arXiv:2305.05899v1 [cs.CV] 10 May 2023

# 基于先验量化的移动图像复原

Shiqi Chen、Jinwen Zhou、Menghao Li、Yueting Chen、Tingting Jiang

**摘要—在数字图像中，光学像差的表现是一种多变量退化，场景的光谱、镜头缺陷和视野共同影响最终结果。除了在硬件层面消除光学像差之外，利用各种先验信息的后处理系统对于校正也至关重要。然而，由于不同先验之间存在内容差异，用于对齐这些因素的处理流程效率有限，复原效果也未得到优化。为此，我们提出了一种先验量化模型，用于校正图像处理系统中的光学像差。为了融合这些信息，我们将各种先验编码到潜在空间中，并通过可学习码本对其进行量化。量化后，将先验编码与图像复原分支融合，从而实现有针对性的光学像差校正。全面的实验结果表明，所提出方法具有良好的灵活性，并验证了其针对特定相机实现目标复原的潜力。此外，我们的模型有望分析各种先验与设备光学像差之间的相关性，这将有助于软硬件协同设计。**

索引词—图像处理，神经网络，光学像差校正，优先级

## I. 引言

ANY数字成像系统存在光学像差，因此要实现精确测量，必须校正像差 [1]。遗憾的是，图像中的光学像差受多种因素影响，其形式化表示如下：

$$J_e = \int \mathcal{C}_e(\lambda) \cdot [I_e(h, w, \lambda) * L_e(h, w, \lambda)] d\lambda + N_e(h, w), \quad (1)$$

这里（$h$，$w$）表示像素的坐标，$\lambda$ 表示波长，$\mathcal{C}_e$ 和 $I_e$ 分别表示光谱响应和点扩散函数（PSF），$L_e$、$J_e$和 $N_e$ 分别表示潜在清晰图像、观测图像和测量噪声。需要注意的是，下标 $e$ 表示测量值，代表传感器接收到的能量。因此，能量色散、光学系统的视场（FoV）线索、光谱特性以及传感器噪声共同构成了数字图像中的光学像差表达。换句话说，这些因素均可作为先验，以促进校正。

尽管已有许多光学像差校正算法，但要实现广泛应用仍面临一些挑战。一个问题是，现有方法通常具有有限的应用范围，例如，反卷积是在处理空间变化核时效率低下，而且深度学习方法是根据数据 [2] 针对特定设备训练的。另一个挑战是复原质量，由于缺乏足够的信息，复原质量通常不尽如人意。这些不灵活方法的固有缺陷在于，它们无法挖掘数字像素与其他光学先验之间的相互作用 [3]。

如上所述，光学像差的表现与多个因素相关。如何设计一种通用方法，利用这些先验进行定向校正，是一个值得探讨的问题。同时，这一基线有助于分析不同先验与光学像差校正之间的相关性，从而指导高端成像系统中硬件配置与后处理流程的协同设计。

在本篇论文中，我们提出了一种先验量化模型，用于校正受多种因素影响的光学退化，其中，与每种因素相对应的不同先验被输入该模型。然而，由于多模态先验之间存在内容差异，无法将其直接整合到模型中以实现高效的后处理。为此，我们将各种先验编码到高维潜在空间中，并利用可学习码本对其进行表征。学习到的编码降低了多模态先验表示的维度，并对辅助信息的全局相互关系进行建模。随后，模型通过将量化先验表示与不同尺度的特征进行融合，将其整合到图像复原分支中。最后，我们在多个尺度上对复原过程进行监督，以确保用于融合的跨尺度信息准确且有效。与黑箱模型不同，该模型能够根据用户的需求调整输入的像素级先验，旨在实现有针对性的改进。此外，学习到的码本在不同先验与复原之间架起了桥梁。因此，所提出的模型具有分析先验利用情况的潜力，这对于成像系统设计具有重要意义。

## II. 所提出的方法

如上所述，光学像差的表现受多种因素影响。因此，模型必须具备感知复杂组合并将其转化为简单形式的能力。直接将图像信号与各个先验结合效率低下，模型会在辅助信息上耗费大量计算开销。因此，我们将这些先验编码到高维空间中，并用可学习码本表示。由于邻域内的像差相对恒定，图像 $x \in \mathcal{R}^{H \times W \times 3}$ 的光学先验可以表示为码本 $z_q \in \mathcal{R}^{h \times w \times d_z}$ 的空间集合，其中 $d_z$ 是编码的维度。为了有效地学习这样的

本研究得到了民用航空航天预研项目 D040107、国家自然科学基金 61424080211 以及浙江实验室重点研究项目 2021MH0AC01 的部分资助。（通讯作者：Tingting Jiang。）Shi

qi Chen, Jinwen Zh你，an d Yueting Chen 是 w ith the Colle ge o f Optical 科学与工程学院，浙江大学，中国杭州 310000（电子邮箱：chenshiqi@zju.edu.cn）。

Tingting Jiang 任职于中国杭州 311100 的浙江实验室智能传感系统研究中心（电子邮箱：eager- jtt@zhejianglab.com）。

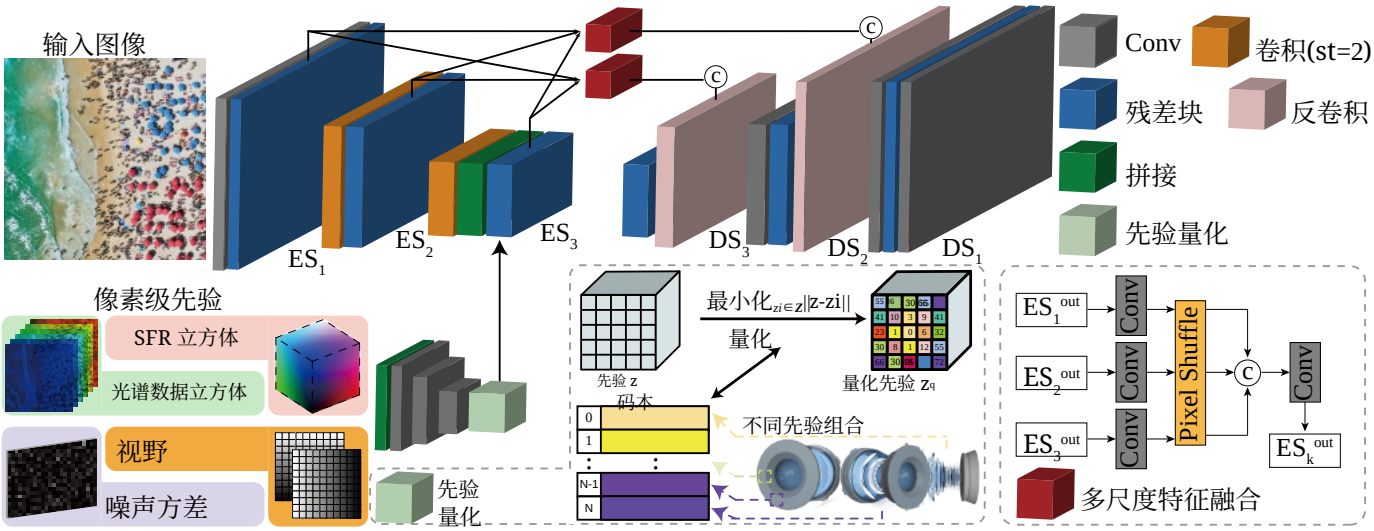


图1。Thede 所提出的先验量化模型的尾部。各层配置以不同颜色的方块表示。 .

码本，我们提出利用卷积的空间提取能力，并将其与神经离散表示学习 [4] 相结合。首先，我们使用卷积神经网络将先验编码到具有相同空间分辨率（$h \times w$）的潜在空间中。然后，我们将编码后的特征嵌入空间编码 $\hat{z}_{ij} \in \mathcal{R}^{d_z}$，并将每个编码量化为离散码本 $\mathcal{Z} = \{z_k\}_{k=1}^{K} \subset \mathcal{R}^{n_z}$ 中与之最近的条目：

$$z_q = \left( \underset{z_k \in \mathcal{Z}}{\arg\min} ||\hat{z}_{ij} - z_k|| \right) \in \mathcal{R}^{h \times w \times d_z}, \tag{2}$$

第三，我们将量化先验表示与复原分支中的特征进行拼接。随后，残差块主动融合图像特征，其中量化后的先验为重建分支提供空间/通道重要性的线索。

在大多数复原模型中，应用由粗到细架构时，跳跃连接仅在单一尺度上进行处理。受多尺度特征之间密集连接的启发，我们设计了一个多尺度特征融合（MFF）模块，用于将量化表示与其他尺度的特征进行融合 [5]。如图 1 右下方所示，该模块接收不同编码尺度（$ES_i$）的输出作为输入。通过卷积调整每个 $ES_i^{out}$ 的通道数后，我们使用像素重排将编码信息转换为相同的空间分辨率，然后进行融合 [6]。MFF 的输出被传递至相应的解码尺度。通过这种方式，每个尺度都能够感知其他尺度的编码信息，尤其是经过量化先验筛选的最低尺度特征，从而提高复原质量。

这里我们讨论模型的损失函数。由于式 2 中的量化操作在反向传播过程中不可微，若直接采用端到端监督训练整个网络，先验的卷积神经网络编码器将无法获得梯度来优化其参数。幸运的是，码本对齐策略使从解码器传播来的梯度能够更新这些卷积神经网络编码器。具体而言，我们使编码后的先验尽可能接近可学习码本的向量。这样一来，量化过程就像一个梯度估计器，即使训练过程中梯度被截断，也能让卷积神经网络编码器估计反向传播并更新参数。因此，我们采用码本对齐损失来实现编码器的优化 [7]:

$$\mathcal{L}_{align} = ||sg[\hat{z}_{ij}] - z_q||_2^2 + ||\hat{z}_{ij} - sg[z_q]||_2^2, \tag{3}$$

其中，$\hat{z}_{ij} = \{E_1(p_1)\ E_2(p_2) \ldots, E_n(p_n)\}$，其中 $p_i$ 和 $E_i$ 分别为 $i^{th}$ 先验和编码器。$\{\cdot\}$ 是用于拼接编码特征的操作。$sg[\cdot]$ 表示停止梯度操作，以确保损失函数仅引导编码后的先验和码本向量相互靠近。

由于我们的网络采用由粗到细的渐进式细化方式，因此我们在不同尺度上监督图像重建。在图像内容监督方面，我们发现，对于光学像差校正，L1损失在定量指标上表现更好。然而，由于内容损失仅衡量像素级差异，模型能够很好地恢复低频信息（例如亮度和颜色，等等。），却难以恢复场景纹理。为避免像素级监督的限制，我们采用傅里叶域监督作为辅助损失 [5]。与梯度约束（例如，总变分）或预训练模型上的特征相似度（例如，感知损失）相比，这一项能够提供有关不同空间频率的更多信息。损失 $\mathcal{L}_{content}$ 的公式如下：

$$\mathcal{L}_{content} = \frac{1}{t_k} \sum_{k=1}^{K} ||\hat{I}_k - I_k||_1 + \lambda ||\mathcal{F}(\hat{I}_k) - \mathcal{F}(I_k)||_1, \tag{4}$$

这里，第 k 个尺度的输出为 $I_k$，其对应的真实值为 $I_k$，其中 $K$ 是尺度的数量。每个尺度的内容损失是所有元素的平均值，元素总数记为 $t_k$。$\mathcal{F}$ 是对不同尺度执行的快速傅里叶变换（FFT）操作，$\lambda$ 通过实验设为 0.1。总体损失函数是 $\mathcal{L}_{content}$ 和 $\mathcal{L}_{align}$ 之和。

表 I 所提模型与其他竞争方法在合成数据和真实数据上的性能。百分比表示相较于最佳模型的相对改进（峰值信噪比先转换为均方根误差，其他指标不变）。

| 方法 \ 测试集 | 合成评估 | | | | | | | | 真实评估 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 峰值信噪比↑ | | 结构相似性↑ | | VIF↑ | | 学习感知图像块相似度↓ | | 盲/无参考图像空间质量评估↓ | | 自然图像质量评估器↓ | |
| SRN | 29.79 | (56.6%) | 0.9247 | (29.1%) | 0.8221 | (11.3%) | 3.428 | (48.2%) | 54.49 | (16.1%) | 5.686 | (21.3%) |
| IRCNN | 30.58 | (48.0%) | 0.9289 | (24.5%) | 0.8314 | (10.0%) | 3.151 | (43.7%) | 53.47 | (14.5%) | 5.572 | (19.7%) |
| 自去模糊 | 32.23 | (24.0%) | 0.9353 | (17.5%) | 0.8544 | (7.05%) | 2.766 | (35.8%) | 50.28 | (9.03%) | 5.323 | (16.0%) |
| GLRA | 32.47 | (19.6%) | 0.9347 | (18.2%) | 0.8662 | (5.60%) | 2.457 | (27.8%) | 49.74 | (8.04%) | 5.476 | (18.3%) |
| FDN | 33.27 | (3.39%) | 0.9402 | (12.2%) | 0.8943 | (2.28%) | 1.944 | (8.69%) | 47.42 | (3.54%) | 5.038 | (11.2%) |
| Deep Wiener | 32.02 | (27.6%) | 0.9463 | (5.7%) | 0.9007 | (1.55%) | 1.928 | (7.94%) | 46.52 | (16.8%) | 4.943 | (9.44%) |
| Ours | 33.42 | (0.0%) | 0.9517 | (0.0%) | 0.9147 | (0.0%) | 1.775 | (0.0%) | 45.74 | (0.0%) | 4.476 | (0.0%) |

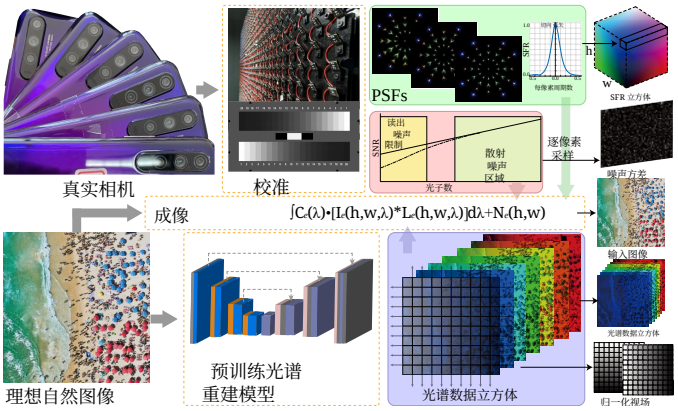


图 2. 训练数据的合成流程以及 priors.

### A. 训练数据和先验的合成流程

由于光学像差与多种因素密切相关，我们构建了一个综合数据集用于校正。训练数据和先验的详细合成流程如图 2 所示。首先，我们对多种手机相机的点扩散函数和噪声因素进行标定，其中点扩散函数用于模拟退化并计算空间频率响应先验。其次，我们使用预训练的 MST++ [8] 将 RGB 图像转换为高光谱数据，以获取光谱先验。第三，视场先验由归一化到 $[-1, 1]$ [9]、[10] 的 $(h,\ w)$ 像素坐标组成。最后，将多光谱数据与光学像差结合，处理过程与公式 1 相同。

## III. 实验结果

### A. 数据集、指标与训练设置

如上图所示，我们使用 DIV8K 中的自然图像合成训练图像对和先验 [11]。训练数据集包含 800 对图像。图像分辨率会缩放至 $3000 \times 4000$，以匹配真实相机。在真实场景评估中，我们使用华为荣耀20 和 iPhone 12 等多款移动终端拍摄了大量照片。对于评估指标，PSNR、SSIM、VIF[12], 和 LPIPS [13] 用于在有参考图像的情况下评估模型。BRISQUE [14], 和 NIQE [15] 用于评估所拍摄照片的复原效果。训练时，我们将整幅图像裁剪为 $256 \times 256$ 像素，并组成每批 8 张图像的小批次。为确保公平比较，我们将先验与图像拼接后输入竞争模型。我们使用 Adam 在 $\beta_1 = 0.5$ 和 $\beta_2 = 0.999$上优化模型。初始学习率为 $10^{-4}$，之后每 50 个周期减半。

### B. 与最先进方法的定量评估 $s$

我们将提出的先验量化模型与针对光学像差校正设计的其他算法进行比较。所有这些方法均使用相同的数据集重新训练，直至收敛。在评估空间变化像差校正效果时，我们在合成数据集上评估定量指标。表 I 报告了各种方法在合成数据集上的性能。盲方法（SRN[16],IRCNN[17], SelfDeblur[18], GLRA[19]）能够成功处理单个相机的光学退化。然而，手机相机种类繁多，光学像差各不相同，仅凭图像特征，盲方法难以获得精确的退化线索。非盲方法（FDN[20], Deep Wiener[21]）将点扩散函数输入预训练模型，以适应特定相机，其思路与我们的方法类似。然而，由于点扩散函数是退化的高维表示，因此仅选取少数具有代表性的点扩散函数作为模型输入（ [21]中仅使用 5 个点扩散函数），这会忽略整个视场内点扩散函数的空间关系。否则，复杂度将极高。与这种方法不同，我们使用重排后的空间频率响应立方体表示每个视场的色散，这是一种维度更低的退化表示。因此，我们的模型在计算开销相对较低的情况下，仍能获得像素级引导，从而实现更好的复原。

### C. 真实复原对比

此外，我们使用真实相机拍摄的退化照片测试了这些算法，结果如图3所示。由于 Deep Wiener 获取的先验与像素邻域并不对应，因此模型需要借助注意力机制来确定输入图像的具体退化情况。这些额外开销增加了处理负担，导致无法实现高效复原。与其他算法和内置图像信号处理器相比，我们的模型能够高效整合各种先验，并通过量化码本向量对其进行表征。因此，我们的方法降低了多任务后处理的复杂度，并全面提升了图像质量。我们的非盲模型还有一个优点：它可以设计成泛化能力更强的后处理系统。

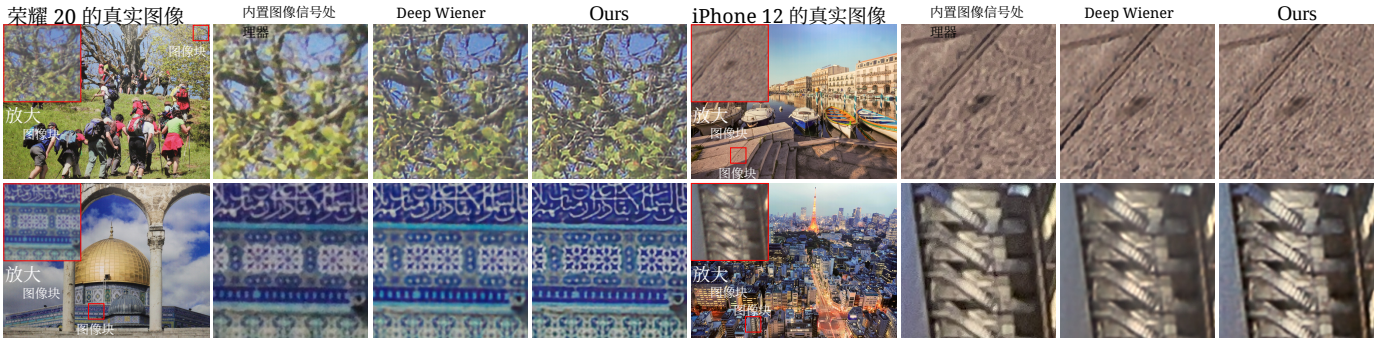


图3. 真实图像复原 c 比较，其中该位置以红色突出显示。见 m 更多实验结果见补充材料 1。

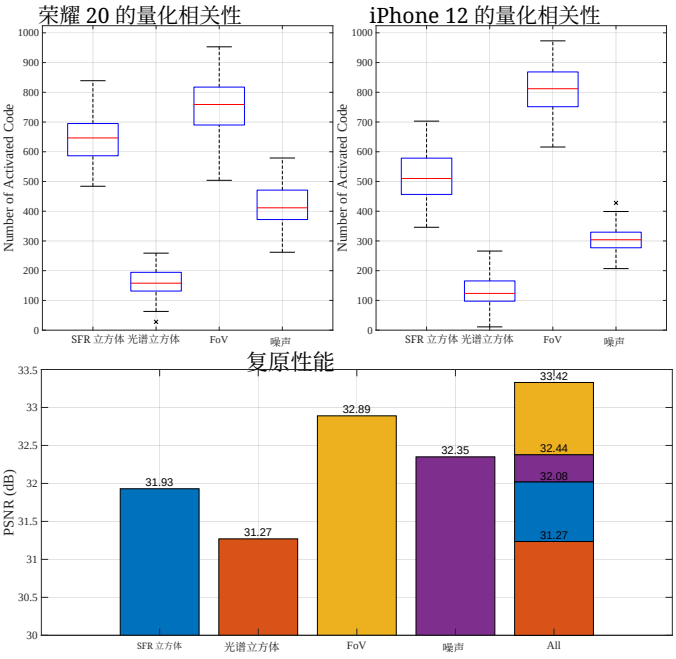


图4. 不同移动相机中先验与光学像差校正之间的相关性。

在实验中，我们使用在仿真数据集上训练的模型，对各种手机相机拍摄的真实图像进行后处理。图3左侧和右侧的结果表明，在合成数据上预训练的模型在特定设备（荣耀 20 Pro 和 iPhone 12 ）上具有出色的泛化能力，无需微调。因此，即插即用特性表明，所学习的模型有望以灵活的模型取代相机专用 ISP 系统。关于详细对比，请读者参阅补充材料，其中提供了更多不同制造样品的复原结果，以及更多有关系统成像质量的定量结果。

### *D.* 先验相关性分析

如上所述，在固定距离成像时，光学像差与多个因素存在明确的相关性，而这些辅助信息有助于图像复原。然而，在实时成像中获取如此多的辅助信息，代价高昂且难以承受。我们使用学习得到的模型以分析引入各项先验对复原的影响。评估某一特定先验时，我们将除该先验之外的所有辅助项置零，并使用不同手机相机拍摄的原始照片测试训练好的模型，统计推理过程中激活的编码数量。激活的码本数量越多，表明该先验与光学像差的相关性越高。评估结果如图4上方所示。我们发现，空间频率响应和视场先验在推理过程中会激活更多编码，这意味着它们与像差高度相关，并在校正中发挥关键作用。与 iPhone 12 相比，荣耀 20 Pro 的复原在评估空间频率响应和噪声先验时会激活更多编码，这归因于其光学元件存在不均匀的光学像差，以及传感器的信噪比较低。该实验还表明，激活数量可反映不同先验的使用情况，为成像系统评估提出了一个全新的问题。此外，我们在合成测试集上评估了复原指标与各项先验之间的相关性（如图4下方所示）。不同于激活编码的数量，我们发现噪声对复原性能的影响更为显著，这可能是因为峰值信噪比是在像素尺度上计算绝对误差。另一方面，这一现象表明，我们可以将噪声与光学像差分离，从而在复原指标和计算效率之间取得平衡。

## IV. 结论

总之，我们开发了一种深度学习模型，以利用多种先验进行光学像差校正。尽管这些先验的内容各不相同，所提出的量化策略仍能有效融合这些因素，并引导光学像差校正。全面的实验表明，整合所有相关信息确实有助于光学像差复原；而且，所提出的模型无需微调即可泛化到不同的移动设备。此外，学习得到的模型量化了不同先验之间的相关性及其在校正过程中的重要性。因此，在获得关键线索后，后处理系统中的像差校正便能更加高效。未来，我们将致力于将所提出的模型应用于不同的成像设备，以实现高效校正并提升泛化能力。

## 参考文献

[1] T. Yue、J. Suo、J. Wang、X. Cao 和 Q. Dai，《通过探索几何与视觉先验进行盲光学像差校正》，IEEE计算机视觉与模式识别会议论文集，第 1684–1692 页，2015 年。[2] T. Eboli、J.-M. Morel 和 G. Facciolo，《快速两步盲光学像差校正》，欧洲计算机视觉会议，第 693–708 页，2022 年。[3] S. Chen、T. Lin、H. Feng、Z. Xu、Q. Li 和 Y. Chen，《面向大规模生产移动终端的计算光学》，IEEE 模式分析与机器智能汇刊，第 1–16 页，2022 年。[4] A. Van Den Oord、O. Vinyals 等，《神经离散表示学习》，神经信息处理系统进展，第 30 卷，2017 年。[5] S.-J. Cho、S.-W. Ji、J.-P. Hong、S.-W. Jung 和 S.-J. Ko，《重新思考单幅图像去模糊中的由粗到细方法》，IEEE/CVF国际计算机视觉会议论文集，第 4641–4650 页，2021 年。[6] W. Shi、J. Caballero、F. Husz´ar、J. Totz、A. P. Aitken、R. Bishop、D. Rueckert 和 Z. Wang，《使用高效亚像素卷积神经网络实时进行单幅图像与视频超分辨率》，IEEE计算机视觉与模式识别会议论文集，第 1874–1883 页，2016 年。[7] J. Duan、L. Chen、S. Tran、J. Yang、Y. Xu、B. Zeng 和 T. Chilimbi，《使用表示码本进行多模态对齐》，IEEE/CVF计算机视觉与模式识别会议论文集，第 15 651–15 660 页，2022 年。[8] Y. Cai、J. Lin、Z. Lin、H. Wang、Y. Zhang、H. Pfister、R. Timofte 和 L. Van Gool，《Mst++：用于高效光谱重建的多阶段光谱变换器》，IEEE/CVF计算机视觉与模式识别会议论文集，第 745–755 页，2022 年。[9] S. Chen、H. Feng、K. Gao、Z. Xu 和 Y. Chen，《基于退化框架的超高质量计算成像》，IEEE/CVF国际计算机视觉会议论文集，第 2632–2641 页，2021 年。[10] Q. Sun、C. Wang、Q. Fu、X. Dun 和 W. Heidrich，《通过可微光线追踪进行端到端复杂镜头设计》，ACM图形学汇刊，第 40 卷，第 4 期，2021 年 7 月。 [在线]。网址：https://doi.org/10 1145 3450626 3459674 。 / 。S. Gu、A. Lugmayr、M. Danelljan、M. Fritsche、J. Lamour 和 R. Timo-fte，《DIV8K：多样化的8K分辨率图像数据集》，载于 2019 IEEE/CVF国际计算机视觉会议研讨会（ICCVW）。IEEE，2019 年，第 3512–3516 页。[12] H. R. Sheikh 和 A. C. Bovik，《图像信息与视觉质量》，IEEE图像处理汇刊，第 15 卷，2006 年，第 2 期，第 430–444 页。[13] R. Zhang、P. Isola、A. A. Efros、E. Shechtman 和 O. Wang，《深层特征作为感知度量的惊人效力》，载于 IEEE计算机视觉与模式识别会议论文集，2018 年，第 586–595 页。[14] A. Mittal、A. K. Moorthy 和 A. C. Bovik，《空间域中的无参考图像质量评估》，IEEE图像处理汇刊，第 21 卷，第 12 期，第 4695–4708 页，2012 年。[15] A. Mittal、R. Soundararajan 和 A. C. Bovik，《打造“完全盲”的图像质量分析器》，IEEE信号处理快报，第 20 卷，第 3 期，第 209–212 页，2012 年。[16] X. Tao、H. Gao、X. Shen、J. Wang 和 J. Jia，《用于深度图像去模糊的尺度循环网络》，IEEE计算机视觉与模式识别会议论文集，第 8174–8182 页，2018 年。[17] K. Zhang、W. Zuo、S. Gu 和 L. Zhang，《学习用于图像复原的深度CNN去噪器先验》，IEEE计算机视觉与模式识别会议论文集，第 3929–3938 页，2017 年。[18] D. Ren、K. Zhang、Q. Wang、Q. Hu 和 W. Zuo，《利用深度先验进行神经盲反卷积》，IEEE/CVF计算机视觉与模式识别会议论文集，第 3341–3350 页，2020 年。[19] W. Ren、J. Zhang、L. Ma、J. Pan、X. Cao、W. Zuo、W. Liu 和 M.-H. Yang，《通过广义低秩近似进行深度非盲反卷积》，神经信息处理系统进展，第 31 卷，2018 年。[20] J. Kruse、C. Rother 和 U. Schmidt，《学习突破高效基于FFT的图像反卷积的极限》，IEEE国际计算机视觉会议论文集，第 4586–4594 页，2017 年。[21] T. Lin、S. Chen、H. Feng、Z. Xu、Q. Li 和 Y. Chen，《通过频率自适应与微调策略进行非盲光学退化校正》，《光学快报》，第 30 卷，第 13 期，第 23 485–23 498 页，2022 年 6 月。 [在线]。网址：https://opg.optica.org/oe/abstract.cfm?URI=oe-30-13-23485