Neural Grey：基于日常相机旋转的实用渐晕校准

Yueze Wang 和 Spencer LaVere Smith

arXiv:2611.28934v1 [eess.IV] 21 Nov 2026

摘要——**Neural Grey 通过拍摄静止的浅色表面，从四张滚转角度各异的图像中估计可重复使用的相机–镜头阴影图。旋转相机会使墙面照明在传感器上移动，而镜头阴影则保持不变。我们利用物理阴影包络和紧凑的有界神经残差，对这两个场以及微小的位姿和曝光误差进行联合拟合。然而，仅凭旋转无法区分径向镜头衰减与墙面径向照明。焦距条件径向选择器可消除此歧义：它根据焦距和传感器尺寸选择宽范围的径向轮廓，同时保留拟合得到的空间残差。组件研究和敏感性研究揭示了场表示、径向选择和数值设置各自的作用。在 100 个合成案例中，估计得到的阴影图平均峰值信噪比为：Neural Grey 达到 25.43 分贝，四次拍摄的 DeVigNet 达到 23.72 分贝。Neural Grey 能精确恢复一个光线追踪镜头轮廓，在其余 11 个上的平均值为 24.49 分贝。最接近的竞争方法（Kim 和 Pollefeys）在全部 12 个案例上的平均值为 21.95 分贝。在 21 组真实拍摄中，Neural Grey 的平均值为 25.38 分贝，比最强竞争方法的平均值高 5.62 分贝，并在七种光学配置中的五种上胜出。**

索引词—**渐晕校准、镜头阴影校正、逆成像、神经场、辐射度校准。**

一、引言

渐晕使图像亮度不仅取决于场景辐射度，也取决于位置。这给光度测量、科学成像以及更换镜头时的一致性校正带来了困难 [1], [2]。有效的阴影响应取决于相机、镜头、光圈和对焦状态 [3], [4]。我们旨在为其中一种配置建立可重复使用的校准映射。

在照明与设备条件合适时，受控平场采集仍是最准确的选择 [5]–[7]。然而，在实际使用中，重复进行这种标定可能很不方便。解析轮廓和数据库查找表效率较高，但可能无法涵盖特定设备的偏心、传感器响应，或未列出的镜头 [3], [4]。大型图像修复网络灵活性更高，但仅凭四张标定采集图像，无法提供种群模型通常所需的训练数据。

作者来自美国加利福尼亚州圣巴巴拉的加利福尼亚大学电气与计算机工程系（电子邮箱：yueze@ucsb.edu；sls@ucsb.edu）。通讯作者：Spencer LaVere Smith。

我们的出发点是一个简单的采集实验：拍摄同一面浅色墙壁，同时滚转相机。墙面照明会在传感器上移动，而相机–镜头阴影则保持不变。因此，将拍摄图像放在一起拟合，便能分离出其中大部分空间结构。但有一个关键限制：径向变暗在滚转后看起来仍然不变，因此仅凭图像无法判断它来自镜头还是墙壁。我们将此称为径向规范歧义。

Neural Grey 在整个设计中都遵循这一差异。余弦幂基函数和小型有界神经场能够拟合阴影与照明，同时避免任意复现墙面纹理。随后，径向选择器利用焦距和传感器尺寸来选择宽尺度衰减轮廓，并保留从拟合后的拍摄图像中得到的空间残差。输出是针对一种相机–镜头–光圈–焦点配置的可复用阴影图。我们的实验分别检验场表示、基于元数据的径向选择以及采集协议。

我们的贡献如下：

- 一种联合辐射度模型，可分离传感器固定阴影与场景固定照明，并阐明相机滚转后仍存在的空间歧义。
- 一种有界物理–神经表示，以及一种由焦点条件控制的径向选择器，可提供滚转采集图像中缺失的先验信息。
- 我们在合成轮廓、光线追踪光学系统、真实采集和实测轮廓压力测试案例上进行评估，并开展组件、简化、敏感性和采集数量测试，以区分各项设计选择的作用。

II. 相关工作

A. 物理与几何校正

经典校正会拟合余弦幂、椭圆形或偶多项式径向轮廓 [3], [5], , 有时还会学习光学中心偏移 [8], [9]。这些模型



图1. Neural Grey 遵循拍摄图像所能揭示和无法揭示的信息。(a) 相机滚转会使墙面照明相对于传感器固定阴影发生移动。(b) 物理基底与有界神经残差联合拟合阴影、场景照明、位姿和曝光。(c) 低频稳定化之后，焦距元数据用于选择模糊径向轮廓；其余空间细节予以保留。由此得到的场图可校正之后在相同光学配置下拍摄的图像。场图用于说明这一机制，并非实测结果。

高效且易于解释，但近似径向的假设无法涵盖所有单机特有偏心、传感器堆栈响应或非对称结构 [10]。回顾性显微校正则通过熵最小化 [11]或多图像模型 [2], [12]，根据图像内容估计共享阴影。辐射度校准还会根据曝光变化[13], [14], 估计相机响应，所用数据来自图像集合 [15], 或人脸 [16]。

B. 纯数据驱动模型

学习式校正包括 UNet 风格模型 [17]–[18]、VCLUT 等查找表（LUT）形式[19]，、频率感知型 DeVigNet [20]，以及离散高斯滤波器估计 [21]。这类群体模型通常需要具有代表性的训练数据，并预测校正后的图像。相比之下，我们的逐配置设置通过四张拍摄图像估计可复用阴影场；第V节比较了预训练基线和逐实例学习基线。相关照明研究还包括经典 Retinex 模型 [22], [23]，、学习式分解 [24], [25]，、无参考曲线与照明估计 [26]–[28], 以及噪声感知增强 [29]。这些方法针对的是图像增强，而非可复用的镜头标定，但都需要将照明与图像内容区分开来。

C. 物理锚定神经场与逆成像

物理信息学习将制约结构置于逆问题中 [30], [31]，而隐式神经场则提供紧凑的连续参数化 [32], [33]。相关表示对三维形状 [34]、连续图像 [36]，和照明相关图像变换 [37]进行建模。针对图像的网络拟合也可以在无需外部训练的情况下充当先验 [38]–[39]。Neural Grey 将这些工具结合到滚转-捕获因式分解中，并明确处理其径向歧义。

III. 所提方法

A. 问题表述

图 1 总结了该方法。我们首先使用独立的传感器场和场景场拟合所有拍摄图像，然后利用焦距元数据选出其中尚未确定的径向分量。两者的区别至关重要：准确重建四张输入图像是必要条件，但这本身并不能保证阴影正确。

令 x 表示二维传感器坐标，u 表示静态校准表面上的坐标。未知的 $V(x) > 0$ 固定于传感器，而表面辐亮度和照明场 $L(u) > 0$ 固定于场景。根据标准辐射度图像形成模型 [1], [10], [40]，拍摄图像 k 满足

$$I_k(x) = g_k\, V(x)\, L(W_k x) + b_k + n_k(x), \qquad (1)$$

其中，$k \in \{1, \ldots, 4\}$ 为稀疏、随意拍摄且近似旋转四分之一圈的图像编索引。相似变换 $W_k x = s_k R_{\alpha_k} x + t_k$、增益 g_k、黑电平偏移 b_k和零均值拍摄噪声 n_k 均为干扰变量；除n_k 外，其他变量均联合估计。通过重叠掩模移除越界坐标。我们使用四张拍摄图像，拍摄对象为静态、浅色且近似朗伯表面的物体，拍摄时相机大致旋转了四分之一圈。强度经过线性化处理；否则，依赖场景的相机处理可能会违反该模型 [41]。

从实际角度看，滚转后，墙壁上的标记或照明梯度会落到不同的传感器像素上，

而镜头阴影则始终位于相同的传感器像素上。公式（1）利用这种差异将这两个场分离，同时考虑轻微的手持抖动和曝光误差。

命题 1（相机旋转下的歧义）。假设 $W_k = R_k$ 是绕同一光学中心的精确旋转。对于每个对所有测得的旋转都保持不变的正值场 Q ， $Q(R_k x) = Q(x)$，该变换

$$V'(x) = V(x)Q(x), \qquad L'(u) = L(u)/Q(u) \tag{2}$$

使公式 (1) 中的所有无噪观测保持不变。证明。根据所述不变性， $V'(x)L'(R_k x) = V(x)Q(x)L(R_k x)/Q(R_k x) = V(x)L(R_k x)$ 对每个k均成立。

因此，任何在所有测得的滚转下看起来都不变的图案，都可以在 V 和 L 之间移动，而不会改变图像。对于理想的四分之一转，这包括每个正值 C_4-不变场，因此也包括每个径向场 $Q(x) = q(\|x\|)$。对数 L的均值中心化只能固定标量缩放；它无法消除这种空间歧义。有界场能够稳定可观测部分，而独立的焦距元数据则能选出主导的径向部分。

B. 有界阴影与照明场

我们将传感器固定阴影建模为平滑的径向基底，并加入有界空间校正：

$$\begin{aligned} V_{phys}(x) &= I_0 \cos^p(\arctan(r_{ell}/f_{norm}))\,, \\ V_{pred}(x) &= V_{phys}(x)\,[1 + \Delta V(x)]. \end{aligned} \tag{3}$$

椭圆半径 r_{ell} 采用学习得到的中心、轴向尺度和旋转，以适应偏心以及水平与垂直方向衰减不一致的情况。指数 $p \in [2, 6]$和正缩放因子 f_{norm} 控制基底轮廓； f_{norm}是一个优化变量，与后续使用的记录焦距不同。

一个紧凑的坐标多层感知器（MLP）通过平滑的 Tanh 激活函数表示相对于物理基底的偏差。其输出有界，并辅以低阶角向结构：

$$\Delta V(x) = \alpha_V \tanh(f_\eta(z)) + e(r_{ell})\mathcal{H}_\omega(\phi). \tag{4}$$

其中 z 结合了位置、椭圆半径和角度特征，而 $\mathcal{H}_\omega$ 包含有界的二阶和四阶谐波。径向包络 e会抑制中心附近的角度项，因为该处的方向定义不明确。两个分支均以零初始化，因此拟合从物理场开始。有界校正可适应空间偏差，同时限制对墙面纹理进行复现的自由度。补充材料中说明了这些特征、残差上限 $\alpha_V = 0.18$以及固定的网络维度。

场景照明场使用二阶多项式表示倾斜和宽域曲率，并加入紧凑的 Tanh 多层感知器，以表示多项式未能捕捉的平滑结构：

$$\begin{aligned} \ell(u) &= q_2(u) + 0.25\tanh(h_\psi(u)), \\ \log L(u) &= \ell(u) - \mathrm{mean}_u\,\ell(u). \end{aligned} \tag{5}$$

其神经项允许约 $[0.78, 1.28]$ 的乘性变化，之后再进行均值中心化。与 V（x）不同，该照明场在每次捕获中都根据变形场景位置 $u = W_k x$ 进行评估。分离的坐标系提供了跨旋转约束；多项式、有界残差和平滑的 Tanh 激活函数则限制了两个场各自能够吸收的变化。我们省略高频傅里叶编码，以优先获得平滑场 [42]，[43]；显式带宽控制是处理这一表示问题的另一种方法 [44]。

生成的型号包含约 24.8K 个参数。其容量在不同光学配置下保持固定；第 IV节将考察相近的网络规模或更简单的场表示是否能改善恢复效果。

C. 联合拟合与尺度归一化

损失函数有三个作用：数据项重建全部四幅图像，场项限制V 和 L所能吸收的结构量，而干扰项则使增益、形变、焦距尺度和中心保持在接近其采集先验的范围内。我们联合优化着色参数 $\{p, f_{norm}, x_c, y_c, a_x, a_y, \varphi\}$、每次捕获的干扰变量 $\{g_k, b_k, W_k\}$以及多项式与多层感知机场L：

$$\begin{aligned} \mathcal{L} = \mathcal{L}_{data} &+ \lambda_{phys}\mathcal{L}_{phys} \\ &+ \lambda_{light}\mathcal{L}_{light} + \lambda_{anchor}\mathcal{L}_{anchor} \\ &+ \lambda_{res}\mathcal{L}_{res} + \lambda_{gb}\mathcal{L}_{gb} \\ &+ \lambda_{warp}\mathcal{L}_{warp} + \lambda_f\mathcal{L}_f + \lambda_c\mathcal{L}_c. \end{aligned} \tag{6}$$

数据项通过 $w_e(x, y) = 1 + 1.25r^2$，确保所有四次捕获之间的一致性，并将其归一化为单位均值：

$$\mathcal{L}_{data} = \frac{1}{4}\sum_k w_e \left\| I_k - g_k \cdot V_{pred} \cdot L(W_k x) - b_k \right\|_2^2 . \tag{7}$$

归一化前，这会将角落的贡献从 1 提高到 2.25，从而抵消中等半径处图像面积的集中，并突出衰减信号最强的位置。这里， $\mathcal{L}_{phys} = \|V_{pred} - V_{phys}\|^2$ 保留物理基底、 $\mathcal{L}_{light} = \|$对数 $L\|^2$以及 $\mathcal{L}_{res} = \|\Delta V\|^2$。它由经过曝光归一化和配准的输入通过稳健融合构建而成，

V_{fused} 提供了一个不依赖任何外部平场参考的内部自监督锚点。我们使用$\mathcal{L}_{anchor} = \text{SmoothL1}(V_{pred}, V_{fused}; \beta = 0.03)$ [45]；透射率为 3% 时的转折点采用二次函数，以处理较小的融合误差；对于较大的墙壁或配准残留误差，则采用线性函数。此外，
$\mathcal{L}_{gb} = \|g-1\|^2 + \|b\|^2$,$\mathcal{L}_{warp}$ 表示平移、旋转和对数尺度偏差的平方，$\mathcal{L}_f$ 表示相对于元数据推导的归一化焦距的相对偏差平方，而 $\mathcal{L}_c$则用于惩罚光学中心偏离其稳健初始化位置。

我们对采样坐标上的场损失取平均，并对各次拍摄的干扰项惩罚取平均。所有数据集和镜头均使用相同的八个正则化权重。这些权重在重建精度、合理的场以及较小的采集误差之间进行平衡。补充材料列出了这些权重的取值，并将每个权重分别设为默认值的一半和两倍进行测试，以评估局部敏感性。

对 $\log L$ 进行均值中心化，可以固定阴影与照明之间的全局缩放关系。受限的逐次拍摄增益用于补偿曝光差异。具体边界和初始化设置见补充材料。

所有参数均使用 Adam [46]和余弦退火 [47]进行联合优化。神经残差采用最快的学习率；全局光学、照明和曝光变量变化得更慢，从而使局部场校正能够在稳定的光学和几何拟合过程中进行调整。补充材料给出了固定的优化计划、坐标采样方式、学习率和参数边界。

D. 选择剩余的径向轮廓

$\widehat{}$联合拟合返回 V，但拍摄模型无法唯一确定其宽尺度径向趋势。因此，我们利用焦距和传感器尺寸单独作出明确选择。这些关于光学配置的独立信息，与相机旋转所提供的空间约束相互补充。

在选择径向轮廓之前，我们会削弱可能从墙壁泄漏到阴影估计中的大型单侧低频结构。反射平均 $\mathcal{P}_{D_2}$会将平滑图与其水平反射、垂直反射和双轴反射进行平均。当不对称性较小时，软门控只施加轻微校正；当不对称性较大时，则施加更强的校正。将此 D_2 先验应用于低频分量后，可以抑制单侧照明泄漏，同时保留细尺度残差。它对反射对称性的偏好也可能衰减真实存在的宽尺度光学不对称性。

我们将稳定化对数图分解为径向分箱中位数 $p_b(r)$和空间残差 $a(x)$。焦距和传感器对角线共同定义了两个解析余弦幂律轮廓：p_w 代表更强的广角衰减，而p_s 则为标准焦距和更长焦距提供第二种轮廓。连续权重取决于焦距以及与拟合轮廓的偏差：

$$p_{out}(r) = (1-\beta_w-\beta_s)p_b(r) + \beta_w p_w(r) + \beta_s p_s(r),$$
$$V_{out}(x) = \mathcal{N}\{\exp[a(x) + p_{out}(r(x))]\}, \tag{8}$$

其中，$\mathcal{N}$ 将峰值设为 1。这只会替换稳定图中的径向轮廓，既不会重新拟合神经场，也不会查询镜头数据库。没有焦距元数据时，我们使用稳定后的估计值。

选择器在归一化传感器坐标中运行，利用径向中位数将宽尺度衰减与空间残差分离。高斯平滑的标准差为图像长边的 3.9%，径向轮廓被离散为 48 个区间。焦距软过渡以 35 mm 和 42 mm 为锚点，采用 35 mm 等效尺度。这些光学先验解决了命题 1 所指出的歧义；其数值是经验性开发选择，并在所有评估中保持不变。补充材料给出了完整规则及经过测试的敏感性范围。

IV. 模型分析与消融研究

本分析分别考察表示能力、径向歧义和采集鲁棒性。一项包含 15 个场景的表示研究由九个合成场景和六个独立的实测剖面场景组成，用于测试移除物理组件和神经组件的效果（表 I）。随后，采用12案例诊断基准评估径向选择、采集数量和局部敏感性（表 II–III）。该基准包含九种基于 Lensfun 的光学配置文件，以及三种带有模拟照明和噪声的解析光学剖面。在每项研究中，我们使用峰值信噪比（PSNR）比较估计的阴影图与参考图。对于 12 案例基准，我们还报告结构相似性指数（SSIM）。SSIM 比较两张图的亮度、对比度和结构，接近 1 的数值表示高度一致。

A. 物理基底与神经残差的互补性

表征诊断包含九个合成场景和六个独立的实测剖面场景，并加入了更强的噪声/抖动。该诊断将移除物理基底或神经残差后的结果，与常数图和二维网格进行比较。表I报告了平均值、最差情况以及各场景的胜出次数。此诊断用于支持表征方式的选择；接下来的 12 案例诊断基准实验将检验完整方法。

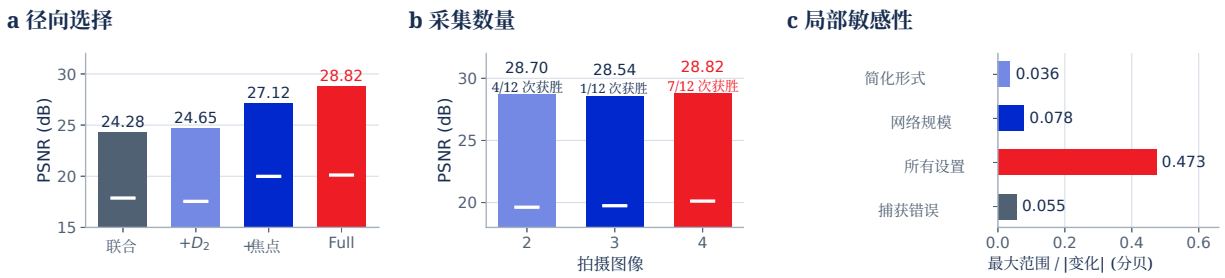

图2. 12案例诊断基准上的设计诊断。（a）固定四次采集拟合下的选择器变体。（b）使用前两次、三次或四次采集进行重新拟合。柱状图表示平均峰值信噪比；白色标记表示最差情况。（c）在简化形式、网络规模、损失权重和表征设置之间，平均峰值信噪比跨度的最大值；捕获错误表示端点绝对变化的最大值。面板（a、b）使用完整拟合计划；面板（c）使用较短的拟合计划来评估局部敏感性。捕获错误扫描对三个场景实现取平均值。补充材料给出了数值设置。

表I 15种表征情景（9种合成 + 6 独立测量剖面案例）的综合消融总结。数值为峰值信噪比（分贝）；下划线标记每个总结列中的最佳值。

变体	Avg	最差值胜出次数	
Neural Grey	31.31	16.11	10/15
不使用物理先验（纯神经网络）	28.50	9.71	0/15
不使用神经残差（纯物理）28.34		18.36	5/15
全部移除（输出 1.0）	23.43	7.35	0/15
纯二维网格	26.52	14.49	0/15

结合物理基底与神经残差可获得最高的平均峰值信噪比（31.31 分贝），并在 15 个案例中赢得 10 个。纯物理模型赢得其余五个案例，且最差情况得分更高（18.36 分贝，而非 16.11 分贝）。在完整研究中，加入有界神经残差可使平均峰值信噪比提高 2.97 分贝，而保留物理基底相比仅使用神经残差的变体可使其提高 2.81 分贝。这些比较共同表明，在光学模型的基础上进行受约束的空间校正是合理的：这两个组件各自带来互补的益处，从而提升平均恢复效果。

B. 焦距元数据与径向选择的作用

表 II 检验径向选择的贡献。仅联合拟合即可达到 24.28 分贝； D_2 稳定化可再提升 0.37 分贝。直接对联合估计应用径向选择器，可达到 27.12 分贝；在此之前进行 D_2稳定化还可再提升 1.69 分贝，达到 28.82 分贝。因此，径向选择带来的增益最大，而 D_2则提供了互补的稳定化效果。

C. 采集数量与方向采样的影响

平均峰值信噪比并不会随着每增加一次拍摄图像而提高：两次、三次和四次拍摄图像分别达到 28.70, 28.54 和 28.82 分贝。四次拍摄图像在 12 种情况中有 7 种表现最佳，相比之下，两次和三次拍摄图像分别为 4/12 和 1/12；四次拍摄图像的平均值最高，且最差情况下的得分也最高（20.12，对比 19.63/19.75 分贝）。四次拍摄图像在 $0°/90°/180°/270°$ 处均匀采样相机方向。四种情况下，两次拍摄图像仍然表现最佳，因此结果支持将其作为默认方案，而非最低采集要求。

D. 对采集扰动和超参数的鲁棒性

我们在 12案例诊断基准上，每次改变一个采集因素，并为每个案例使用三个场景随机种子，以测试捕获错误。在所测试的最强扰动下，平均峰值信噪比与该因素的零水平相差不到 0.056 分贝（表 III）。补充材料提供了采集生成设置。

针对数值敏感性，我们将每个损失权重调整为原来的两倍，并在固定其他设置的情况下测试相近的网络规模和表征设置。将残差上限改为 0.12/0.24、移除谐波或简化照明后，径向选择前的平均峰值信噪比跨度为 0.039 分贝，之后为 0.036 分贝。在 50轮协议下的 12 案例诊断基准上，网络规模跨度达到 0.078 分贝，而最大的损失权重跨度为残差惩罚的 0.473 分贝（图2c）。容量效应较小，支持采用紧凑固定网络；残差权重效应较大，则表明正则化仍然很重要。

选择器常量在 Enhanced-36 上单独测试；这是一个基于 Lensfun、加入了噪声和位姿扰动的基准。大多数跨度都低于 0.15 分贝，但更改宽配置余弦指数会产生 0.612 分贝的跨度。该指数直接控制广角径向先验的强度，因此这一

表II：12案例诊断基准上的径向选择、元数据和捕获数量消融。除最差值和胜出次数外，所有条目均为均值； Δ 相对于固定四次捕获完整模型计算。选择器/元数据各行使用相同的联合拟合；捕获数量各行则对前两次、三次或四次拍摄图像重新拟合。两次/三次拍摄图像的锚点为逐像素中位数；四次拍摄图像则使用截尾均值。指标采用固定的全帧域，且不进行真实值尺度拟合。

变体	峰值信噪比 ↑	Δ （分贝）	结构相似性指数 ↑	R^2 ↑	最差值 ↑	Wins
仅联合拟合	24.28	-4.54	0.9829	0.730	17.87	–
$+D_2$ 稳定	24.65	-4.17	0.9839	0.740	17.55	–
+径向选择器（无 D_2）	27.12	-1.69	0.9913	0.816	19.99	–
完整模型	**28.82**	**0.00**	**0.9932**	**0.838**	**20.12**	–
无焦距元数据（仅D_2 ）	24.65	-4.17	0.9839	0.740	17.55	–
2 张拍摄图像（0°/90°）	28.70	-0.12	0.9929	0.783	19.63	4/12
3张拍摄图像（0°/90°/180°）	28.54	-0.27	0.9930	0.789	19.75	1/12
4张拍摄图像（0°/90°/180°/270°）	28.82	0.00	0.9932	0.838	20.12	7/12

表 III 单因素采集容差在12案例诊断基准上的表现。峰值信噪比为12个案例和三个随机种子在最强端点处的平均值； Δ 与同一扫描的零水平进行比较。

因素	扫描范围 σ	PSNR	Δ与0相比
Roll	0–3°	28.618	−0.016
平移	0–8 像素	28.587	−0.055
缩放	0–3%	28.669	+0.041
对数曝光	0–0.15	28.628	−0.005

选择比在测试范围内调整分箱或平滑更为重要。因此，敏感性结果支持采用紧凑的固定字段，同时指出径向先验强度和残差正则化是影响更大的选择。

表 IV SYNTHETIC-100。平均峰值信噪比/结构相似性指数/平均绝对误差；N 为评估案例数量。下划线标记每个指标列中的最佳值，包括并列情况。

型号	N	峰值信噪比↑ 对误差↓	结构相似性指数↑	平均绝
Neural Grey	100	<u>25.43</u>	<u>1.00</u>	<u>0.06</u>
Zheng 等人 [5]	100	13.58	0.96	0.18
Goldman [10]	100	12.53	0.96	0.20
Kim 和 Pollefeys [9]	100	17.88	0.98	0.11
Cao 等人 [49]	100	17.31	0.92	0.14
DeVigNet（4-Shot） [20]	100	23.72	<u>1.00</u>	<u>0.06</u>
DeVigNet（1-Shot） [20]	100	16.08	0.91	0.14
Abdelhamed 等人 [21]	100	16.50	0.98	0.12
CNN-4Shot	100	15.60	0.96	0.16
UNet-4Shot [18]	100	13.48	0.95	0.21
解析径向（4F）	100	15.88	0.93	0.16
受 VCLUT 启发 [19]	100	16.79	0.98	0.12
解析径向（1F）	100	17.03	0.97	0.13

V. 不同光学配置下的评估

我们评估了四个互为补充的基准：Synthetic-100 包含合成阴影和照明；RayTraced-12 包含模拟的光学架构；Real-21 包含真实的校准拍摄；Enhanced-36 则结合了实测的 Lensfun 配置文件与模拟的采集扰动。我们将其与 Zheng 等人进行比较 [5], Cao 等人 [49], Goldman [10], Kim 和 Pollefeys [9], DeVigNet（4-Shot） [20], Abdelhamed 等人 [21], CNN-4Shot、UNet-4Shot [18], 解析径向（1F）、解析径向（4F）以及受 VCLUT 启发的 [19]基线方法。后缀 1F/4F 分别表示单帧和四帧变体；此外，还报告了 DeVigNet（1-Shot）在 Synthetic-100 上的结果。我们在统一的全帧域中报告峰值信噪比、结构相似性指数 [48], 和平均绝对误差。合成和射线追踪配置文件没有对应的实际焦距/传感器组合，因此用于测试联合拟合；真实相机基准则用于测试完整架构。

单帧方法使用标准的合成／射线采集，并在 Real-21 和 Enhanced-36 上使用四个独立估计结果的均值；多次采集方法则接收相同的四张图像。补充材料介绍了各项适配方法，并提供了逐次拍摄的实验方案。学习式方法比较涵盖固定预训练推理（DeVigNet（4-Shot））、稀疏的逐实例拟合（CNN-4Shot 和 UNet-4Shot）、查找表建模（受 VCLUT 启发）以及 Abdelhamed 高斯族适配器。

A. 合成阴影与光照

Synthetic-100 涵盖 100 个案例中的十种退化类别。阴影包括余弦幂、偶次多项式、非单调多区域、偏心椭圆和非参数样条等类型。照明从平滑渐变到硬边缘和高频纹理不等。联合扰动达到标准

表 V RAYTRACED-12：12 种光学架构。Neural Grey 的 24.49 分贝是 11 个案例的有限值均值，因为其中一个零误差案例的峰值信噪比为无穷大；结构相似性指数和平均绝对误差则保留全部 12 个案例。下划线标记各指标列中的最佳值，包括并列情况。

型号	N ↓	峰值信噪比↑	结构相似性指数↑	平均绝对误差↓
Neural Grey	12	<u>24.49</u>	<u>0.90</u>	<u>0.06</u>
Zheng 等人 [5]	12	16.71	0.80	0.14
Cao 等人 [49]	12	18.99	0.89	0.11
Goldman [10]	12	10.32	0.75	0.35
Kim 和 Pollefeys [9]	12	21.95	0.89	0.09
DeVigNet（4-Shot） [20]	12	14.18	0.78	0.20
Abdelhamed 等人 [21]	12	17.10	0.82	0.18
CNN-4Shot	12	17.72	0.87	0.14
UNet-4Shot [18]	12	17.28	0.82	0.15
解析径向（4F）	12	20.51	0.80	0.14
受 VCLUT 启发 [19]	12	10.07	0.74	0.32
解析径向（1F）	12	15.82	0.84	0.20

差： 2.2° 滚转、14 像素平移、1.4% 缩放和 5% 曝光，并包含光响应非均匀性 (PRNU)、读出噪声和散粒噪声。每种基线方法都涵盖全部 100 个案例。

Neural Grey 以 25.43 分贝位居第一，比 DeVigNet（4-Shot）高 1.70 分贝。最强的经典方法结果是 Kim 和 Pollefeys 的 17.88 分贝。补充材料中的类别级表格显示，在边缘极端和不对称性混乱的条件下，物理正则化最为有效。

B. 光线追迹光学架构上的验证

为验证阴影行为在物理上的合理性，RayTraced-12 通过光线追迹模拟评估了 12 种经典数学镜头架构（例如单透镜、库克三片式镜头、双高斯镜头、卡塞格林反射镜） [50]。表V总结了架构层面的比较结果。

Neural Grey 的有限值光线追迹均值为 24.49 分贝，领先 Kim 和 Pollefeys（21.95 分贝） +2.54 分贝，也高于解析径向（4F）的 20.51 分贝。在模拟的卡塞格林应力剖面上，其环形响应违背了光学中心最亮的假设，Neural Grey 达到 17.00 分贝，而 DeVigNet（4-Shot）达到 12.99 分贝。结合下文的商用镜头基准，这些模拟涵盖了多样的光学结构。

C. 商用镜头上的真实场景性能

Real-21 包含 21 次手持拍摄，来自七种 Sigma、Canon 和 Nikon 配置，涵盖定焦、微距、广角和远摄设计。每次拍摄均包含四张标定采集图像。真实值图来自单独的受控平场采集，且估计器无法获取这些图像。每张参考 RAW 图像都根据记录的黑电平和白电平，按拜耳位置进行线性化，再合成为线性亮度，去除脉冲噪声，以长边的 1.2% 为标准差进行高斯平滑，并通过稳健中心峰值估计进行归一化。四张随手拍摄的标定图像与参考图像使用相同的相机–镜头–光圈–焦点配置，且没有共用任何帧。表 VI中的所有数值均为三次拍摄的平均绝对峰值信噪比，未进行事后标量对齐。长边为 1024 像素时，Neural Grey 达到 25.38 分贝，比表现最强的竞争方法（Cao 等人，19.76 分贝）高出 5.62 分贝。这一优势并非源于高峰值信噪比的远摄案例：去除全部六次 105 毫米和 200 毫米拍摄后，Neural Grey 在剩余 15 次拍摄中的平均值仍为 21.26 分贝，而表现最强的剩余竞争方法（Cao 等人）为 15.74 分贝，优势为 $+5.52$ 分贝。在七种配置中，它有五种排名第一，包括 Sigma 20 毫米和 Nikon 28 毫米广角配置，以及 80 毫米、105 毫米和 200 毫米配置。在 Canon FL 55毫米配置下，Neural Grey 达到 20.87 分贝，比 Cao 等人（21.32 分贝）低 0.45 分贝；在 Micro 60毫米配置下，达到 12.98 分贝，比 UNet-4Shot（13.28 分贝）低 0.30 分贝。Micro 60毫米配置下，各方法的绝对峰值信噪比均较低。诊断标量校正只能消除 Neural Grey 的少量误差，表明剩余误差主要是空间误差（参见补充材料）。

Neural Grey 在 21 次拍摄中有 14 次表现最佳；与 Cao 等人（完整基准均值最高的竞争者）进行配对比较时，21 次中赢得 19 次，与 Zheng 等人比较时则赢得 18 次。重复拍摄用于测试捕获变化，而 Enhanced-36 则通过加入噪声和抖动的实测 Lensfun 配置文件，拓展了光学和捕获条件。每次拍摄的结果见补充材料。

图 3展示了 Neural Grey 在 Sigma 20 毫米镜头上表现最出色的一次拍摄（单次拍摄结果为 26.88 分贝；图中分辨率为 1000 像素时为 26.84 分贝），并与 Lensfun 以及按 Real-21 完整平均峰值信噪比排名前三的非 Neural Grey 方法进行了对比。在另一项受控纹理图像实验中，将每张测得的阴影图应用于同一张干净照片，再使用估计出的阴影图进行校正。在 21 次拍摄中，完整估计器达到 31.40 分贝，高于联合拟合的 29.00 分贝和未校正时的 19.58 分贝（见补充材料）。主要评估目标是以独立的平场参考为基准，恢复可重复使用的阴影图；这项受控图像实验展示了阴影图的精度如何转化为亮度校正效果。

表 VI REAL-21：七种商用光学配置，每种拍摄三次。数值为绝对峰值信噪比（分贝），未进行逐预测标量对齐；下划线标记每列的最佳结果。括号内为测试的变焦设置。所有配置均以最大光圈拍摄：从左到右，$f/1.4$, $f/3.5$, $f/5.6$, $f/1.2$, $f/2.8$ $f/2.5$ $f/2.8$ 28 80 ,,和 。尼康 – 毫米为 AF-D 版本。

型号	Sigma 20毫米	尼康28–80毫米 (28毫米)	尼康28–80毫米 (80毫米)	Canon FL 55毫米	尼康微距 60毫米	Nikon 105 毫米	尼康 80–200 毫米 (200 毫米)	Avg.
Neural Grey	26.17	23.85	22.43	20.87	12.98	35.41	35.95	25.38
Kim 和 Pollefeys [9]	12.15	14.16	15.43	13.46	13.23	21.70	15.78	15.13
Cao 等人 [49]	12.04	12.42	20.35	21.32	12.56	25.82	33.84	19.76
Zheng 等人 [5]	12.28	12.57	19.71	21.07	12.44	26.95	32.99	19.72
解析径向（1F）	11.47	13.36	19.57	19.85	12.48	25.69	31.51	19.13
Goldman [10]	3.52	7.18	10.34	7.43	7.63	15.82	9.03	8.71
DeVigNet（4-Shot） [20]	9.65	10.38	21.09	18.48	12.70	12.15	18.81	14.75
Abdelhamed 等人 [21]	12.85	12.01	20.63	19.93	12.15	20.98	31.92	18.64
受 VCLUT 启发 [19]	7.33	12.93	14.29	16.15	11.13	21.80	15.15	14.11
CNN-4Shot	11.84	16.39	21.48	13.75	13.25	22.91	27.25	18.13
UNet-4Shot [18]	13.00	16.43	17.95	14.04	13.28	14.47	21.81	15.85

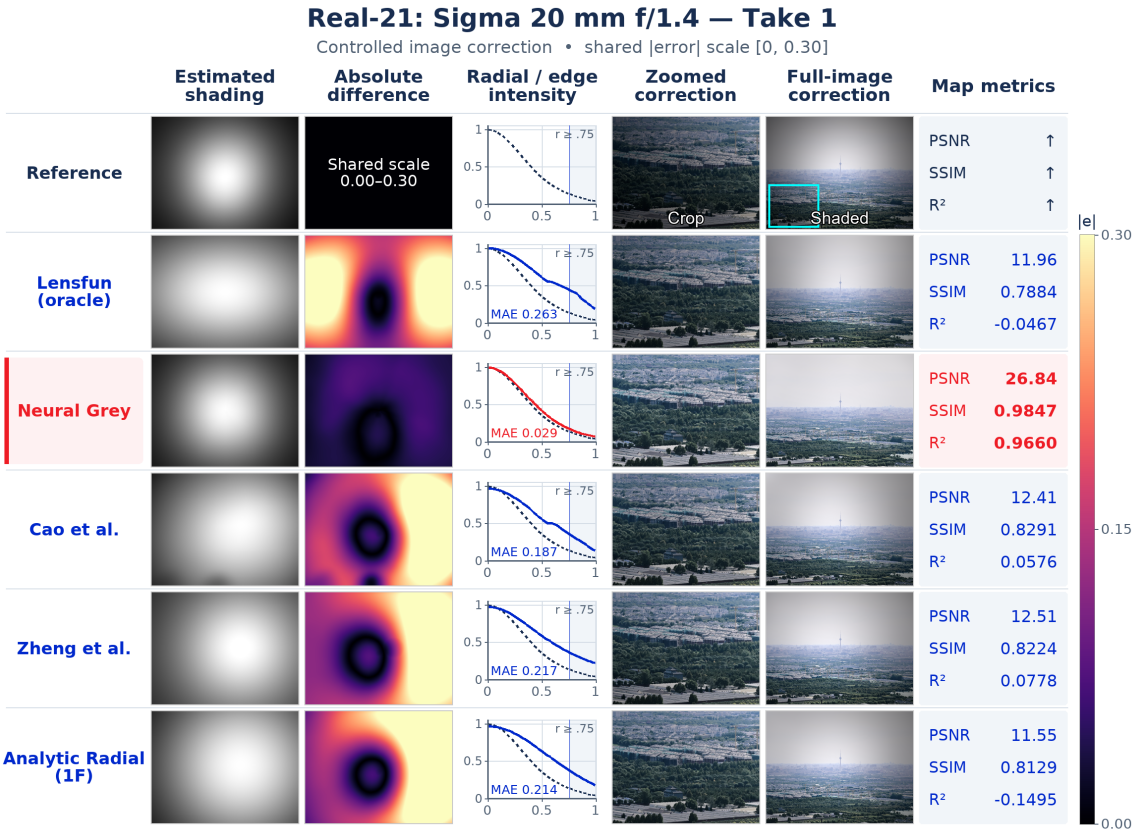

图 3。Real-21 适马 20 毫米 f/1.4，第 1 次拍摄。差异图使用固定的 $[0, 0.30]$ 线性强度尺度。纹理图像面板为受控模拟：在应用测得的阴影后，将其置于线性 RGB 中，并使用各估计结果将其移除。映射指标使用所显示的 1000 像素估计结果和独立的平场参考；定量表格使用 1024 像素的长边。单帧定性基线使用第 1 次捕获。Lensfun 使用依据真实值（第 V-D节）选择的插值。

D. 与标准查找表软件的比较

Lensfun 代表摄影软件中常见的、由元数据驱动的查找表校正：一款镜头相机身份、焦距、光圈和对焦距离用于从数据库中选取插值后的镜头配置文件。对于每个已涵盖的镜头，平场真实值会选出最佳插值的 Lensfun 配置文件。

表 VII 增强版强版-36。每个镜头列均为六种固定墙壁峰值信噪比（dB）的平均绝对值，涵盖六种固定墙壁/照明配置；全部36 种情况均采用表格VI协议。下划线标记每列的最佳结果。

型号	iPhone XS 4.25 毫米	Sigma 19 毫米 $f/2.8$ DN	Pentax-DA 18–135 毫米 (100 毫米)	Sigma 60–600 mm +1.4× 转换器 (840 毫米)	Laowa 85 mm 2× 超微距	奥林巴斯 75 mm $f/1.8$	Avg.
Neural Grey	<u>25.02</u>	<u>21.11</u>	<u>17.81</u>	18.97	<u>20.62</u>	<u>20.06</u>	<u>20.60</u>
Kim 和 Pollefeys [9]	20.09	16.82	17.57	17.59	17.02	17.94	17.84
Cao 等人 [49]	18.50	14.60	10.12	13.32	13.00	11.77	13.55
Zheng 等人 [5]	19.42	15.19	10.48	14.64	13.63	12.36	14.28
解析径向（4F）	15.27	9.48	12.36	12.38	12.70	6.57	11.46
Goldman [10]	6.51	11.52	15.80	15.98	13.16	15.29	13.04
DeVigNet（4-Shot） [20]	15.70	17.94	12.55	16.19	17.67	12.93	15.50
Abdelhamed 等人 [21]	19.93	17.66	17.18	17.94	15.29	18.51	17.75
受 VCLUT 启发 [19]	12.64	17.47	16.36	<u>19.11</u>	17.13	17.34	16.68
CNN-4Shot	19.86	14.14	9.85	12.36	12.17	11.44	13.30
UNet-4Shot [18]	20.76	12.32	9.10	11.19	11.11	9.97	12.41

在 Sigma 20 毫米和 Micro 60 AF-D 的精确记录、光学等效的 AF/AF-D 80–200 毫米记录，以及尼康 105 毫米的代际候选记录中，即使由预言机选出的查找表，其表现仍不及 Neural Grey 的原位估计。Canon FL 55毫米没有数据库配置文件。完整数值见补充材料。这表明，即使采用预言机进行配置文件选择，我们基于特定配置的原位估计仍优于标准数据库查找表。

E. Enhanced-36

Enhanced-36 使用表 VII中的六种 Lensfun 镜头配置文件，每种配置均采用六种模拟拍摄条件。每个案例都结合了测得的 Lensfun 阴影配置文件、不规则朗伯墙面、独立照明、半对角线长度 0.5–3% 的偏心、轻微的滚转/位姿和曝光变化，以及散粒噪声/读出噪声/行噪声。表 VII采用相同的比较流程，对表 VI中的所有方法以及 Abdel-hamed 等人的方法进行了评估。

Neural Grey 在六个镜头列中的五列以及 36 种情况的平均值上均居首，达到 20.60 分贝。Kim 和 Pollefeys 的方法以 17.84 分贝位居第二，其次是 Abdelhamed 等人，达到 17.75 分贝。在剩余配置中，即搭配 1.4×增距镜、焦距为 840 毫米的 Sigma 60–600 mm 镜头，Neural Grey 达到 18.97 分贝，与 VCLUT 启发的方法（19.11 分贝）仅差 0.14 分贝。在六种光学配置文件及其采集扰动下，同一固定配置取得最高平均值，并在五组镜头中领先，证明了该方法适用于所测试的各种光学条件。

VI. 结论

Neural Grey 结合四滚转辐射因子分解、有界物理–神经场和焦距元数据，区分传感器固定阴影与场景固定照明，并据此选出原本模糊的径向轮廓。该方法在合成基准和光线追踪基准上均排名第一，在 21 组商用镜头拍摄的图像上达到 25.38 分贝，比最强竞品的真实场景平均值高 5.62 分贝。相同的 24.8K 参数方案涵盖 20–200 毫米配置，并生成可重复使用的特定设备阴影图。

参考文献

[1] P. E. Debevec 和 J. Malik，“恢复 HDR 辐射图”，载于 *SIGGRAPH*，1997年。[2] K. Smith、Y. Li、F. Piccinini、G. Csucs、C. Balazs、A. Bevilacqua 和 P. Horvath，“CIDRE：一种用于光学显微镜的照明校正方法”，*NatureMethods*，第 12 卷，第 404–406 页，2015年。[3] E. Hecht，《光学》，第 5 版。美国马萨诸塞州波士顿：Pearson，2016年。[4] Cooke Optics Ltd.，《i 技术参考手册：智能镜头元数据协议》，技术报告，2021年。[5] Y. Zheng、S. Lin 和 S. B. Kang，“单图像渐晕校正”，*IEEE*模式分析与机器智能汇刊，2009年。[6] W. Yu，“实用抗渐晕方法”，*IEEE TCE*，2004年。[7] P. Kask、K. Palo、C. Hinnah 和 T. Pommerencke，“ 用于荧光样本高通量成像的平场校正 ”，*J. Microscopy*，第 263 卷，第 3 期，第 328–340 页，2016年。[8] S. B. Kang 和 R. Weiss，“我们能否利用一张平坦、无纹理的朗伯表面图像校准相机？ ”，载于*ECCV*，2000年。[9] S. J. Kim 和 M. Pollefeys，“鲁棒辐射度校准与渐晕校正”，*IEEE*模式分析与机器智能汇刊，第 30 卷，576，2008年4月，562，第 4 期，第 562 页。[10] D. B. Goldman，“渐晕与曝光校准及补偿”，*IEEE*模式分析与机器智能汇刊，2010年。[11] B. Likar、J. B. A. Maintz、M. A. Viergever 和 F. Pernuš，“基于熵最小化的回顾性阴影校正”，*J. Microscopy*，第 197 卷，第 3 期，第 285–295 页，2000年。

[12] T. Peng、K. Thorn、T. Schroeder、L. Wang、F. J. Theis、C. Marr 和 N. Navab，“用于光学显微图像背景与阴影校正的 BaSiC 工具”，Nature Communications，第 8 卷，文章编号 14836，2017年。[13] T. Mitsunaga 和 S. K. Nayar，“辐射度自校准”，载于 IEEE计算机视觉与模式识别会议论文集，第 1 卷，第 374–380 页，1999年。[14] M. D. Grossberg 和 S. K. Nayar，“从图像确定相机响应：哪些信息是可知的？”，IEEE模式分析与机器智能汇刊，第 25 卷，第 11 期，第 1455–1467 页，2003年。[15] Z. Mo、B. Shi、S.-K. Yeung 和 Y. Matsushita，“互联网照片集的辐射度校准”，载于 IEEE计算机视觉与模式识别会议论文集，第 4738–4746 页，2017年。[16] C. Li、S. Lin、K. Zhou 和 K. Ikeuchi，“基于图像中人脸的辐射校准”，载于 IEEE计算机视觉与模式识别会议论文集，第 3117–3126 页，2017年。[17] O. Ronneberger、P. Fischer 和 T. Brox，“U-Net：用于生物医学图像分割的卷积网络”，载于医学图像计算与计算机辅助干预会议，2015年。[18] C. Chen、Q. Chen、J. Xu 和 V. Koltun，“学习在黑暗中看见”，载于计算机视觉与模式识别会议，2018年。[19] Y. Li 等，“先验视觉引导的自监督学习实现彩色渐晕校正”，IEEE生物医学与健康信息学汇刊，2024年。[20] S. Luo 等，“DeVigNet：通过双重聚合融合变换器实现高分辨率渐晕去除”，载于人工智能促进协会年会，2024年。[21] A. Abdelhamed、J. Yim、A. Punnappurath、M. S. Brown、J. Choe 和 K. Kim，“从照片中提取渐晕和颗粒滤镜效果”，载于 IEEE/CVF冬季计算机视觉应用会议论文集，2022年。[22] E. H. Land 和 J. J. McCann，“明度与视网膜皮层理论”，《美国光学学会杂志》，第 61 卷，第 1 期，第 1–11 页，1971年。[23] D. J. Jobson、Z. Rahman 和 G. A. Woodell，“用于弥合彩色图像与人类对场景观察之间差距的多尺度视网膜皮层算法”，IEEE图像处理汇刊，第 6 卷，976，1997年7月，第 965 页。[24] C. Wei、W. Wang、W. Yang 和 J. Liu，“用于低照度增强的深度视网膜皮层分解”，载于英国机器视觉会议论文集，文章编号 155，2018年。[25] Y. Zhang、J. Zhang 和 X. Guo，“点亮黑暗：一种实用的低照度图像增强器”，载于 ACM多媒体会议论文集，第 1632–1640 页，2019年。[26] C. Guo、C. Li、J. Guo、C. C. Loy、J. Hou、S. Kwong 和 R. Cong，“用于低照度图像增强的零参考深度曲线估计”，载于 IEEE/CVF计算机视觉与模式识别会议论文集，第 1780–1789 页，2020年。[27] R. Liu、L. Ma、J. Zhang、X. Fan 和 Z. Luo，“结合协同先验架构搜索的视网膜皮层启发式展开，用于低照度图像增强”，载于 IEEE/CVF计算机视觉与模式识别会议论文集，第 10561–10570 页，2021年。[28] L. Ma、T. Ma、R. Liu、X. Fan 和 Z. Luo，“迈向快速、灵活且鲁棒的低照度图像增强”，载于 IEEE/CVF计算机视觉与模式识别会议论文集，第 5637–5646 页，2022年。[29] X. Xu、R. Wang、C.-W. Fu 和 J. Jia，“信噪比感知低照度图像增强”，载于 IEEE/CVF计算机视觉与模式识别会议论文集，第 17714–17724 页，2022年。[30] M. Raissi、P. Perdikaris 和 G. E. Karniadakis，“物理信息神经网络：一种深度学习框架”，《计算物理学杂志》，2019年。[31] G. E. Karniadakis、I. G. Kevrekidis、L. Lu、P. Perdikaris、S. Wang 和 L. Yang，“物理信息机器学习”，Nature Reviews Physics，第 3 卷，第 6 期，第 422–440 页，2021 年。[32] B. Mildenhall 、 P. P. Srinivasan、M. Tancik、J. T. Barron、R. Ramamoorthi 和 R. Ng，“神经辐射场：将场景表示为神经辐射场以合成视图”，载于 ECCV，2020 年。[33] V. Sitzmann、J. N. P. Martel、A. W. Bergman、D. B. Lindell 和 G. Wetzstein，“具有周期激活函数的隐式神经表示”，载于神经信息处理系统大会，2020年。

[34] J. J. Park, P. Florence, J. Straub, R. Newcombe 和 S. Lovegrove，《DeepSDF：学习用于形状表示的连续有符号距离函数》，载于*IEEE/CVF*计算机视觉与模式识别会议论文集，第 165–174 页，2019年。[35] L. Mescheder、M. Oechsle、M. Niemeyer、S. Nowozin 和 A. Geiger，《占用网络：在函数空间中学习三维重建》，载于*IEEE/CVF*计算机视觉与模式识别会议论文集，第 4460–4470 页，2019年。[36] Y. Chen、S. Liu 和 X. Wang，《利用局部隐式图像函数学习连续图像表示》，载于*IEEE/CVF*计算机视觉与模式识别会议论文集，第 8628–8638 页，2021年。[37] S. Yang、M. Ding、Y. Wu、Z. Li 和 J. Zhang，《用于协同低照度图像增强的隐式神经表示》，载于*IEEE/CVF*国际计算机视觉大会论文集，第 12918–12927 页，2023年。[38] D. Ulyanov、A. Vedaldi 和 V. Lempitsky，《深度图像先验》，载于*IEEE*计算机视觉与模式识别会议论文集，第 9446–9454 页，2018年。[39] Y. Gandelsman、A. Shocher 和 M. Irani，《Double-DIP：通过耦合深度图像先验进行无监督图像分解》，载于*IEEE/CVF*计算机视觉与模式识别会议论文集，第 11026–11035 页，2019年。[40] M. D. Grossberg 和 S. K. Nayar，《相机响应函数空间建模》，*IEEE*模式分析与机器智能汇刊，第 26 卷，第 10 期，第 1272–1282 页，2004年。[41] S. Nam 和 S. J. Kim，《利用深度神经网络对相机中的场景依赖成像进行建模》，载于*IEEE*国际计算机视觉大会论文集，第 1717–1725 页，2017年。[42] N. Rahaman、A. Baratin、D. Arpit、F. Draxler、M. Lin、F. A. Hamprecht、Y. Bengio 和 A. Courville，《论神经网络的频谱偏置》，载于国际机器学习会议，2019年。[43] M. Tancik、P. P. Srinivasan、B. Mildenhall、S. Fridovich-Keil、N. Raghavan、U. Singhal、R. Ramamoorthi、J. T. Barron 和 R. Ng，《傅里叶特征使网络能够在低维域中学习高频函数》，载于神经信息处理系统大会，2020年。[44] D. B. Lindell、D. Van Veen、J. J. Park 和 G. Wetzstein，《BACON：用于多尺度场景表示的带限坐标网络》，载于*IEEE/CVF*计算机视觉与模式识别会议论文集，第 16252–16262 页，2022年。[45] P. J. Huber，《稳健估计》，统计数学年刊，1964年。[46] D. P. Kingma 和 J. Ba，《Adam：一种随机优化方法》，载于国际学习表征会议，2015年。[47] I. Loshchilov 和 F. Hutter，《SGDR：带热重启的随机梯度下降》，载于国际学习表征会议，2017年。[48] Z. Wang 等，《图像质量评估：从误差可见性到结构相似性》，*IEEE*图像处理汇刊，2004年。[49] H. Cao、X. Gu、M. Zhang、H. Zhang 和 X. Chen，《面向面阵传感器的和谐二维高斯滤波器渐晕校正》，*IEEE*计算成像汇刊，第 8 卷，第 576–584 页，2022年。[50] W. Smith，《现代镜头设计》。麦格劳-希尔教育，2004年。