集。为了缓解高低分辨图像间的不对齐问题, 选择将手机固定在单反相机上方拍摄具有微小运动的场景。

2) RealBSR。针对 BurstSR 的数据中仍然存在不对齐和不同设备获得的图像之间存在域偏移的问题, Wei 等人(2023)提出了 RealBSR (real-world burst image super-resolution) 数据集, 使用 Sony Alpha 7R 相机进行拍摄。经过光学变焦策略得到高低分辨率的数据对之后, 以高分辨率图像为参考, 在低分辨率图像上使用尺度不变特征变换匹配算法 (scale invariant feature transform, SIFT) 算法进行数据对齐。该数据集同时包含像感域和颜色域的数据对。

### 2.2.3 Raw 域视频超分辨率数据集

1) RawVD (raw video dataset) (Liu 等, 2021) 是一个合成的像感域视频超分辨率数据集, 高分辨率像感域视频使用佳能 5D3 单反相机拍摄。得到相机拍摄的高分辨率数据集后, 经过散焦模糊、降采样和加噪声等操作模拟降质过程, 得到低分辨率的像感域视频序列。

2) Real-RawVSR (real-world raw video super-

resolution)(Yue等,2022)是第一个真实降质的像感域视频超分辨率数据集。为采集该数据集,利用两台不同焦距的佳能60D相机构造双相机系统,配置了带有反射涂层(反射率50%)和透射涂层(透射率50%)的分束器,以使两个相机接收来自同一视点的同强度自然光和场景。为了解决高低分辨率之间的数据不对齐问题,使用全局单应(homography)变换以及局部的Deepflow变换进行配准对齐,获得了较好的对齐效果。

### 2.3 总结与讨论

综上,像感域图像视频超分辨率按照网络模型的不同可分为单阶段超分重建(Raw2RGB)、双阶段

超分重建(Raw2RGB,通过颜色校正矩阵恢复色彩)以及只在像感域重建(Raw2Raw)等3类。考虑到超分辨率与颜色恢复是两个不同的优化任务,且颜色复原错误对网络损失值影响更大。因此网络会过多关注颜色复原程度而非细节修复,如何增强超分任务中的细节复原值得探索。另外,多帧超分对不同帧之间的对齐精度有更高要求,因此,有诸多方法探索对齐策略。根据对齐方法的不同,可将其分为光流法对齐、金字塔式可变形变换对齐以及注意力式对齐等3种类型。探索更为精确的对齐方法将有助于进一步提升多帧超分辨率性能。图4总结了像感域不同的超分辨率方法。

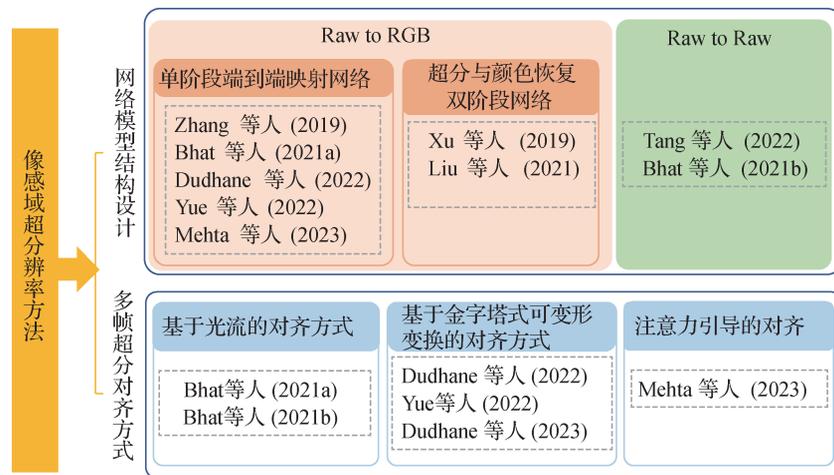


图4 像感域超分辨率方法

Fig. 4 Raw image(video) super resolution methods

## 3 Raw域HDR重建

### 3.1 HDR重建问题概述

在自然场景中,物体的反射亮度具有极大的动态范围,通常约为 $10^6:1$ 或者更高。然而,目前大多数数字摄影传感器仅能测量和存储有限范围内的亮度变化,例如图像位数大多限制在8 bits,性能较高的传感器可以达到10 bits或12 bits,但仍然无法涵盖自然场景广泛的亮度范围。因此,产生的低动态范围(low dynamic range, LDR)图像常常存在曝光过度或曝光不足等问题,导致了亮区和暗区的细节损失。高动态范围(HDR)技术可以从低动态范围(LDR)图像中恢复高光区域和暗光区域的细节信息,从而丰富单幅图像的场景信息,使其更贴近人类视觉的感知。

根据处理的数据类型,可将其分为HDR图像重建和HDR视频重建。基于多曝光图像的HDR图像重建方法是目前的主流方法,其通过对同一场景拍摄多幅不同曝光的LDR图像,经融合得到HDR图像。所得HDR图像的动态范围与拍摄的LDR图像数量或者图像间的曝光差异呈正比。然而在动态场景中,由于不同曝光LDR图像间存在位移,重建的HDR图像结果容易出现鬼影(ghosting effect),严重影响了图像的视觉质量。因此,学界提出了多种传统及深度学习方法来应对。

视频HDR重建与基于多曝光的图像重建类似,主要通过拍摄同一场景下多个不同曝光的LDR视频序列来重建HDR视频,包括基于硬件的方法以及基于后处理的软件方法。基于硬件的方法依赖于特殊传感器或多传感器实现,且都在像感域进行处理。例如行交织传感器(Çoğalan and Akyüz, 2020)和四

元组传感器(Jiang等,2021)等能够在一帧图像内的不同空间位置记录不同曝光设置下的像素大小,然后将其融合为HDR视频,但由于在有限的传感器尺寸上复用多种曝光,牺牲了空间分辨率。另一种方式是利用分光镜和多个传感器同时捕获不同曝光的视频序列(Hu等,2013;Unger等,2016;Froehlich等,2014),通过将照入镜头的同一束光分别投射到不同传感器上,所得LDR视频完全对齐且无分辨率损失。但系统构型过于复杂且臃肿,难以应用到普通相机或手机等移动设备中。在后处理算法方面,近年基于深度学习的HDR重建方法取得了出色的重建效果。

### 3.2 Raw域HDR重建

如上所述,内嵌到相机中的传统HDR重建算法均在像感域进行处理。然而学界大部分HDR重建工作都是基于颜色域的LDR图像进行重建。但颜色域数据相较于像感域数据,存在非线性变化和细节损失问题。像感域数据拥有更高的位深度和更广的亮度范围,保留了更多的细节,可弥补场景信息的不足,因此基于像感域数据的HDR重建技术相对于颜色域数据具有更大的潜力和优势。

#### 3.2.1 Raw域HDR图像重建

在传统像感域HDR重建方面,Kronander等人(2013)针对以往像感域HDR算法需要去马赛克、去噪、重采样及融合等多个步骤的问题,提出了一个基于空间自适应的统一框架进行HDR重建。该方法计算复杂度低,实时性强。Lecouat等人(2022)提出了基于多幅不同曝光连拍图像的像感域超分辨率与HDR重建方法,采用了物理成像模型,通过迭代优化算法解决成像逆问题,获得了快速高效的HDR重建。近几年,基于深度学习的像感域HDR重建逐渐

获得关注。Chaudhari等人(2019)利用一个网络实现了多幅像感域LDR图像到颜色域HDR图像的映射,避免了HDR重建与ISP变换级联带来的误差传播。Dai等人(2024)提出利用小波变换将运动伪影和细节在频域分开,分别实现细节恢复和鬼影去除,取得了比颜色域模型更好的效果。Zou等人(2023)通过学习曝光掩膜来区分高动态场景的曝光不良和良好区域,利用像感域不同颜色通道的数据分布特点,对曝光不良区域利用信息更丰富的通道引导细节重建,取得了出色的重建效果。Liu等人(2023)首次开发了通过使用手机摄像头的HDR图像数据集,并覆盖了不同噪声水平的白天和夜间场景。

#### 3.2.2 Raw域HDR视频重建

Kronander等人(2013)基于传统算法使用像感域数据进行HDR视频重建。由于缺少相关数据集,基于深度学习的像感域HDR视频重建鲜有研究。针对此,Yue等人(2023c)基于Stagger传感器构建了一个包含像感域和颜色域的LDR-HDR视频对的数据集。交错式(staggered)传感器可近乎同时捕获两幅交替曝光的图像,有效避免了空域、时域难以对齐给HDR重建带来的问题。在此数据集基础上提出了像感域HDR重建算法,包括预先白平衡、对齐、融合和后处理模块等模块。

### 3.3 Raw域HDR数据集

目前像感域HDR数据集仍比较有限。本文选取4个代表性的开源数据集进行介绍,统计对比见表4。Chen等人(2021)提出的DeepHDRVideo数据集包含281组带有HDR真值的图像对,使用BasleracA4095-30uc相机拍摄。该数据集为交替曝光类型,每5帧的中间帧对应拍摄一张HDR真值数据,曝光类型包括长、中、短3种。Zou等人(2023)提

表4 Raw域HDR重建数据集

Table 4 Public datasets for Raw domain HDR reconstruction

数据集	类型	重建方法	曝光种类	输入/输出	发表会议	图像数量/幅	训练集	分辨率/像素	拍摄设备
DeepHDRVideo	视频	交替曝光	两/三次曝光	12 bits/-	ICCV2021	281	无	4 096 × 2 168	Basler acA4095-30uc
Hanji	图像	单图重建	一次曝光	14 bits/-	SIGGRAPH2022	183	有	1 920 × 1 280	Canon 5D Mark III
Mobile-HDR	图像	交替曝光	两次曝光	10 bits/16 bits	CVPR 2023	251	有	4 000 × 3 000	OPPO A95
								4 096 × 3 072	HONOR 70 Pro
RawHDR	图像	单图重建	一次曝光	14 bits/20 bits	ICCV2023	324	有	6 720 × 4 480	Canon 5D Mark IV

注:SIGGRAPH:Special Interest Group for Computer Graphics,“-”表示无对应数据。

出的单图 HDR 重建数据集(RawHDR)包含 324 个 Raw/HDR 图像对,使用 Canon 5D Mark IV 相机拍摄。Hanji 等人(2022)提出的单图 HDR 重建数据集包含 183 个 Raw/HDR 图像对,使用 Canon5DMark III 相机拍摄。Liu 等人(2023)通过多款智能手机捕获了首个真实世界运动 HDR 图像数据集(Mobile-HDR)。

### 3.4 总结与讨论

综上,像感域 HDR 重建的核心是如何融合不同曝光图像中互补的视觉信息。由于不同图之间存在运动,融合后易出现鬼影问题,因此高效去鬼影是 HDR 重建重点关注的问题。考虑到过曝区域以及极暗区域细节少,难以提供对齐所需要的纹理先验,因此当前方法在该类区域的 HDR 重建效果尚存在提升空间。后续可通过深入挖掘像感域图像视频在该类困难区域所提供的信息来缓解该问题。

## 4 Raw 域去摩尔纹

### 4.1 去摩尔纹问题定义

当两个结构相似、频率相近的细密网格重叠时,就会产生摩尔纹,如图 5 所示。在拍摄细密织物或屏幕时,容易出现摩尔纹,而且摩尔纹会随着两个网格之间的角度及距离变化而呈现出不同的形态。尤其是对于屏摄图像的摩尔纹,镜头或显示器的旋转、相机传感器感光阵列及显示屏二极管阵列的分辨率、相机拍摄时的参数设置等都会影响摩尔纹的形态、颜色和强度,导致其在表征空间上与正常图像信号存在交叠。因此,在保留图像细节的同时去除摩尔纹极具挑战。传统摩尔纹去除方法大多是在相机中增加防混叠滤波器,但该措施难以解决屏摄以及大范围细密纹理等挑战场景。目前的深度学习去摩尔纹方法与数据集均针对颜色域图像。然而,颜色域中的摩尔纹由两部分因素导致,一部分是由于相机感光网格阵列和显示面板阵列之间的频率混叠产生的;另一部分是由于 ISP 去马赛克模块插值不准确而产生。与之不同,像感域图像的摩尔纹未经 ISP 非线性变换,成因仅为频率混叠,其特性相对而言更容易刻画。因此,像感域去摩尔纹相较于颜色域更具优势。此外,拍摄屏幕经常导致颜色失真,如何在去摩尔纹的同时校正图像色差也是一个值得探索的方向。本节综述像感域去摩尔纹方向的进展与数据集情况。

### 4.2 Raw 域去摩尔纹方法

目前大多数算法均面向颜色域的摩尔纹去除,仅有少量研究探索像感域摩尔纹去除。Yue 等人(2023b)提出了一种基于像感域的屏摄图像去摩尔纹网络 RDNet(raw demoiréing network),该网络可以在像感域实现摩尔纹的有效去除,并通过预训练的 ISP 网络将去摩尔纹的像感域图像转换至颜色域,得到颜色域的去摩尔纹图像。网络整体结构主要由内容分类模块、多尺度特征融合编码模块、分离解码模块以及 ISP 模块组成。考虑到文字图像与自然图像的摩尔纹特性及纹理特性不同,RDNet 对这两类分别处理,提升了文字区域的去摩尔纹效果。通过共享编码、分离解码的方式提升了重建图像的纹理细节恢复质量。同时该工作公开了首个像感域去摩尔纹数据集。在此基础上,Yue 等人(2023a)进一步揭示了像感域摩尔纹的频率和空间分布在颜色通道中的差异性,提出了通道和空间调制(channel and spatial modulation, CSMod)方法以充分利用颜色分离特征和颜色混合特征,取得了出色的去摩尔纹效果。不同于端到端的单阶段去摩尔纹方法(Raw2RGB),Xu 等人(2023b)提出了双阶段去摩尔纹方法。一个支路用于像感域去摩尔纹,另一个支路用于颜色域去摩尔纹,之后用颜色域结果引导像感域的 ISP 过程,取得了出色的颜色恢复效果。

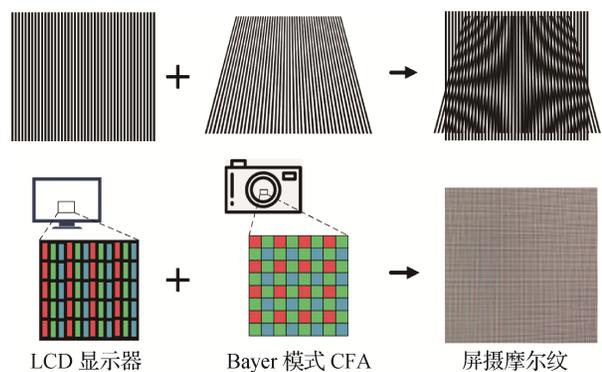


图5 摩尔纹形成过程

Fig. 5 Formation process of moiré patterns

### 4.3 去摩尔纹数据集

现有的去摩尔纹数据集包括合成数据集和真实数据集。现有的合成去摩尔纹数据集包括 LCD-moiré(Yuan 等,2019)和 CFAMoiré(Yuan 等,2020),其合成纹理单一,与真实摩尔纹存在较大的差异。sRGB 颜色域真实去摩尔纹数据集包括 TIP2018(Sun

等,2018)、MRBI(moiré removal and brightness improvement)(Yue等,2021)、FHDMi(full high definition moiré image)(He等,2020)和VDMoiré(video demoiréing dataset)(Dai等,2022)等,这些数据集难以兼顾摩尔纹多样性和对齐准确性。见表5,像感域去摩尔纹数据集依赖时域和空域对齐技术实现成对图像的获取,并利用多种拍摄设备提高多样性。本节介绍现有的像感域图像和视频去摩尔纹数据集的构建方法,统计信息见表5。

1)RawDM(raw demoiréing)(Yue等,2023b)是第1个真实像感域图像去摩尔纹数据集。该数据通过使用4部手机和3个显示器组成共8组配置进行拍摄。为解决拍摄的摩尔纹图像和显示的原图之间的空域不对齐的问题,该方法提出使用全局单应变换(homography transform)和Deepflow算法进行后处理。切块后,该数据集包含63 180幅分辨率为 $256 \times 256$ 像素的训练集图像和408幅 $512 \times 512$ 像素的测试集图像。

2)RawVDM(raw video demoiréing)(Yue等,2023a)是第一个真实像感域视频去摩尔纹数据集。该数据集通过使用4组手机—显示屏组合进行拍摄,包含丰富的视频内容。为解决拍摄视频和显示视频之间的时域对齐问题,提出利用显示视频前后插入的白帧定位视频起止位置,并检测扩展斜线图案的斜率获得时域对齐帧。为解决拍摄视频和显示视频之间的空域对齐问题,使用透视变换和Deepflow算法进行后处理。该数据集包含300个720 p的视频,每个视频包含60帧。

#### 4.4 总结与讨论

综上,像感域去摩尔纹尚处于研究初期,目前仅利用了图像类别先验以及像感域的通道先验。由于目前的去摩尔纹工作采用截屏或者源图作为真值图

像,所以去摩尔纹过程需完成色调转换任务。色调恢复与摩尔纹去除是两个不同的任务,如何平衡这两者的性能尚需探索。另外一个可行的方式为保持翻拍的色调,仅聚焦摩尔纹去除任务。

## 5 Raw域多任务联合重建

成像过程常面临着欠采样、噪声等多种降质耦合共存的情况,需要完成去马赛克、超分辨率和去噪等多种任务以获得高质量图像重建。在实际应用中,将这些问题分离考虑,并在应用时直接串联使用,会导致误差累积。因此,近期研究人员开始探索Raw域上的多任务联合重建问题。例如,考虑到噪声对于去马赛克过程的干扰,Liu等人(2020)利用绿色通道具有采样率高、噪点少的优势,设计了绿色通道指引重建,利用优先恢复的绿色通道的先验信息指导其余通道的恢复过程,实现了像感域图像去噪和去马赛克任务的联合重建。延续这一思路,Guo等人(2021)进一步利用绿色通道的先验信息估计多帧图像之间的偏移,在Burst图像的去噪及去马赛克任务上取得了较好的表现。鉴于联合进行去马赛克和超分辨率任务能够去除将两者分别进行时产生的图像瑕疵,Xu等人(2020b)设计了预去马赛克网络与密集残差网络,通过级联机制促进了预去马赛克结果与最终重建结果之间的信息交互,共同提升了去马赛克和超分辨率的重建质量。考虑到多个任务依次进行将导致误差累积,Qian等人(2022)提出了端到端的网络TENet++以规避该问题,实现了联合的去马赛克、去噪及超分辨率。针对Burst图像的HDR与超分辨率任务,Lecouat等人(2022)利用基于特征学习的图像配准算法得到更具鲁棒性的多帧信息融合权重,并基于半二次分裂算法对融合过程进

表5 图像/视频去摩尔纹数据集

Table 5 Public datasets for image and video demoiréing

数据集	域(类型)	良好对齐	组合数	分辨率/像素	图像数量/幅
TIP2018	sRGB(图像)	是	9	$\leq 700 \times 700$	135 000
FHDMi	sRGB(图像)	否	6	$1\,920 \times 1\,080$	12 000
RawDM	Raw+sRGB(图像)	是	8	$1.8\text{ K} \times 1.3\text{ K}, 512 \times 512$	948
VDMoiré	sRGB(视频)	否	2	$1\,280 \times 720$	17 400
RawVDM	Raw+sRGB(视频)	是	4	$1\,280 \times 720$	18 000

行迭代优化,高效地解决了HDR、超分辨率、去马赛克以及去噪任务。针对RGBW(red, green, blue, white)这一新兴的高信噪比CFA阵列,Fan等人(2023)设计了包含去噪和去马赛克的两阶段网络,分别使用基于Half-Shuffle Transformer结构的全通道动态卷积以及空间压缩Transformer结构,在去噪与去马赛克联合任务中取得了出色的性能。

综上,多任务联合重建方法可通过一个网络联合解决多个成像问题,但多任务相耦合也导致模型的可解释性变差。该类重建方法是否适合部署至终端尚需进一步验证。

## 6 Raw域数据生成

基于像感域的视觉重建均取得了优于颜色域的视觉重建效果。基于像感域数据的深度学习方法需要大量的像感域数据进行训练,但目前图像、视频大多以sRGB形式存储传输,难以直接获取大量Raw数据。因此,近年开始探索基于逆处理的方法(inverse ISP),从颜色域图像中直接生成像感域图像。

Brooks等人(2019)率先提出了基于Unprocessing的策略,对色调映射、伽马矫正、去马赛克、颜色矫正和白平衡的逆过程由可逆、可微函数进行近似,该逆ISP过程灵活、可解释,可适配各种像感域数据处理问题,被广为采用。Zamir等人(2020)提出CycleISP网络,利用两个不同的深度网络分别拟合Raw至sRGB再反向到Raw的过程,该网络学习与设备无关的转换,无需相机参数如颜色校正矩阵和白平衡增益等先验信息,更加具有通用性。Xing等人(2021)提出Invertible ISP网络,利用归一化流的可逆性质,基于归一化流构建单一可逆网络,同时学习ISP和逆ISP过程。除了RAW数据重建,该网络还可以应用于RAW数据压缩和HDR重建等任务。为了融合物理模型优化与深度学习的优势,Conde等人(2022)提出结合基于模型和基于数据驱动的混合逆ISP方法。该可逆模型对ISP过程中的镜头阴影校正、白平衡、颜色空间转换、伽马矫正和色调映射等过程进行建模,并通过网络学习出可解释的变换参数(字典)以控制上述过程,实现了像感域与颜色域之间的双向映射,取得了出色的逆变换效果。但由于学习模型大多受困于所用数据集,如何提升该类方法

对不同Raw数据的适配能力是一个值得探索的方向。

综上,现有研究可基于某种相机的训练数据生成面向该相机的像感域数据或者通过改变参数反演出更广的像感域数据。尚无研究探索如何利用现有的大量颜色域数据集生成适配多类传感器的像感域数据。目前扩散模型的生成能力在多个任务中得到了验证,后续有望通过扩散模型生成多类型的高质量像感域数据。另外,如何生成连续的像感域视频也值得进一步探索。

## 7 Raw域视觉重建应用现状

随着深度学习的迅猛发展,像感域视觉重建技术在效果上大幅超过传统方法,逐步应用于智能手机、安防监控和无人驾驶等多个领域。本文从像感域视觉重建任务出发,结合使用场景需求,综述其在智能手机、安防监控和无人驾驶等典型场景的广泛应用情况。

### 7.1 智能手机

在大规模数据和人工智能技术的驱动下,像感域视觉重建成为智能手机摄像系统中的重要技术。传统ISP算法在像感域去噪、HDR重建等任务上的性能趋于饱和。同时,智能重建算法能灵活满足新的应用需求,带来巨大应用价值,例如:夜景拍照、美颜拍照和风格调整等。这些因素进一步促进了基于人工智能(artificial intelligence, AI)计算的像感域视觉重建方法在智能手机摄像系统中的应用。智能手机由于孔径和传感器尺寸受限,更容易受到随机噪声和过饱和失真的影响,带来了动态范围低和图像质量退化等问题。Liu等人(2023)构建了第1个面向智能手机的像感域HDR重建数据集。随着硬件架构的发展和演进,众多研发机构也从像感域视觉重建的角度出发,研制了智能ISP芯片,如OPPO的马里亚纳、小米的澎湃C1、荣耀的爱芯元智等。随着神经处理器的架构演进,算力和数据吞吐量的提升,面向移动平台的像感域图像重建算法与芯片化研发将成为智能手机产业的热点方向。

### 7.2 监控安防

相比于智能手机,监控安防相机具有更小的传感器尺寸,更容易受到噪声和过饱和的影响。尤其是在低照度场景下,图像传感器感光能力急剧下降,导致彩色图像清晰度快速下降,无法同时兼顾图像

细节和色彩。补充光源可以提升光照度,但同时造成光污染甚至影响驾驶员的行车安全。另一种可行途径是采用多传感器获取更多信息,利用像感域视觉重建方法提升成像质量。Dong 等人(2022)通过将 Raw 图像与合成的单色图像融合来实现微光图像增强。产业界在安防监控领域研发了多款智能 ISP 产品。例如,华为越影 ISP 引擎利用深度学习来估计海量低照度场景图像的噪声和信号分布特性,实现了低照度场景下的智能降噪;采用像感域 HDR 技术,在逆光场景下将动态范围提升了 12 dB,并且实现场景快速跟踪,有效避免了图像画面瞬间过暗或者过曝的问题。

### 7.3 无人驾驶

无人驾驶场景对视觉成像和处理系统提出了更高的要求。手机摄影可以允许较大的延时(1~2 s),但是无人驾驶需要将延时控制在 ms 级,对图像视频数据的处理与传输速度提出了严格的要求。特斯拉跳过相机 ISP,直接从多机位的像感域数据提取特征,使用多种检测头实现车辆、行人、车道线、深度图、速度/方向等多类重要目标的检测。ARM 公司(Tseng 等, 2019)和 Algolux 公司(Beiker, 2023)也采用了类似的解决方案,不再进行 ISP 处理,而是采用 Raw 图像作为高级视觉任务的输入,实现端到端深度神经网络融合和处理。然而,受限于成像器与成像条件,车载像感域图像数据通常包含严重的噪声和干扰。因此,针对车载像感域数据特性设计轻量级重建是无人驾驶场景提升低光、强光和雨雾等恶劣成像条件感知能力的底层技术基础。

随着嵌入式芯片算力的提升,未来面向像感域的智能 ISP 模块(例如降噪、增强等)在矿井、安全应急和远洋船只等极端场景中将会有更多应用涌现。

## 8 结 语

### 8.1 总结

本文从 6 个方面综述了像感域视觉重建相关方法与数据集的发展路线。像感域去噪增强、超分辨率、HDR 重建、去摩尔纹和多降质联合重建等任务各有其独特难点。在视频降噪方面,如何构建兼顾真实运动与细粒度纹理的真实降噪数据集尚难以解决;在多帧超分辨率方面,如何构建高效精准的多帧对齐模块是其核心难点;在 HDR 重建方面,如何提

升困难区域(极暗区域或者曝光区域)的去鬼影效果仍需探索;在去摩尔纹方面,如何平衡色调恢复与摩尔纹去除效果需进一步研究;在多降质联合重建方面,如何提升模型的可解释、可调性是一个重点问题。

以上任务也存在诸多共性问题。例如,由于结果需要在颜色域观看,以上任务均涉及颜色转换问题,如何在完成各自任务的同时恢复正确的色彩信息是其共性问题。在构建数据集方面,多数数据集存在空间不对齐问题,如何精准对齐降质图像与真值图像是以上 5 类任务的共性问题。

### 8.2 展望

相较于颜色域视觉重建的繁荣与紧凑,像感域视觉重建相关研究还有巨大的空间和广阔的天地。像感域视觉重建领域的发展将有助于各类重建算法落地到成像端,进而显著提升传感器的成像能力。基于以上各节的总结和分析,关于像感域视觉重建,在未来有 4 个研究方向值得进一步推进。

1) 像感域数据特性挖掘。现有视觉重建算法大多利用像感域数据线性、宽位深的优势,有些工作开始探索像感域数据的独特特性。例如,像感域图像摩尔纹在不同颜色通道的频率特性不同;像感域不同颜色通道响应数值亮度不同,绿色通道数值相较于红蓝通道更高;像感域 Bayer 的非连续性在 Transformer 及卷积网络设计中的应用。如何在超分辨率及去噪等应用更广泛的任务中挖掘像感域数据的优势更具吸引力,也是亟待探索的方向。

2) 大规模像感域数据获取。数据匮乏是像感域视觉重建领域研究难以大面积展开的主要原因。虽然在多种像感域重建任务中均已构建相应的数据集,但这些数据集的规模依然远小于 ImageNet、LSDIR 等颜色域数据集。此外,像感域数据的特性依赖于传感器,提升传感器的多样性也是大规模数据采集的挑战。增强像感域数据的可获得性有助于快速提升数据集规模,并且便于用户直接评测。一个可行的途径是引入颜色域图像辅助的像感域图像压缩技术,通过端对端学习,利用极少量 meta 数据即可高质量解码出 Raw 图像(Wang 等, 2023)。该技术可复用相机已有的 JPEG 硬编码器,极大减少传输带宽,帮助目前相机同时直出像感域图像和颜色域图像。另一种可行途径是探索更为有效的 inverse ISP 方法,从颜色域图像中反演像感域数据。但是,



Yue 等人(2023a)的去摩尔纹工作表明现有 inverse ISP 方法难以恢复像感域不同颜色通道的数据特性,如何针对重建任务研发数据特性保持的 inverse ISP 方法值得进一步探索。

3)数据偏差消除。由于不同相机所采集的像感域数据具有较大的特性差异,基于特定相机采集的图像数据所训练的模型难以直接适配其他相机,在泛化性测试中性能均存在不可忽略的下降。如何消除数据偏差(data-bias)对像感域视觉重建方法的落地应用具有重要意义。一个可行的方向是采用物理模型和数据模型共同驱动的学习方式,使得某些参数具有物理可调性。另一个可行方向是挖掘不同相机像感域图像之间的关系,构建参数化映射模型,以消除域差异。

4)大模型加持下的像感域视觉重建。当前大模型发展如火如荼,如何基于大模型开展像感域视觉重建是一个值得探索的方向。由 SAM(segment anything model)到 MobileSAM 的发展可知,数据规模是提升模型性能的关键所在,而大模型知识的重利用将有助于显著降低模型复杂度。因此,亟需采集大规模像感域数据,构建视觉重建基座模型,探索面向终端部署模型的蒸馏技术,从而在计算资源受限的边端场景突破各类图像传感器的成像质量瓶颈,为视觉传感产业带来跨代次的革新。

**致 谢** 本文由中国图象图形学学会图像视频通信专委会组织撰写,该专委会链接为 <https://www.csig.org.cn/16/201612/49312.html>。

## 参考文献(References)

- Abdelhamed A, Brubaker M and Brown M. 2019. Noise flow: noise modeling with conditional normalizing flows//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 3165-3173 [DOI: 10.1109/ICCV.2019.00326]
- Abdelhamed A, Lin S and Brown M S. 2018. A high-quality denoising dataset for smartphone cameras//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 1692-1700 [DOI: 10.1109/CVPR.2018.00182]
- Anaya J and Barbu A. 2018. RENOIR——A dataset for real low-light image noise reduction. *Journal of Visual Communication and Image Representation*, 51: 144-154 [DOI: 10.1016/j.jvcir.2018.01.012]
- Beiker S. 2023. Next-Generation Sensors for Automated Road Vehicles. No. EPR2023003. SAE International, USA [DOI: 10.4271/EPR2023003]
- Bhat G, Danelljan M, Van Gool L and Timofte R. 2021a. Deep burst super-resolution//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 9209-9214 [DOI: 10.1109/CVPR46437.2021.00909]
- Bhat G, Danelljan M, Yu F, Van Gool L and Timofte R. 2021b. Deep reparametrization of multi-frame super-resolution and denoising//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 2460-2470 [DOI: 10.1109/ICCV48922.2021.00246]
- Brooks T, Mildenhall B, Xue T F, Chen J W, Sharlet D and Barron J T. 2019. Unprocessing images for learned raw denoising//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 11028-11037 [DOI: 10.1109/CVPR.2019.01129]
- Chang K C, Wang R, Lin H J, Liu Y L, Chen C P, Chang Y L and Chen H T. 2020. Learning camera-aware noise models//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 343-358 [DOI: 10.1007/978-3-030-58586-0\_21]
- Chaudhari P, Schirmmacher F, Maier A, Riess C and Köhler T. 2019. Merging-ISP: multi-exposure high dynamic range image signal processing//Proceedings of the 43rd DAGM German Conference on Pattern Recognition. Bonn, Germany: Springer [DOI: 10.1007/978-3-030-92659-5\_21]
- Chen C, Chen Q F, Do M and Koltun V. 2019. Seeing motion in the dark//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 3184-3193 [DOI: 10.1109/ICCV.2019.00328]
- Chen C, Chen Q F, Xu J and Koltun V. 2018. Learning to see in the dark//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 3291-3300 [DOI: 10.1109/CVPR.2018.00347]
- Chen G Y, Chen C F, Guo S, Liang Z T, Wong K Y K and Zhang L. 2021. HDR video reconstruction: a coarse-to-fine network and a real-world benchmark dataset//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 2502-2511 [DOI: 10.1109/ICCV48922.2021.00250]
- Conde M V, McDonagh S, Maggioni M, Leonardis A and Pérez-Pellitero E. 2022. Model-based image signal processors via learnable dictionaries//Proceedings of the 36th AAAI Conference on Artificial Intelligence. Virtual: AAAI: 481-489 [DOI: 10.1609/aaai.v36i1.19926]
- Çoğalan U and Akyüz A O. 2020. Deep joint deinterlacing and denoising for single shot dual-ISO HDR reconstruction. *IEEE Transactions on Image Processing*, 29: 7511-7524 [DOI: 10.1109/tip.2020.3004014]
- Dai P, Yu X, Ma L, Zhang B H, Li J, Li W B, Shen J J and Qi X J. 2022. Video demoiréing with relation-based temporal consistency//Proceedings of 2022 IEEE/CVF Conference on Computer Vision

- and Pattern Recognition. New Orleans, USA: IEEE: 17622-17631 [DOI: 10.1109/CVPR52688.2022.01710]
- Dai T H, Wei L, Cao X L, Liu J Z, Jia X, Leonardis A, Yan Y L and Yuan S X. 2024. Wavelet-based network for high dynamic range imaging. *Computer Vision and Image Understanding*, 238: #103881 [DOI: 10.1016/j.cviu.2023.103881]
- Dong X B, Xu W Y, Miao Z H, Ma L, Zhang C, Yang J W, Jin Z, Teoh A B J and Shen J J. 2022. Abandoning the Bayer-filter to see in the dark//*Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. New Orleans, USA: 17410-17419 [DOI: 10.1109/CVPR52688.2022.01691]
- Du S L, Dang H, Zhao M H and Shi Z H. 2023. Low-light image enhancement and denoising with internal and external priors. *Journal of Image and Graphics*, 28(9): 2844-2855 (都双丽, 党慧, 赵明华, 石争浩). 2023. 结合内外先验知识的低照度图像增强与去噪算法. *中国图象图形学报*, 28(9): 2844-2855 [DOI: 10.11834/jig.220707]
- Dudhane A, Zamir S W, Khan S, Khan F S and Yang M H. 2022. Burst image restoration and enhancement//*Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans, USA: IEEE: 5759-5768 [DOI: 10.1109/CVPR52688.2022.00567]
- Dudhane A, Zamir S W, Khan S, Khan F S and Yang M H. 2023. Burstormer: burst image restoration and enhancement Transformer//*Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver, Canada: IEEE: 5703-5712 [DOI: 10.1109/CVPR52729.2023.00552]
- Fan Z H, Wu X, Meng F Q, Wu Y Q and Zhang F. 2023. OTST: a two-phase framework for joint denoising and remosaicing in RGBW CFA//*Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver, Canada: IEEE: 2832-2841 [DOI: 10.1109/CVPRW59228.2023.00284]
- Feng H S, Wang L Z, Wang Y Z, Fan H Q and Huang H. 2024. Learnability enhancement for low-light raw image denoising: a data perspective. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(1): 370-387 [DOI: 10.1109/TPAMI.2023.3301502]
- Foi A, Trimeche M, Katkovnik V and Egiazarian K. 2008. Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data. *IEEE Transactions on Image Processing*, 17(10): 1737-1754 [DOI: 10.1109/TIP.2008.2001399]
- Froehlich J, Grandinetti S, Eberhardt B, Walter S, Schilling A and Brendel H. 2014. Creating cinematic wide gamut HDR-video for the evaluation of tone mapping operators and HDR-displays//*Proceedings of SPIE 9023, Digital Photography X*. San Francisco, USA: SPIE [DOI: 10.1117/12.2040003]
- Fu Y, Wang Z C, Zhang T and Zhang J. 2023. Low-light raw video denoising with a high-quality realistic motion dataset. *IEEE Transactions on Multimedia*, 25: 8119-8131 [DOI: 10.1109/TMM.2022.3233247]
- Guo S, Liang Z T and Zhang L. 2021. Joint denoising and demosaicking with green channel prior for real-world burst images. *IEEE Transactions on Image Processing*, 30: 6930-6942 [DOI: 10.1109/TIP.2021.3100312]
- Hanji P, Mantiuk R, Eilertsen G, Hajisharif S and Unger J. 2022. Comparison of single image HDR reconstruction methods—the caveats of quality assessment//*Proceedings of 2022 ACM SIGGRAPH Conference Proceedings*. Vancouver, Canadian: ACM: #1 [DOI: 10.1145/3528233.3530729]
- He B, Wang C, Shi B X and Duan L Y. 2020. FHDe<sup>2</sup>Net: full high definition demoiréing network//*Proceedings of the 16th European Conference on European Conference on Computer Vision*. Glasgow, UK: Springer: 713-729 [DOI: 10.1007/978-3-030-58542-6\_43]
- Hu J, Gallo O, Pulli K and Sun X B. 2013. HDR Deghosting: how to deal with saturation?//*Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition*. Portland, USA: IEEE: 1163-1170 [DOI: 10.1109/cvpr.2013.154]
- Huang H F, Yang W H, Hu Y Y, Liu J Y and Duan L Y. 2022. Towards low light enhancement with RAW images. *IEEE Transactions on Image Processing*, 31: 1391-1405 [DOI: 10.1109/TIP.2022.3140610]
- Jiang H Y and Zheng Y Q. 2019. Learning to see moving objects in the dark//*Proceedings of 2019 IEEE/CVF International Conference on Computer Vision*. Seoul, Korea (South): IEEE: 7323-7332 [DOI: 10.1109/ICCV.2019.00742]
- Jiang J J, Cheng H, Li Z Y, Liu X M and Wang Z Y. 2023. Deep learning based video-related super-resolution technique: a survey. *Journal of Image and Graphics*, 28(7): 1927-1964 (江俊君, 程豪, 李震宇, 刘贤明, 王中元). 2023. 深度学习视频超分辨率技术综述. *中国图象图形学报*, 28(7): 1927-1964 [DOI: 10.11834/jig.220130]
- Jiang Y T, Choi I, Jiang J and Gu J W. 2021. HDR video reconstruction with tri-exposure Quad-Bayer sensors [EB/OL]. [2024-01-19]. <https://arxiv.org/pdf/2103.10982.pdf>
- Jin X, Han L H, Li Z, Guo C L, Chai Z and Li C Y. 2023a. DNF: decouple and feedback network for seeing in the dark//*Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Vancouver, Canada: 18135-18144 [DOI: 10.1109/CVPR52729.2023.01739]
- Jin X, Xiao J W, Han L H, Guo C L, Zhang R X, Liu X L and Li C Y. 2023b. Lighting every darkness in two pairs: a calibration-free pipeline for RAW denoising//*Proceedings of 2023 IEEE/CVF International Conference on Computer Vision*. Paris, France: IEEE: 13229-13238 [DOI: 10.1109/ICCV51070.2023.01221]
- Joze H R V, Zharkov I, Powell K, Ringler C, Liang L M, Roulston A, Lutz M and Pradeep V. 2020. ImagePairs: realistic super resolution dataset via beam splitter camera rig//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. Seattle, USA: IEEE: 2190-2200 [DOI: 10.1109/

- CVPRW50498.2020.00267]
- Kronander J, Gustavson S, Bonnet G and Unger J. 2013. Unified HDR reconstruction from raw CFA data//Proceedings of 2013 IEEE International Conference on Computational Photography (ICCP). Cambridge, USA: IEEE: 1-9 [DOI: 10.1109/iccp.2013.6528315]
- Lamba M and Mitra K. 2021. Restoring extremely dark images in real time//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE: 3486-3496 [DOI: 10.1109/CVPR46437.2021.00349]
- Lehouat B, Eboli T, Ponce J and Mairal J. 2022. High dynamic range and super-resolution from raw image bursts. *ACM Transactions on Graphics*, 41(4): #38 [DOI: 10.1145/3528223.3530180]
- Lehouat B, Ponce J and Mairal J. 2021. Lucas-Kanade reloaded: end-to-end super-resolution from raw image bursts//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 2350-2359 [DOI: 10.1109/ICCV48922.2021.00237]
- Li Y W, Zhang K, Liang J Y, Cao J Z, Liu C, Gong R, Zhang Y L, Tang H, Liu Y, Demandolx D and Ranjan R. 2023. LSDIR: a large scale dataset for image restoration//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Vancouver, Canada: IEEE: 1775-1787 [DOI: 10.1109/CVPRW59228.2023.00178]
- Liang Z T, Guo S, Gu H, Zhang H Q and Zhang L. 2020. A decoupled learning scheme for real-world burst denoising from raw images//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 150-166 [DOI: 10.1007/978-3-030-58595-2\_10]
- Liu J M, Wu C H, Wang Y Z, Xu Q, Zhou Y Q, Huang H B, Wang C, Cai S F, Ding Y F, Fan H Q and Wang J. 2019. Learning raw image denoising with Bayer pattern unification and Bayer preserving augmentation//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Long Beach, USA: IEEE: 2070-2077 [DOI: 10.1109/CVPRW.2019.00259]
- Liu L, Jia X, Liu J Z and Tian Q. 2020. Joint demosaicing and denoising with self guidance//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 2237-2246 [DOI: 10.1109/CVPR42600.2020.00231]
- Liu S Z, Zhang X D, Sun L C, Liang Z T, Zeng H and Zhang L. 2023. Joint HDR Denoising and fusion: a real-world mobile HDR image dataset//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, Canada: IEEE: 13966-13975 [DOI: 10.1109/CVPR52729.2023.01342]
- Liu X H, Shi K D, Wang Z and Chen J. 2021. Exploit camera raw data for video super-resolution via hidden Markov model inference. *IEEE Transactions on Image Processing*, 30: 2127-2140 [DOI: 10.1109/TIP.2021.3049974]
- Ma L, Ma T Y and Liu R S. 2022. A comprehensive review of low-light image enhancement algorithms. *Journal of Image and Graphics*, 27(5): 1392-1409 (马龙, 马腾宇, 刘日升. 2022. 低光照图像增强算法综述. 中国图象图形学报, 27(5): 1392-1409) [DOI: 10.11834/jig.210852]
- Maggioni M, Huang Y B, Li C, Xiao S, Fu Z Q and Song F L. 2021. Efficient multi-stage video denoising with recurrent spatio-temporal fusion//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 3465-3474 [DOI: 10.1109/CVPR46437.2021.00347]
- Mehta N, Dudhane A, Murala S, Zamir S W, Khan S and Khan F S. 2023. Gated multi-resolution transfer network for burst restoration and enhancement//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 22201-22210 [DOI: 10.1109/CVPR52729.2023.02126]
- Plotz T, Roth S. 2017. Benchmarking denoising algorithms with real photographs//Proceedings of 2017 IEEE conference on computer vision and pattern recognition. Honolulu, USA: IEEE: 1586-1595 [DOI: 10.1109/cvpr.2017.294]
- Qian G C, Wang Y H, Gu J J, Dong C, Heidrich W, Ghanem B and Ren J S. 2022. Rethinking learning-based demosaicing, denoising, and super-resolution pipeline//Proceedings of 2022 IEEE International Conference on Computational Photography (ICCP). Pasadena, USA: IEEE: 1-12 [DOI: 10.1109/ICCP54855.2022.9887682]
- Sun Y J, Yu Y Z and Wang W P. 2018. Moiré photo restoration using multiresolution convolutional neural networks. *IEEE Transactions on Image Processing*, 27(8): 4160-4172 [DOI: 10.1109/TIP.2018.2834737]
- Tang C Z, Yang Y Q, Zeng B, Tan P and Liu S C. 2022. Learning to zoom inside camera imaging pipeline//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 17531-17540 [DOI: 10.1109/CVPR52688.2022.01703]
- Tang L F, Huang H, Zhang Y F and Li F. 2022. Spatial aware channel attention guided high dynamic image reconstruction. *Journal of Image and Graphics*, 27(12): 3581-3595 (唐凌峰, 黄欢, 张亚飞, 李凡. 2022. 空间感知通道注意力引导的高动态图像重建. 中国图象图形学报, 27(12): 3581-3595 [DOI: 10.11834/jig.211039])
- Tseng E, Yu F L, Yang Y T, Mannan F, Arnaud K S T, Nowrouzezahrai D, Lalonde J F and Heide F. 2019. Hyperparameter optimization in black-box image processing using differentiable proxies. *ACM Transactions on Graphics*, 38(4): #27 [DOI: 10.1145/3306346.3322996]
- Unger J, Hajisharif S and Kronander J. 2016. Unified reconstruction of RAW HDR video data//Dufaux F, Le Callet P, Mantiuk R K, Mrak M, eds. *High Dynamic Range Video: From Acquisition to Display and Applications*. Amsterdam: Academic Press: 63-83 [DOI: 10.1016/b978-0-08-100412-8.00002-4]
- Wang R X, Xu X G, Fu C W, Lu J B, Yu B and Jia J Y. 2021. Seeing

- dynamic scene in the dark: a high-quality video dataset with mechatronic alignment//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 9680-9689 [DOI: 10.1109/ICCV48922.2021.00956]
- Wang X T, Chan K C K, Yu K, Dong C and Loy C C. 2019. EDVR: video restoration with enhanced deformable convolutional networks//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Long Beach, USA: IEEE: 1954-1963 [DOI: 10.1109/cvprw.2019.00247]
- Wang Y F, Yu Y, Yang W H, Guo L Q, Chau L P, Kot A C and Wen B H. 2023. Raw image reconstruction with learned compact metadata//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 18206-18215 [DOI: 10.1109/CVPR52729.2023.01746]
- Wang Y Z, Huang H B, Xu Q, Liu J M, Liu Y Q and Wang J. 2020. Practical deep raw image denoising on mobile devices//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 1-16 [DOI: 10.1007/978-3-030-58539-6\_1]
- Wei K X, Fu Y, Yang J L and Huang H. 2020. A physics-based noise formation model for extreme low-light raw denoising//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual Conference: IEEE: 2758-2767 [DOI: 10.1109/cvpr42600.2020.00283]
- Wei K X, Fu Y, Zheng Y Q and Yang J L. 2022. Physics-based noise modeling for extreme low-light photography. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(11): 8520-8537 [DOI: 10.1109/TPAMI.2021.3103114]
- Wei L Y and Sun B Y. 2022. The review of demosaicing methods for Bayer color filter array image. Journal of Image and Graphics, 27(9): 2683-2696 (魏凌云, 孙帮勇. 2022. Bayer阵列图像去马赛克算法综述. 中国图象图形学报, 27(9): 2683-2696)[DOI: 10.11834/jig.211059]
- Wei P X, Sun Y J, Guo X B, Liu C, Li G B, Chen J, Ji X Y and Lin L. 2023. Towards real-world burst image super-resolution: benchmark and method//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 13187-13196 [DOI: 10.1109/ICCV51070.2023.01217]
- Xing Y Z, Qian Z A and Chen Q F. 2021. Invertible image signal processing//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 6283-6292 [DOI: 10.1109/CVPR46437.2021.00622]
- Xu K, Yang X, Yin B C and Lau R W H. 2020a. Learning to restore low-light images via decomposition-and-enhancement//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 2278-2287 [DOI: 10.1109/CVPR42600.2020.00235]
- Xu S N, Song B B, Chen X Y and Zhou J T. 2023b. Image demoiréing in RAW and sRGB domains[EB/OL]. [2024-01-19]. <https://arxiv.org/pdf/2312.09063.pdf>
- Xu X, Ye Y F and Li X. 2020b. Joint demosaicing and super-resolution (JDSR): network design and perceptual optimization. IEEE Transactions on Computational Imaging, 6: 968-980[DOI: 10.1109/TCI.2020.2999819]
- Xu X Y, Ma Y R and Sun W X. 2019. Towards real scene super-resolution with raw images//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 1723-1731 [DOI: 10.1109/CVPR.2019.00182]
- Yuan S X, Timofte R, Leonardis A, Slabaugh G, Luo X T, Zhang J T, Qu Y Y, Hong M, Xie Y, Li C H, Xu D J, Chu Y H, Sun Q Y, Liu S, Zong Z Y, Nan N, Li C H, Kim S, Nam H, Kim J, Jeong J, Cheon M, Yoon S J, Kang B, Lee J, Zheng B L, Liu X H, Dai L H, Chen J, Cheng X, Fu Z Y, Yang J, Lee C, Vien A G, Park H, Nathan S, Beham M P, Roomi S M M, Lemarchand F, Pelcat M, Nogues E, Puthussery D, Hrishikesh P S, Jiji C V, Sinha A and Zhao X. 2020. NTIRE 2020 challenge on image demoiréing: methods and results//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Seattle, USA: IEEE: 1882-1893 [DOI: 10.1109/CVPRW50498.2020.00238]
- Yuan S X, Timofte R, Slabaugh G, Leonardis A, Zheng B L, Ye X, Tian X, Chen Y W, Cheng X, Fu Z Y, Yang J, Hong M, Lin W Y, Yang W J, Qu Y Y, Shin H K, Kim J Y, Ko S J, Dong H, Guo Y, Wang J, Ding X, Han Z Y, Das S D, Purohit K, Kandula P, Suin M and Rajagopalan A N. 2019. AIM 2019 challenge on image demoiréing: methods and results//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). Seoul, Korea (South): IEEE: 3534-3545 [DOI: 10.1109/ICCVW.2019.00438]
- Yue H J, Cao C, Liao L, Chu R H and Yang J Y. 2020. Supervised raw video denoising with a benchmark dataset on dynamic scenes//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 2298-2307 [DOI: 10.1109/CVPR42600.2020.00237]
- Yue H J, Cheng Y J, Liu X and Yang J Y. 2023a. Recaptured Raw Screen Image and Video Demoiréing via Channel and Spatial Modulations//Proceedings of the 37th Conference on Neural Information Processing Systems. New Orleans, USA: NeurIPS: 1-12
- Yue H J, Cheng Y J, Mao Y, Cao C and Yang J Y. 2023b. Recaptured screen image demoiréing in raw domain. IEEE Transactions on Multimedia, 25: 5589-5600[DOI: 10.1109/TMM.2022.3198333]
- Yue H J, Mao Y, Liang L P, Xu H T, Hou C P and Yang J Y. 2021. Recaptured screen image demoiréing. IEEE Transactions on Circuits and Systems for Video Technology, 31(1): 49-60 [DOI: 10.1109/TCSVT.2020.2969984]
- Yue H J, Peng Y B, Yu B T, Yin X W, Zhou Z Y and Yang J Y. 2023c. HDR video reconstruction with a large dynamic dataset in raw and sRGB domains [EB/OL]. [2024-01-19]. <https://arxiv.org/pdf/2304.04773.pdf>

- Yue H J, Zhang Z M and Yang J Y. 2022. Real-RawVSR: real-world raw video super-resolution with a benchmark dataset//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv, Israel: Springer: 608-624 [DOI: 10.1007/978-3-031-20068-7\_35]
- Zamir S W, Arora A, Khan S, Hayat M, Khan F S, Yang M H and Shao L. 2020. CycleISP: real image restoration via improved data synthesis//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 2693-2702 [DOI: 10.1109/CVPR42600.2020.00277]
- Zamir S W, Arora A, Khan S, Khan F S and Shao L. 2021. Learning digital camera pipeline for extreme low-light imaging. *Neurocomputing*, 452: 37-47 [DOI: 10.1016/j.neucom.2021.04.076]
- Zhang F, Xu B, Li Z Q, Liu X R, Lu Q B, Gao C X and Sang N. 2023. Towards General Low-Light Raw Noise Synthesis and Modeling//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 10820-10830 [DOI: 10.1109/ICCV51070.2023.00993]
- Zhang X E, Chen Q F, Ng R and Koltun V. 2019. Zoom to learn, learn to zoom//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 3762-3770 [DOI: 10.1109/CVPR.2019.00388]
- Zhang Y, Li D S, Law K L, Wang X G, Qin H W and Li H S. 2022. IDR: self-supervised image denoising via iterative data refinement//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 2088-2097 [DOI: 10.1109/CVPR52688.2022.00214]
- Zhang Y, Qin H W, Wang X G, Li H S. 2021. Rethinking noise synthesis and modeling in raw denoising//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision, Virtual: IEEE: 4593-4601 [DOI: 10.1109/iccv48922.2021.00455]
- Zou Y H and Fu Y. 2022. Estimating fine-grained noise model via contrastive learning//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 12672-12681 [DOI: 10.1109/CVPR52688.2022.01235]
- Zou Y H, Yan C G and Fu Y. 2023. RawHDR: high dynamic range image reconstruction from a single raw image//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 12300-12310 [DOI: 10.1109/ICCV51070.2023.01133]

### 作者简介

岳焕景,女,副教授,主要研究方向为底层视觉处理。

E-mail: huanjing.yue@tju.edu.cn

杨敬钰,通信作者,男,教授,主要研究方向为计算成像与重建。E-mail: yjy@tju.edu.cn

杨文瀚,男,副研究员,主要研究方向为底层视觉处理。

E-mail: yangwh@pcl.ac.cn

李重仪,男,教授,主要研究方向为底层视觉处理。

E-mail: lichongyi@nankai.edu.cn

杨铀,男,教授,主要研究方向为视觉感知与计算。

E-mail: yangyou@hust.edu.cn

刘文予,男,教授,主要研究方向为计算机视觉。

E-mail: liuwuy@hust.edu.cn