

# PhoCoLens：无透镜成像中的真实感和一致重建

Xin Cai<sup>1,2</sup>, Zhiyuan You<sup>1</sup>, Hailong Zhang<sup>3</sup>, Wentao Liu<sup>2,4</sup>, Jinwei Gu<sup>1</sup>, Tianfan Xue<sup>1,2</sup>

<sup>1</sup>The Chinese University of Hong Kong, <sup>2</sup>Shanghai Artificial Intelligence Laboratory,

<sup>3</sup>Tsinghua University, <sup>4</sup>SenseTime

{cx023, yz023, tfxue}@ie.cuhk.edu.hk, jwgu@ie.cuhk.edu.hk,

{zhanghl21}@mails.tsinghua.edu.cn, liuwentao@sensetime.com

## Abstract

与传统的基于镜头的系统相比，无镜头相机在尺寸、重量和成本方面具有显着的优势。由于没有聚焦镜头，无镜头相机依靠计算算法从多重测量中恢复场景。然而，当前的算法面临着不准确的前向成像模型和先验不足的问题，无法重建高质量图像。为了克服这些限制，我们引入了一种新颖的两阶段方法，用于一致且逼真的无透镜图像重建。我们方法的第一阶段通过专注于使用随空间变化的反卷积方法精确重建低频内容来确保数据一致性，该方法可适应相机视场中点扩散函数（PSF）的变化。第二阶段通过结合预先训练的扩散模型的生成先验来增强照片真实感。通过对第一阶段检索到的低频内容进行调节，扩散模型有效地重建了通常在无透镜成像过程中丢失的高频细节，同时还保持了图像保真度。与现有方法相比，我们的方法在数据保真度和视觉质量之间实现了卓越的平衡，正如两种流行的无透镜系统 PhlatCam 和 DiffuserCam 所证明的那样。项目网站：[phocolens.github.io](http://phocolens.github.io)。

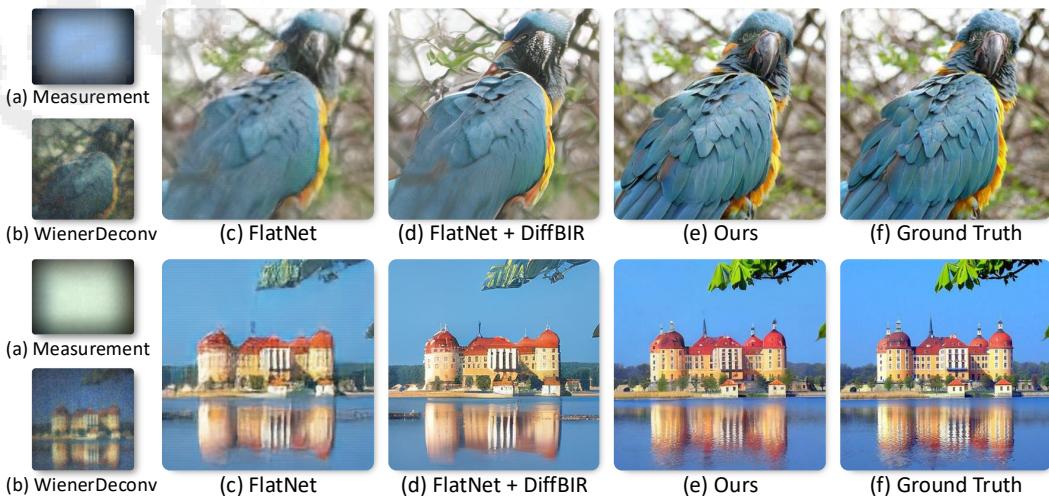


图 1：我们引入了 PhoCoLens，这是一种无透镜重建算法，与现有方法相比，它可以实现更好的视觉质量和与地面事实的一致性。与传统的重建算法 (b) 和 (c) 相比，我们的方法恢复了更多的细节，并且与生成方法 (d) 相比，还保持了对地面事实更好的保真度。



w/o spatially varying deconv      with spatially varying deconv

图 2: 没有 (左) 和有 (右) 空间变化反卷积的结果之间的比较。

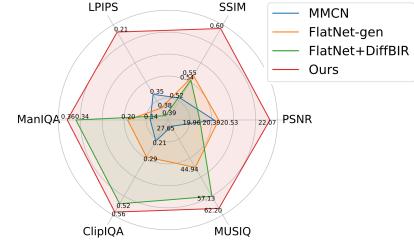


图 3: PhlatCam [17] 数据集的保真度和真实性评估。

## 1 简介

无透镜成像系统 [6] 已成为超紧凑、轻便且经济高效的成像的突破性解决方案。它们用靠近传感器放置的振幅 [3, 4] 或相位掩模 [1, 5, 20] 代替透镜来调制入射光。这种设计显着减小了相机的尺寸和重量，并实现了创新的传感器形状，例如球形 [14] 或圆柱形。

考虑到无镜头相机的原始测量结果通常是模糊且无法识别的，因此很难恢复原始场景的高质量图像。图 1 左栏显示了 WienerDeconv [44] 的无透镜测量和恢复图像的示例。由于缺乏聚焦元件，无透镜相机无法直接记录场景，而是将其编码为复杂的衍射图案。该编码过程可以近似为与大点扩散函数 (PSF) 的卷积。PSF 的作用就像应用于场景的低通滤波器，从而引入模糊性，这意味着单次测量可能有多种可能的恢复。因此，强大的算法对于无透镜系统中的高质量重建至关重要。

无镜头重建的首要挑战是同时实现真实感和一致性。真实感要求重建质量高、细节丰富，而一致性则进一步要求重建内容与原始场景保持一致。传统技术，如 WienerDeconv，可以重建与地面真实情况一致的图像，但视觉质量显着下降（图 1b）。基于学习的方法（例如 FlatNet-gen [17]）尝试通过配对图像和无透镜测量进行训练来增强视觉质量，但它们通常无法恢复高频细节（图 1c）。使用生成先验的恢复算法可以进一步提高视觉质量，例如 DiffBIR [23]（图 1d）。虽然这些恢复方法可以注入丰富的细节，但它们也可能会改变图像内容或插入不存在的对象，从而破坏一致性。例如，在图 1d 中，与地面真实情况（图 1f）相比，顶行的喙形状错误，而底行的叶子看起来是假的。

此外，不准确的成像过程模拟也可能损害数据的一致性。大多数现有的重建算法 [17, 27] 将成像过程简化为具有平移不变 PSF 的卷积。然而，实际上，PSF 在空间上是变化的，特别是当入射角增加时。因此，这些区域的重建与原始场景的相似度明显下降，尤其是在外围视野中。

为了实现照片真实感和一致性，我们提出了一种基于距离零空间分解的两阶段无透镜重建[36]。根据这种分解，重建图像由两个分量组成，一个来自“距离空间”，可以直接从无透镜测量中计算出来（图1a），另一个来自“零空间”，这是无透镜测量过程中丢失的详细纹理成像过程。因此，第一阶段通过恢复“范围空间”分量来优先考虑数据一致性，“范围空间”分量是与真实值匹配的低频内容。第二阶段的重点是照片真实感，通过从“零空间”添加更多高频细节，同时保持第一阶段建立的一致性。

在第一阶段，为了提高一致性，我们提出了一种新颖的空间变化反卷积来重建“范围空间”内容。与假设平移不变 PSF 的传统方法不同，我们的方法以数据驱动的方式自动适应相机视场中 PSF 的空间变化。这项创新更准确地模拟了前向成像过程，从而更好地重建了低频组件的结构完整性和细节。图 2 显示了我们的方法（右）相对于使用单核进行反卷积的方法（左）的改进。

在第二阶段，为了提高照片真实感，我们使用预先训练的扩散模型集成生成先验，将真实的细节插入到第一阶段的输出中。该模型专门针对恢复无透镜成像过程中丢失的高频细节。我们根据第一阶段重建的低频内容来调节扩散模型，引导其合并“零空间”中缺失的高频元素。我们的监督方法确保最终图像与实际测量一致，并实现逼真的质量。

结合这两个阶段，我们的无透镜重建在重建图像的数据一致性和视觉质量之间实现了良好的平衡，超越了现有的方法，如图1e所示。为了验证这一点，我们使用评估保真度和视觉质量的指标，将我们的方法与两种类型的无镜头相机上的其他方法进行比较：PhlatCam [17] 和 DiffuserCam [27]。PhlatCam 数据集上的定性比较显示了两个方面的改进，如图 3 所示，突出了我们的技术相对于当前方法的优越性。

## 2 相关工作

传统上，无透镜成像是通过解决卷积成像模型的正则化最小二乘问题来解决的，通常会结合梯度或频域中的稀疏性约束[1,5,21,31]。深度学习通过图像到图像网络实现可学习的反卷积参数和感知增强，彻底改变了无透镜成像。最近的创新包括深度展开技术 [19,27,49]，以及图像空间 [17] 或特征空间 [22] 中的前馈反卷积方法。然而，这些方法通常在成像过程中假设恒定的点扩散函数。我们的建议引入了一种空间变化的反卷积方法来克服这一限制并实现更精确的图像重建。

由于 PSF 在 FoV 上变化的成像系统普遍存在，空间变化的反卷积已成为一个深入研究的领域 [11,24,46]。然而，这些方法[2,20,25,29,47]通常速度慢，计算量大，并且导致图像质量差，尤其是在复杂系统中。最近，MultiWienerNet [48] 引入了一种用于快速、空间变化反卷积的深度学习方法，但它需要繁琐的多位置 PSF 校准。相比之下，我们的空间变化反卷积方法专为无透镜成像而设计，利用单个初始 PSF 并自动学习整个图像的变化，从而无需进行大量校准。

使用扩散模型的逆成像可以分为监督成像或零样本成像。监督方法使用配对图像训练条件扩散模型 [8, 13, 12, 32, 50]，以弥合输入和所需输出之间的差距，利用逆向成像 [35, 39] 的生成先验和受控生成 [28, 34]。另一方面，零样本方法 [37] 采用指导来解决各种一般逆问题 [9, 38]。这些技术通常依赖于预定义的条件进行指导。DDRM [15] 和 DDNM [41] 等特定模型侧重于数学分解以改善扩散过程。我们的工作整合了两种方法的优势，将监督微调与基于距离零空间分解的理论框架相结合，旨在为逆向成像挑战提供稳健的解决方案。

## 3 Preliminary

### 3.1 范围-零空间分解

无透镜成像过程可以表示为线性变换[10]。具体来说，给定场景平面  $x \in \mathbb{R}^{M \times 2}$ ，传感器测量  $\hat{y} \in \mathbb{R}^{N \times 2}$  表示为  $\hat{y} = Ax + n$ 。其中， $A \in \mathbb{R}^{N \times 2 \times M \times 2}$  为无透镜成像系统的传递矩阵， $n$  为传感器噪声。传输矩阵  $A$  本质上编码了来自场景平面上每个点的光如何影响每个传感器像素。给定一个校准的相机系统，传输矩阵  $A$  已知，无透镜重建的目标是从测量（如图 1a）中恢复场景平面图像  $x$ （如图 1f）。为简单起见，下面我们考虑无噪声情况，即  $y = Ax$ 。

然后我们引入范围零空间分解。设  $A^\dagger \in \mathbb{R}^{M \times 2 \times N \times 2}$  为线性矩阵  $A$  的伪逆，满足  $AA^\dagger A = A$ 。操作  $A^\dagger A$  可以解释为到  $A$  的范围空间上的投影，因为对于任何样本  $x$ ，我们有  $AA^\dagger A = A$ 。相反，运算  $(I - A^\dagger A)$  充当  $A$  零空间上的投影，假设  $A(I - A^\dagger A)x = 0$ 。因此，任何样本  $x$  都可以分解为两个正交分量：那在于

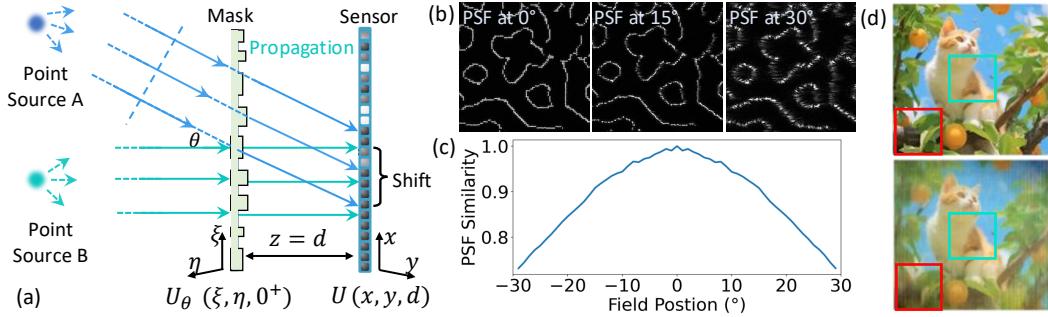


图 4: 无镜头相机中 PSF 的表征。(a) 无透镜相机中光传播的图示: 无限远处的两个点源 A 和 B 发射平行光束。光源 A 相对于光轴以  $\theta$  角发射, 导致传感器平面上发生 PSF 偏移。该 PSF 偏移取决于入射角  $\theta$  以及传感器与掩模之间的距离  $d$ 。(b)  $0^\circ$ 、 $15^\circ$  和  $30^\circ$  角度光源的模拟 PSF。(c) 不同场位置的轴上 PSF 和离轴 PSF 之间的内积相似性。(d) 使用  $0^\circ$  处的 PSF 进行重建, 外围 (红色框) 的退化比中心 (绿色框) 更显着。

A 的值域空间, 另一个在 A 的零空间。数学上, 这种分解是:

$$\mathbf{x} \equiv \mathbf{A}^\dagger \mathbf{A} \mathbf{x} + (\mathbf{I} - \mathbf{A}^\dagger \mathbf{A}) \mathbf{x}, \quad (1)$$

其中  $\mathbf{A}^\dagger \mathbf{A} \mathbf{x}$  是范围空间分量,  $(\mathbf{I} - \mathbf{A}^\dagger \mathbf{A}) \mathbf{x}$  是零空间分量。

通过这种分解, 我们得出了表示一致性和照片真实感的两个不同元素。“范围空间”项  $\mathbf{A}^\dagger \mathbf{A} \mathbf{x}$  确保了一致性, 因为它乘以  $\mathbf{A}$  得到测量值  $\mathbf{y}$ , 满足一致性条件  $\mathbf{A} \mathbf{x} = \mathbf{y}$ 。相反, “零空间”项  $(\mathbf{I} - \mathbf{A}^\dagger \mathbf{A}) \mathbf{x}$  确保重建图像  $\mathbf{x}$  与自然图像统计数据一致。

### 3.2 卷积无透镜成像模型的不匹配

在无透镜文献中, 大多数研究人员[1,5,17,27]将成像过程简化为卷积, 因为完整的传输矩阵  $\mathbf{A} \in \mathbb{R}^{N_2 \times M_2}$  太大而无法计算。具体来说, 测量结果为  $\mathbf{y} = \mathbf{h} * \mathbf{x}$ , 其中  $\mathbf{h}$  是无透镜系统的点扩散函数 (PSF),  $*$  表示卷积。以前的大多数重建算法都是基于这个假设。

然而, 真正的无透镜成像模型并不是如图所示的简单空间不变卷积。4a. 考虑一个无透镜相机, 由放置在纵向位置  $z = 0$  的掩模和放置在  $z = d$  的传感器组成。令  $U(x, y, z)$  为复标量波场, 是横向坐标  $x, y$  和纵向位置  $z$  的复函数。将从无穷远点源以入射角  $\theta$  传播通过掩模后的波场表示为  $U_\theta(\psi, \eta, 0^+)$ 。利用惠更斯-菲涅耳原理 [10], 强度图案  $p_\theta(x, y)$  传感器捕获的数据为:

$$p_\theta(x, y) = |U_\theta(x, y, d)|^2 = \left| \frac{d}{j\lambda r^2} \iint U_\theta(\xi, \eta, 0^+) \exp(jkr) d\xi d\eta \right|^2, \quad (2)$$

where the distance  $r$  is given by  $r = \sqrt{d^2 + (x - \xi)^2 + (y - \eta)^2}$  and  $\lambda$  is the wavelength of light.

因此, 根据方程 2, 空间不变卷积成像模型仅满足菲涅尔近似下[5, 10]:

$$r = \sqrt{d^2 + (x - \xi)^2 + (y - \eta)^2} \approx d + (x - \xi)^2/(2d) + (y - \eta)^2/(2d). \quad (3)$$

仅当无透镜掩模和传感器之间的距离  $d$  足够大以满足  $d \gg (x - \xi)^2 + (y - \eta)^2$  时, 该近似才有效。然而, 大多数无透镜遮罩都非常靠近传感器 (在典型的无透镜相机中,  $d = 2\text{mm}$ ), 打破了这一假设。

这种不匹配可能导致卷积成像模型不准确[26,43,52]。为了证明这一点, 我们使用方程 1 来模拟典型无透镜相机 (Phlatcam [5]) 在不同入射角  $\theta$  上的 PSF。(2)。如图 4b 所示, 不同角度 ( $0^\circ$ 、 $15^\circ$  和  $30^\circ$ ) 的 PSF 在视觉上是不同的。定量地, 图 4c 显示了中心 PSF 和  $-30^\circ$  从 1.0 下降到 0.7 之间的相似性。

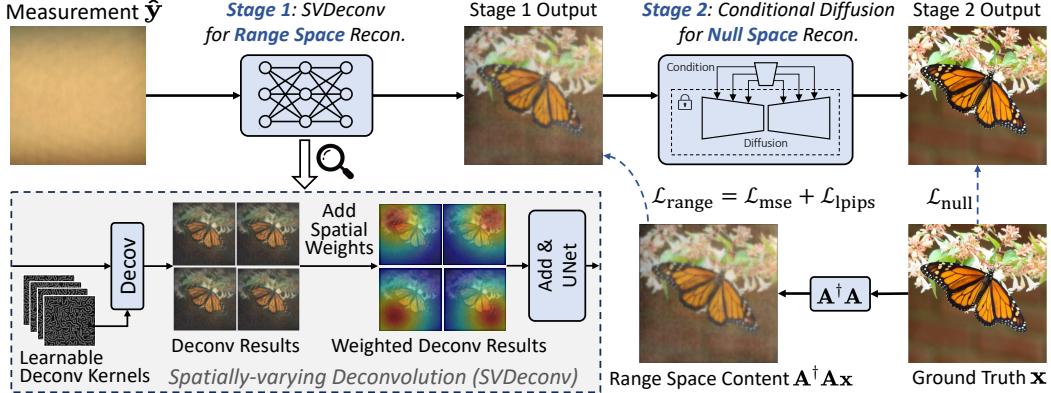


图 5：系统概览。两级管道从空间变化的反卷积网络开始，将无透镜测量映射到范围空间。然后，用于零空间恢复的条件扩散模型使用第一阶段的输出来细化细节，实现最终的重建。

为了进一步展示这种不匹配的后果，我们从干净的场景图像中模拟无透镜捕获（图 4d 顶部），并使用基于方程的精确空间变化 PSF。（2），但使用以 PSF 为中心的简单卷积模型来解决逆成像问题。图 4d 底部显示了结果，边界（红色框）比中心（绿色框）有更多的伪影。这是因为实际 PSF 和用于反卷积的单个 PSF 之间的不匹配在外围场更显着的观点。

#### 4 方法

在本节中，我们将介绍 PhoCoLens，这是一种在无透镜成像中实现真实感和一致重建的方法。如图 5 所示，PhoCoLens 由两个主要阶段组成：距离空间重建和零空间恢复。我们将首先介绍整个框架，然后对每个阶段进行详细说明。

给定输入无镜头测量  $\hat{y} = Ax + n$ ，其中  $y = Ax$  是无镜头测量的无噪声部分， $A$  是无镜头相机的传输矩阵， $x$  是我们要重建的原始场景， $n$  是传感器噪声。无透镜图像重建的目标是恢复满足  $Ax = y$  的逼真图像  $x$ 。

受 3.1 节中的范围零空间分解的启发，我们可以将任何潜在的解决方案  $x$  分解为两个正交分量：保持一致性的范围空间分量和最大化真实感的零空间分量。形式上，该分解由  $x = A^\dagger Ax + (I - A^\dagger A)x$  给出，其中  $A^\dagger Ax$  是范围空间分量， $(I - A^\dagger A)x$  是零空间分量。请注意， $x$  的任何选择都满足方程  $Ax = y$ ，因为  $A(I - A^\dagger A)x = 0$ 。

通过这种分解，我们分两个阶段重建距离空间和零空间分量，如图 5 所示。在第一阶段，我们从有噪声的无透镜测量中恢复由  $A^\dagger Ax$  表示的距离空间内容。具体来说，我们提出了 SVDeconv，一种受物理启发的空间变化反卷积网络，可以学习有效地重建范围空间内容。第二阶段重点是添加由  $(I - A^\dagger A)x$  表示的零空间内容。为此，我们引入了零空间扩散，这是一种为零空间恢复而设计的条件扩散模型。通过以第一阶段获得的距离空间重建为条件，零空间扩散通过合并零空间内容来增强这些图像，从而使它们具有更真实的外观。下面我们分别介绍一下各个阶段。

#### 4.1 Spatially-varying Deconvolution Network

在第一阶段，我们使用 SVDeconv，这是一种新颖的网络，旨在反转前向无透镜成像模型，并根据无透镜测量  $\hat{y} = Ax + n$ （无噪声  $n$ ）重建距离空间内容  $A^\dagger Ax$ 。SVDeconv 由两个主要组件组成：可微分的多核反卷积层和细化 U-Net，如图 5 左下角的灰色框所示。

传统的无透镜成像方法通常采用单个 PSF 来对成像系统进行建模，这对于大视场 (FoV) 场景来说是不准确的，如第 3.2 节中分析的那样。为了解决这个问题，SVDeconv 利用一组可学习的  $K \times K$  PSF 内核，来解释整个视场的空间变化。具体来说，我们将目标图像区域划分为  $K \times K$  段的网格，并且对于每个段，我们应用单独的 PSF 内核。该过程导致  $K \times K$  反卷积运算，产生  $K \times K$  中间反卷积图像，如图 5 所示。多核反卷积在数学上被描述为傅立叶域中的 Hadamard 乘积：

$$\mathbf{x}_{\text{de}}^{(i)} = \mathcal{F}^{-1}(\mathcal{F}(\mathbf{p}^{(i)}) \odot \mathcal{F}(\hat{\mathbf{y}})), i = 1, 2, \dots, K \times K, \quad (4)$$

其中  $\mathbf{x}_{\text{de}}^{(i)}$  是第  $i$  个可学习 PSF 核  $\mathbf{p}^{(i)}$  的第  $i$  次反卷积结果， $\mathcal{F}(\cdot)$  和  $\mathcal{F}^{-1}(\cdot)$  是 DFT 和逆 DFT 运算， $\odot$  表示 Hadamard 乘积。

从这个反卷积操作中，我们获得了  $K \times K$  中间反卷积图像。假设每幅图像准确地重建了与其 PSF 场点相对应的目标图像的特定区域，我们建议使用创新的插值方法将这些图像集成为单个统一的中间图像。插值表示为：

$$\mathbf{x}_{\text{int}}(u, v) = \sum_{i=1}^{K^2} w_i(u, v) \mathbf{x}_{\text{de}}^{(i)}(u, v), \quad (5)$$

其中  $u, v$  是坐标。每个去卷积图像  $\mathbf{x}_{\text{de}}^{(i)}$  的权重  $w_i(u, v)$  与点  $(u, v)$  与每个相应 PSF 的焦点中心之间的距离成反比，正式定义为：

$$w_i(u, v) = \frac{d_i^{-\frac{1}{2}}(u, v)}{\sum_{j=1}^{K^2} d_j^{-\frac{1}{2}}(u, v)}, \text{ with } d_i(u, v) = (u - u_i)^2 + (v - v_i)^2. \quad (6)$$

这样，它确保  $w_i(u, v)$  基于逆欧几里德距离对每个反卷积图像到每个点  $(u_i, v_i)$  的贡献进行归一化，该点表示对应于第  $i$  个可学习 PSF 核  $\mathbf{p}^{(i)}$  的 FoV 中心（我）。该加权和有效地对中间图像进行插值，以形成统一的表示，并在整个视场中增强清晰度和细节。

这种重量设计的优点之一是它显着简化了校准过程。以前的多核方法，例如[48]，通常需要在不同焦点中心进行多个校准的 PSF，导致校准过程非常耗时。相反，我们的方法只需要一个校准的 PSF 来初始化所有内核。这是因为我们预先定义了 PSF 的中心并允许模型自动学习其变化。

插值后，中间图像  $\mathbf{x}_{\text{int}}$  被输入细化 U-Net [33]，消除重建中的噪声和伪影，以近似范围空间内容  $\mathbf{A}^\dagger \mathbf{Ax}$ 。我们采用 MSE 损失和 LPIPS 损失[51]的组合来训练可学习的 PSF 和 U-Net。

#### 4.2 零空间内容恢复

通过重建的范围空间内容，我们的目标是恢复保持一致性并提高照片真实感的最终图像。为了确保一致性，最终图像和范围空间内容（残留内容）之间的差异应驻留在零空间中，这确保最终图像与原始无透镜测量对齐。为了提高照片真实感，零空间和范围空间内容的组合应该显示为逼真的现实世界图像。

为了实现这两个目标，我们提出了零空间扩散，它将重建的距离空间内容作为条件并生成符合这两个约束的图像。它确保剩余内容位于零空间中，同时保持其生成逼真图像的能力。

为了确保零空间扩散的输出满足一致性要求，我们训练模型，使这些样本的残留内容落在零空间内。从数学上讲，如果我们将输出表示为  $\mathbf{x}$ ，则当以  $\mathbf{A}^\dagger \mathbf{Ax}$  为条件时，它应满足以下等式： $\mathbf{A}(\mathbf{x} - \mathbf{A}^\dagger \mathbf{Ax}) = 0 \Rightarrow \mathbf{A}\mathbf{x} - \mathbf{A}\mathbf{Ax} = 0 \Rightarrow \mathbf{A}(\mathbf{x} - \mathbf{x}) = 0$ 。(7)

回想一下， $\mathbf{A}$  是无镜头相机的传输矩阵， $\mathbf{A}^\dagger$  是其伪逆矩阵。该方程表明将算子  $\mathbf{A}$  应用于生成的样本  $\mathbf{x}$  和原始样本  $\mathbf{x}$  之间的差值为零。这确保了剩余内容驻留在  $\mathbf{A}$  的零空间中。

基于此，我们设计零空间扩散如下。给定图像  $x$  及其相应的范围空间内容  $c = A^\dagger Ax$ ，我们逐渐向图像添加噪声，从而产生噪声图像  $x_t$ ，其中  $t$  是噪声添加迭代。与其他条件扩散模型[50]类似，我们训练网络  $\epsilon_\theta$  来预测噪声图像  $x_t$  上添加的噪声，优化目标为： $\min_{\theta} \dots$

这是为了与方程式中的目标保持一致而得出的。(7)。更多细节在附录中。此外，为了确保零空间扩散根据给定条件产生逼真的图像，我们利用预训练的扩散模型，例如稳定扩散[32]，其权重被冻结。我们专注于训练补充调节模块使用范围空间条件指导生成过程，同时保留其生成逼真图像的强大能力。具体来说，我们遵循 StableSR [40] 结构，其中涉及使用条件编码器从范围空间条件中提取多尺度特征，然后使用它们来调制扩散模型中残差块的中间特征图。这种调制可以将生成的图像与范围空间条件对齐，从而增强输出图像的保真度和真实感。

## 5 实验

在本节中，我们使用两个无镜头成像数据集来评估我们提出的方法的性能，这两个数据集是由两种不同类型的无镜头相机在现实环境中收集的：PhlatCam [5] 和 DiffuserCam [1]。PhlatCam 采用相位掩模，而 DiffuserCam 的无透镜掩模采用漫射器。我们将我们的方法与其他方法进行比较，并进行全面的消融研究以评估我们设计的有效性。

### 5.1 数据集和指标

PhlatCam 数据集 [17] 包含 1,000 个类别的 10,000 张图像，大小调整为  $384 \times 384$  像素。使用无透镜 PhlatCam [5] 显示和捕获图像，生成  $1280 \times 1480$  像素的图像。我们按照原始协议使用 990 个类进行训练，使用 10 个类进行测试。

DiffuserCam 数据集 [27] 包含使用标准镜头相机（地面实况）和基于掩模的无镜头相机 DiffuserCam [1] 同时捕获的 25,000 个配对图像。这些图像被分成 24,000 张图像用于训练，1,000 张图像用于测试。两款相机均采用原始分辨率为  $1080 \times 1920$  像素的传感器，首先将其下采样至  $270 \times 480$  像素，然后进一步裁剪至  $210 \times 380$  像素的最终分辨率，以实现正确的显示。

为了评估一致性（保真度）和真实感（视觉质量），我们采用两个指标集：

- 评估一致性的完整参考指标。PSNR、SSIM[42] 和 LPIPS[51] 这三个全参考指标用于评估网络输出与真实值之间的距离，这表明重建的保真度。
- 非参考指标用于评估照片真实感。三个非参考指标 ManIQA [45]、ClipIQA [39] 和 MUSIQ [16] 用于评估重建图像的视觉质量。

### 5.2 实施细节

对于 SVDeconv，我们利用  $3 \times 3$  PSF 内核进行反卷积。我们使用来自 DiffuserCam 和 PhlatCam 数据集的单个校准 PSF 初始化 9 个内核。SVDeconv 中的 U-Net 组件改编自 Le-ADMM-U 中用于 DiffuserCam 训练的 U-Net，以及用于 PhlatCam 的 FlatNet-gen 中的 U-Net。我们使用 Adam 优化器 [18] 将网络训练 100 个时期，批量大小为 5。两个数据集中 U-Net 训练的学习率均设置为  $3e-5$ 。对于反卷积核训练，PhlatCam 的学习率设置为  $4e-9$ ，DiffuserCam 的学习率设置为  $3e-5$ 。我们将 MSE 和 LPIPS 损失权重分别设置为 1 和 0.05。在 DiffuserCam 数据集中，测量值因传感器的有限尺寸而被裁剪，我们采用复制填充 [17]，将测量值的宽度和高度扩大两倍。

对于零空间扩散，我们使用 SVDeconv 重建的范围空间内容作为输入条件。我们按照 StableSR[39] 中使用的训练和推理设置对其进行 200 个 epoch 的训练。

在第一阶段，使用从地面实况图像导出的范围空间内容来训练 SVDeconv。SVDeconv 由两个主要参数组件组成：一个初始化的可学习反卷积内核

表1: DiffuserCam 和 PhlatCam 方法的性能比较

Dataset	PhlatCam						DiffuserCam					
Metrics	PSNR	SSIM	LPIPS↓	ManIQA	ClipIQA	MUSIQ	PSNR	SSIM	LPIPS↓	ManIQA	ClipIQA	MUSIQ
<i>Ground Truth</i>	—	—	—	0.431	0.583	63.40	—	—	—	0.160	0.333	31.21
WienerDeconv [44]	12.19	0.270	0.922	0.111	0.149	15.47	10.59	0.275	0.843	0.184	0.180	19.43
ADMM [7]	13.45	0.301	0.877	0.132	0.159	16.45	12.87	0.305	0.705	0.141	0.136	16.51
Le-ADMM-U [27]	20.12	0.515	0.405	0.138	0.180	24.64	22.35	0.668	0.253	0.110	0.185	20.16
MMCN [49]	20.39	0.524	<u>0.346</u>	0.145	0.206	27.65	24.09	0.744	<u>0.238</u>	0.121	0.183	21.47
UPDN [19]	20.48	0.533	0.352	0.158	0.215	29.32	<b>24.67</b>	<u>0.747</u>	0.256	0.139	0.178	22.56
FlatNet-gen [17]	<u>20.53</u>	<u>0.549</u>	0.375	0.203	0.287	44.94	21.43	0.696	0.254	0.134	0.244	23.15
DDNM+[41]	15.22	0.485	0.623	0.297	0.412	45.32	18.36	0.539	0.516	0.162	0.201	27.83
FlatNet+DiffBIR [23]	19.96	0.544	0.391	<u>0.335</u>	<u>0.523</u>	57.13	18.97	0.517	0.505	<b>0.312</b>	<b>0.548</b>	<b>51.20</b>
<b>PhCoLens(Ours)</b>	<b>22.07</b>	<b>0.601</b>	<b>0.215</b>	<b>0.357</b>	<b>0.565</b>	<b>62.20</b>	<b>24.12</b>	<b>0.748</b>	<b>0.161</b>	0.172	0.339	32.84

具有已知的 PSF，以及使用标准权重初始化且无需预训练的 U-Net。一旦训练完成，SVDeconv 就会处理输入的无透镜测量结果，以估计训练样本的距离空间内容。随后，我们使用这个估计的范围空间内容作为输入条件，通过零空间扩散进行微调。在扩散微调期间，我们利用具有冻结权重的预训练扩散模型。我们只训练像StableSR [35]这样的补充调节模块，以有效地指导重建过程。

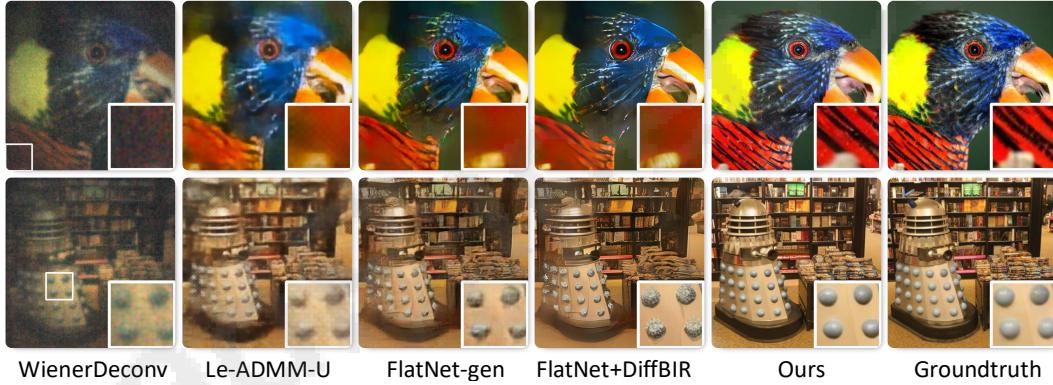


Figure 6: Qualitative comparison between our method and others on the PhlatCam dataset.

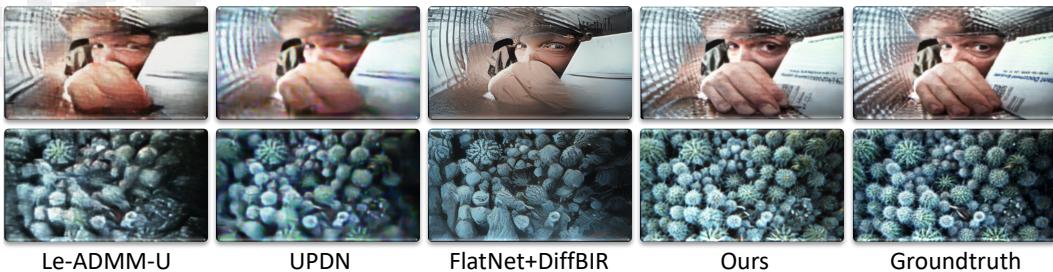


图7: 我们的方法与其他方法在 DiffuserCam 数据集上的定性比较。5.3 与其他方法的比较 et.

我们通过将我们提出的方法与传统方法和基于学习的方法进行比较，对两个数据集使用定性和定量测量来评估我们提出的方法的性能。

我们比较了传统方法，如傅里叶域中的 Tikhonov 正则化重建 (WienerDeconv [44]) 和通过 ADMM [7] 进行的全变分正则化。此外，我们还评估了基于学习的方法，例如展开网络 Le-ADMM-U [27]、MMCN [49] 和 UPDN [19] 以及前馈反卷积方法 FlatNet-gen。此外，我们评估了两种基于扩散的方法：一种使用预训练的稳定扩散[32]和零样本逆成像采样方法 DDNM+ [41]，另一种使用预训练的盲法增强 FlatNet-gen 的输出

Table 2: Comparison of deconvolution methods for range space reconstruction and original content reconstruction.

Reconstruction Target	Method	PhlatCam			DiffuserCam		
		PSNR	SSIM	LPIPS $\downarrow$	PSNR	SSIM	LPIPS $\downarrow$
Range Space	Le-ADMM-U	23.87	0.835	0.307	24.24	0.748	0.282
	SingleDeconv	25.61	0.874	0.263	24.94	0.799	0.230
	MultiWienerNet	26.43	0.880	0.256	26.32	0.823	0.189
Original Content	SVDeconv (Ours)	<b>26.97</b>	<b>0.894</b>	<b>0.253</b>	<b>28.84</b>	<b>0.856</b>	<b>0.144</b>
	Le-ADMM-U	20.12	0.515	0.405	22.35	0.668	0.253
	SingleDeconv	20.43	0.545	0.381	21.97	0.689	0.262
	MultiWienerNet	20.82	0.562	0.374	23.02	0.725	0.214
	SVDeconv (Ours)	<b>21.01</b>	<b>0.571</b>	<b>0.363</b>	<b>25.55</b>	<b>0.781</b>	<b>0.179</b>

表3: 零空间恢复方法的比较

Method	PSNR	SSIM	LPIPS $\downarrow$	ManIQA	ClipIQA	MUSIQ
StableSR [39]	14.93	0.446	0.624	0.64	0.235	33.84
DiffBIR [23]	16.21	0.432	0.502	0.353	0.512	56.65
DDNM [41]	17.62	0.539	0.517	0.148	0.201	27.82
Ours	<b>22.07</b>	<b>0.601</b>	<b>0.215</b>	<b>0.357</b>	<b>0.565</b>	<b>62.20</b>



图9: 我们的零空间扩散之间的比较和其他方法。

图像恢复模型DiffBIR [23]。定量结果总结于表中。1. PhlatCam 和 DiffuserCam 的定性比较分别如图 6 和 7 所示。

根据表 1, 在 PhlatCam 数据集上, 我们的方法在保真度 (全参考指标) 和视觉质量 (非参考指标) 方面均实现了最佳性能。1, 图 6 中的定性结果进一步支持了这一点, 其中我们的重建最接近真实情况。

对于 DiffuserCam 数据集, 我们的方法在一致性指标 (全参考指标、SSIM 和 LPIPS) 方面实现了卓越的性能。虽然 UPDN 实现了稍高的 PSNR, 但其输出的视觉质量受到伪影的影响, 如图 7 所示。我们的方法中 PSNR 的轻微下降可能是由于 DiffuserCam 地面实况本身内的伪影 (例如莫尔图案) 造成的。尽管这些伪像具有良好的视觉一致性, 但它们会以较低的 PSNR 来惩罚高质量图像。图 7 中的视觉比较进一步证明了这一点。相反, FlatNet +DiffBIR, 尽管实现了更高的质量指标 (非参考指标、ManIQA、ClipIQA) 和 MUSIQ), 无法保持数据一致性 (低全参考指标、PSNR、SSIM 和 LPIPS)。另外, 如图所示。7、FlatNet+DiffBIR 有时会严重扭曲原始图像内容, 例如将人手转移到顶行的衣服中。总的来说, 我们的方法实现了数据一致性和视觉质量之间的最佳权衡。并且由于 SVDeconv 的引入, 我们的方法可以比其他方法更好地重建视场外围。

#### 5.4 分析与讨论

空间变化反卷积的有效性。为了验证所提出的 SVDeconv 的有效性, 我们将其与其他可学习的反卷积方法进行了比较。基线包括 Le-ADMM-U (实现 ADMM 的深度展开网络)、SingleDeconv (使用单个反卷积内核) 和 MultiWienerNet [48] (采用多个反卷积内核, 无需加权插值, 与我们使用单个 PSF 的方法类似地初始化)。我们关注两种类型的内容重建: 范围空间内容重建, 其中范围空间内容是目标, 以及原始内容

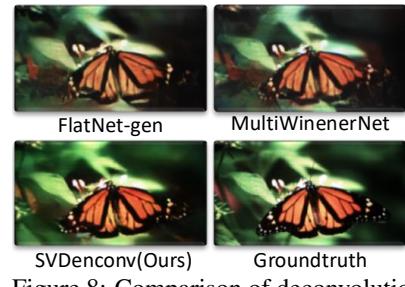


Figure 8: Comparison of deconvolution methods on a DiffuserCam example.

表4: 不同扩散条件的比较

Condition	PhlatCam			DiffuserCam		
	PSNR	SSIM	LPIPS $\downarrow$	PSNR	SSIM	LPIPS $\downarrow$
WienerDeconv	20.38	0.551	0.301	20.42	0.632	0.282
FlatNet-gen	20.89	0.556	0.286	21.67	0.647	0.230
SVD-OC	21.16	0.572	0.276	22.43	0.679	0.185
Ours	<b>22.07</b>	<b>0.601</b>	<b>0.215</b>	<b>24.12</b>	<b>0.748</b>	<b>0.161</b>

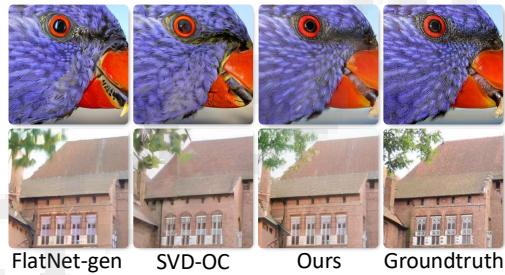


图10: 不同扩散方式的比较 conditions.

重建，使用原始的地面实况作为目标。为了公平比较，所有方法都使用相同的 U-Net 进行去噪和感知增强，并使用相同的损失进行训练。结果见表。图2显示了所提出的空间变化反卷积在范围空间内容和原始内容的忠实重建中的功效。此外，图 8 显示我们的方法优于其他方法，特别是在重建外围的精细细节（由白色虚线框圈出）方面。

零空间扩散的有效性。重建范围空间内容后，我们根据先前重建的范围空间内容执行零空间内容恢复。我们将零空间扩散模型与其他方法进行比较，包括 DiffBIR [23] 和 StableSR [39] 等条件扩散模型作为零样本逆成像方法DDNM [41]。表中的定量结果。图 3 和图 9 中的定性比较表明，我们的零空间扩散在与地面实况的一致性和视觉质量方面表现最好。

范围内容条件的有效性。我们框架的一个重要设计是使用第一阶段重建作为第二阶段扩散的条件。为了评估这些条件的有效性，我们进行了消融研究，用 WienerDeconv、FlatNet-gen 和 SVD-OC（用原始内容训练 SVDencov）的输出替换重建的范围内容，并使用这些重新训练条件扩散模型状况。从图中可以看出10，与其他人相比，我们的范围内容条件显着提高了重建一致性。这是因为使用 FlatNet-gen 等模型的输出会引入额外的伪影，阻碍第二阶段恢复原始图像。此外，表中的定量分析。4 证实了这一观察结果。我们的方法在所有保真度指标中都优于这些条件。

## 6结论与局限性

我们推出 PhoCoLens，这是一种创新的两阶段方法，可在无透镜成像中实现逼真且一致的重建。该方法利用了两种互补方法的优点：范围空间的空间变化反卷积确保了一致性，而零空间的零空间扩散保证了真实感。我们在两台无镜头相机上进行的实验证明了 PhoCoLens 在重建现实场景同时保持保真度方面的有效性。我们方法的一个局限性是 PhoCoLens 的两阶段性质和扩散模型的采样时间阻碍了其实时适用性。此外，尽管 PhoCoLens 实现了最佳保真度，但我们使用的扩散模型仍然可能会引入偏离原始场景的高频细节，特别是对于缺乏细节的平滑场景。未来的工作将集中在两个主要方面：加速扩散模型采样过程，以使用无镜头相机实现实时照片和视频捕获，以及探索 3D 空间变化点扩散函数 (PSF) 效果以捕获 3D 场景信息。

## Acknowledgment

我们感谢审稿人的宝贵意见。这项工作得到上海人工智能实验室、RGC 早期职业计划 (ECS) 编号24209224和香港中文大学直接资助 (RCFUS) 编号4055189的支持。

## References

- [1] Nick Antipa, Grace Kuo, Reinhard Heckel, Ben Mildenhall, Emrah Bostan, Ren Ng, and Laura Waller. Diffusercam: lensless single-exposure 3d imaging. *Optica*, 5(1):1–9, 2018.
- [2] Muthuvvel Arigovindan、Joshua Shaevitz、John McGowan、John W Sedat 和 David A Agard。基于主成分分析的 3D 宽场显微镜中表示深度变化点扩散函数的并行积卷积方法。光学快报，18(7): 6461–6476, 2010。
- [3] M Salman Asif、Ali Ayrem lou、Aswin Sankaranarayanan、Ashok Veeraraghavan 和 Richard GBaraniuk。Flatcam：使用编码光圈和计算的薄型无镜头相机。IEEE 计算成像交易, 3(3): 384–397, 2016。
- [4] Eric Bezzam, Martin Vetterli, and Matthieu Simeoni. Privacy-enhancing optical embeddings for lensless classification. *arXiv preprint arXiv:2211.12864*, 2022.

- [5] Vivek Boominathan、Jesse K Adams、Jacob T Robinson 和 Ashok Veeraraghavan。 Phlatcam: 设计的基于相位掩模的薄型无镜头相机。 IEEE 模式分析和机器智能汇刊, 42(7):1618–1629, 2020。
- [6] Vivek Boominathan, Jacob T Robinson, Laura Waller, and Ashok Veeraraghavan. Recent advances in lensless imaging. *Optica*, 9(1):1–16, 2022.
- [7] Stephen Boyd、Neal Parikh、Eric Chu、Borja Peleato、Jonathan Eckstein 等。通过乘法器的交替方向方法进行分布式优化和统计学习。机器学习的基础和趋势®, 3(1):1–122, 2011。
- [8] Jooyoung Choi, Sungwon Kim, Yonghyun Jeong, Youngjune Gwon, and Sungroh Yoon. Ilvr: Conditioning method for denoising diffusion probabilistic models. *arXiv preprint arXiv:2108.02938*, 2021.
- [9] Hyungjin Chung, Byeongsu Sim, Dohoon Ryu, Jong Chul Ye。使用流形约束改进反问题的扩散模型。神经信息处理系统的进展, 35: 25683–25696, 2022 年。
- [10] 约瑟夫·W·古德曼。傅里叶光学简介。罗伯茨和公司出版商, 2005 年。
- [11] Felix Heide, Mushfiqur Rouf, Matthias B Hullin, Bjorn Labitzke, Wolfgang Heidrich, and Andreas Kolb. High-quality computational imaging through simple lenses. *ACM Transactions on Graphics (ToG)*, 32(5):1–14, 2013.
- [12] 乔纳森·何 (Jonathan Ho)、阿杰·贾恩 (Ajay Jain) 和彼得·阿贝尔 (Pieter Abbeel)。去噪扩散概率模型。神经信息处理系统的进展, 33: 6840–6851, 2020。
- [13] Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022.
- [14] 华毅, 赵勇毅, Aswin C Sankaranarayanan。用于球形传感器上无透镜成像的角度敏感像素。arXiv 预印本 arXiv:2306.15953, 2023。
- [15] Bahjat Kawar, Michael Elad, Stefano Ermon 和 Jiaming Song。去噪扩散恢复模型。神经信息处理系统的进展, 35: 23593–23606, 2022。
- [16] 柯俊杰, 王启飞, 王一林, 佩曼·米兰法尔, 杨峰。Musiq: 多尺度图像质量转换器。IEEE/CVF 国际计算机视觉会议记录, 第 5148–5157 页, 2021 年。
- [17] Salman Siddique Khan, Varun Sundar, Vivek Boominathan, Ashok Veeraraghavan 和 Kaushik Mitra。Flatnet: 通过无透镜测量实现逼真的场景重建。IEEE 模式分析和机器智能交易, 44(4): 1934–1948, 2020。
- [18] 迪德里克·P·金马和吉米·巴。Adam: 一种随机优化方法。arXiv 预印本 arXiv: 1412.6980, 2014。
- [19] Oliver Kingshott, Nick Antipa, Emrah Bostan, and Kaan Akşit. Unrolled primal-dual networks for lensless cameras. *Optics Express*, 30(26):46324–46335, 2022.
- [20] Grace Kuo、Fanglin Linda Liu、Irene Grossrubatscher、Ren Ng 和 Laura Waller。带有随机微透镜扩散器的片上荧光显微镜。光学快报, 28(6):8384–8399, 2020。
- [21] 李成波, 尹窝涛, 江红, 张银。一种有效的增强拉格朗日方法及其在总变差最小化中的应用。计算优化和应用, 56: 507–530, 2013。
- [22] 李英, 李正岱, 陈凯宇, 郭友明, 饶昌辉。Mwdns: 无透镜成像的多尺度特征空间重建。光学快报, 31(23): 39088–39101, 2023。
- [23] 林新奇, 何静文, 陈子彦, 吕兆阳, 本飞, 戴波, 欧阳万里, 乔宇, 董超。Diffbir: 利用生成扩散先验实现盲图像恢复。arXiv 预印本 arXiv:2308.15070, 2023。

- [24] 罗军, 聂云峰, 任文琪, 曹晓春, 杨明轩。通过深度感知点扩散函数校正光学像差。IEEE 模式分析和机器智能交易, 2024 年。
- [25] Elie Maalouf、Bruno Colicchio 和 Alain Dieterlen。使用泽尼克矩的荧光显微镜三维深度变点扩散函数插值。乔萨 A, 28(9): 1864–1870, 2011。
- [26] 松岛恭二.用于离轴数值传播的平移角谱方法。光学快报, 18(17): 18453–18463, 2010。
- [27] 克里斯蒂娜·莫纳霍娃、约书亚·尤特塞弗、格蕾丝·郭、尼克·安提帕、凯罗洛斯·扬尼和劳拉·沃勒。学习了实用的基于掩模的无透镜成像的重建。光学快报, 27(20): 28075–28090, 2019。
- [28] Chong Mou, Xintao Wang, Liangbin Xie, Yanze Wu, Jian Zhu, Zhongang Qi, and Ying Shan. T2i-adapter: 学习适配器以挖掘文本到图像扩散模型的更多可控能力。AAAI 人工智能会议记录, 第 38 卷, 第 4296-4304 页, 2024 年。
- [29] 努尔穆罕默德·帕特瓦里和 Chrysanthe Preza。使用标准正交基础对三维荧光显微镜进行图像恢复, 以有效表示深度变点扩散函数。生物医学光学快报, 6 (10) : 3826–3841, 2015。
- [30] 肯·佩林。改善噪音。第 29 届计算机图形学和交互技术年会论文集, 第 681-682 页, 2002 年。
- [31] Dikpal Reddy、Ashok Veeraraghavan 和 Rama Chellappa。P2c2: 用于高速成像的可编程像素压缩相机。IEEE/CVF 计算机视觉和模式识别会议记录, 第 329-336 页。IEEE, 2011。
- [32] 罗宾·隆巴赫、安德烈亚斯·布拉特曼、多米尼克·洛伦茨、帕特里克·埃瑟和比约恩·奥默。使用潜在扩散模型进行高分辨率图像合成。IEEE/CVF 计算机视觉和模式识别会议记录, 第 10684-10695 页, 2022 年。
- [33] 奥拉夫·罗纳伯格、菲利普·费舍尔和托马斯·布洛克斯。U-net: 用于生物医学图像分割的卷积网络。医学图像计算和计算机辅助干预 - MICCAI 2015: 第 18 届国际会议, 德国慕尼黑, 2015 年 10 月 5 日至 9 日, 会议记录, 第 III 部分 18, 第 234-241 页。施普林格, 2015。
- [34] Chitwan Saharia, William Chan, Huiwen Chang, Chris Lee, Jonathan Ho, Tim Salimans, David Fleet, and Mohammad Norouzi. Palette: Image-to-image diffusion models. In *ACM SIGGRAPH 2022 conference proceedings*, pages 1–10, 2022.
- [35] Chitwan Saharia、Jonathan Ho、William Chan、Tim Salimans、David J Fleet 和 Mohammad Norouzi。通过迭代细化实现图像超分辨率。IEEE 模式分析和机器智能汇刊, 45(4):4713–4726, 2022 年。
- [36] 约翰内斯·施瓦布、斯蒂芬·安霍尔泽和马库斯·哈尔特迈尔。逆问题的深度零空间学习: 收敛分析和速率。反问题, 35 (2) : 025008, 2019。
- [37] 杨松, 斯特凡诺·埃尔蒙。通过估计数据分布的梯度进行生成建模。神经信息处理系统的进展, 2019 年 32 月。
- [38] 杨松, 沉丽月, 雷星, Stefano Ermon。使用基于评分的生成模型解决医学成像中的逆问题。arXiv 预印本 arXiv:2111.08005, 2021。
- [39] 王建一, Kelvin CK Chan, 和 Chen Change Loy。探索剪辑以评估图像的外观和感觉。AAAI 人工智能会议记录, 第 37 卷, 第 2555-2563 页, 2023 年。
- [40] Jianyi Wang, Zongsheng Yue, Shangchen Zhou, Kelvin CK Chan, and Chen Change Loy. Exploiting diffusion prior for real-world image super-resolution. *arXiv preprint arXiv:2305.07015*, 2023.

- [41] 王银怀, 余继文, 张健。使用去噪扩散零空间模型的零样本图像恢复。国际学习表征会议, 2022 年。
- [42] 周旺, 艾伦·C·博维克, 哈米德·R·谢赫, 埃罗·P·西蒙切利。图像质量评估: 从错误可见性到结构相似性。 IEEE 图像处理交易, 13(4):600–612, 2004。
- [43] 魏浩宇, 刘鑫, 郝翔, Edmund Y Lam, 彭一凡。使用最小采样角谱方法对离轴衍射进行建模。光学, 10 (7) : 959–962, 2023。
- [44] 诺伯特·维纳。平稳时间序列的外推、内插和平滑: 工程应用。麻省理工学院出版社, 1949 年。
- [45] 杨思迪, 吴天河, 石树伟, 劳珊珊, 龚原, 曹明灯, 王家豪, 杨玉久。Maniqa: 用于无参考图像质量评估的多维注意力网络。 IEEE/CVF 计算机视觉和模式识别会议论文集, 第 1191-1200 页, 2022 年。
- [46] 杨欣格, 付强, Mohamed Elhoseiny, Wolfgang Heidrich。像差感知景深。 IEEE 模式分析和机器智能汇刊, 2023 年。
- [47] Kyrollos Yanny, Nick Antipa, William Liberti, Sam Dehaeck, Kristina Monakhova, Fan-glin Linda Liu, Konlin Shen, Ren Ng, and Laura Waller. Miniscope3d: optimized single-shot miniature 3d fluorescence microscopy. *Light: Science & Applications*, 9(1):171, 2020.
- [48] Kyrollos Yanny, Kristina Monakhova, Richard W Shuai, and Laura Waller. Deep learning for fast spatially varying deconvolution. *Optica*, 9(1):96–99, 2022.
- [49] 曾天骄, Edmund Y Lam。通过深度学习进行稳健重建, 以处理无透镜成像中的模型失配。 IEEE 计算成像学报, 7: 1080–1092, 2021。
- [50] 张吕敏, 饶安仪, Maneesh Agrawala。向文本到图像扩散模型添加条件控制。 IEEE/CVF 国际计算机视觉会议论文集, 第 3836-3847 页, 2023 年。
- [51] Richard 张、Phillip Isola、Alexei A Efros、Eli Shechtman 和 Oliver Wang。深度特征作为感知度量的不合理有效性。 IEEE 计算机视觉和模式识别会议论文集, 第 586-595 页, 2018 年。
- [52] 张文辉, 张浩, 松岛恭二, 金国凡。用于离轴衍射计算的移带扩展角谱方法。光学快报, 29 (7) : 10089–10103, 2021。

## Appendix

### A PhlatCam Simluation

在 3.2 节中，我们模拟 PhlatCam 来解释卷积无透镜成像模型中不匹配的细节。在这里我们解释如何模拟像 PhlatCam 这样的无镜头相机。

继之前的工作[5]之后，我们的模拟首先基于随机柏林噪声生成的轮廓构建所需的点扩散函数（PSF）[30]。给定掩模和传感器之间的特定距离，我们优化相位掩模以产生与目标匹配的PSF。对于此优化，我们采用近场相位检索算法（NfPR）[5]。该迭代算法在掩模平面和传感器平面处的场之间交替，在每个位置强制执行特定约束：掩模平面处的场幅度设置为统一，并且传感器平面处的场强度符合目标PSF。该算法利用前向菲涅尔传播来计算从掩模到传感器平面的场，并应用后向菲涅尔传播来计算从传感器返回到掩模的场。相位掩模优化所需的输入包括目标 PSF、掩模到传感器的距离（在我们的设置中为 1mm）以及光波长（在本例中设置为 532 nm，这是中可见光谱中的典型波长）。

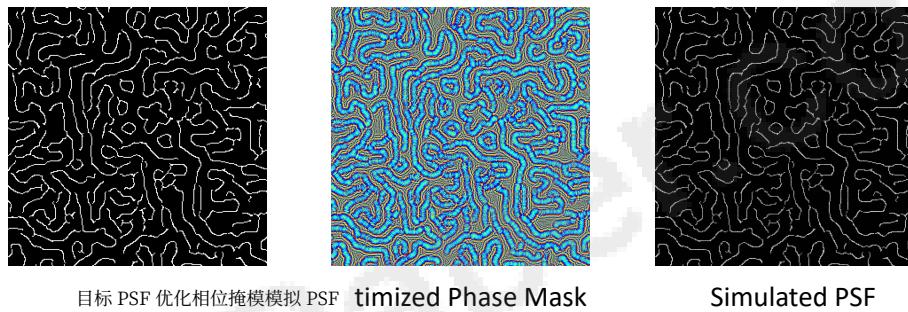


图 A1：范围空间内容模拟结果的可视化示例。

图 A1 说明了目标 PSF、优化的相位掩模以及将菲涅尔传播方法应用于设计的相位掩模所产生的模拟 PSF。给定所需的相位掩模，我们可以准确地模拟各种情况下的无透镜成像过程。

#### B范围空间内容近似

在我们方法的第一阶段（第 4.1 节），我们希望从无透镜测量  $\hat{y}$  重建距离空间投影  $A^\dagger Ax$ 。这种重建需要已知的目标  $A^\dagger Ax$  对应于给定的图像  $x$ 。然而，对于现实世界的无透镜成像系统，直接获得传递矩阵  $A \in \mathbb{R}^{N/2 \times M/2}$  由于其规模相当大，在计算上是不可行的。因此，我们采用  $A$  的近似形式来模拟现实的无透镜成像过程，并仿真其伪逆  $A^\dagger$  通过反转该近似值。

在校准单点扩散函数 (PSF)  $h$  的情况下，我们将前向成像过程近似如下： $y_{sim} = C(h * x) + n, (A1)$

其中  $C$  表示由于传感器尺寸有限而进行的裁剪操作， $*$  表示卷积算子， $n$  表示传感器和量化噪声。在我们的实际实现中，模拟的  $y_{sim}$  被量化为 12 位，并引入传感器噪声以实现 30 的信噪比 (SNR)。

$A^\dagger$  的近似值是通过反转前向过程得出的，特别是通过应用重复填充来抵消裁剪并采用维纳反卷积来反转卷积： $x_{range} = D(P(y_{sim}), h), (A2)$

其中  $P$  代表重复填充,  $D$  代表维纳反卷积。

如果在不同视场 (FoV) 上有多个校准的 PSF, 则通过采用低秩模型来模拟空间变化的卷积, 可以更好地近似成像前向过程。相应地, 可以利用 4.1 节中讨论的空间变化反卷积模型来模拟  $A^\dagger$ 。

在我们的 PhlatCam 和 DiffuserCam 设置中, 我们只有一个校准的 PSF。因此, 我们使用简单的卷积构建图像  $x$  的范围空间内容。值得注意的是, 即使没有多个校准的 PSF 来进行更准确的模拟, 生成的距离空间内容仍然是真实对应物的良好近似值。如第 5.4 节所示, 与第一阶段的原始图像内容相比, 我们的模拟仍然具有更好的监督作用。



图 A2: 范围空间内容模拟结果的可视化示例。



图 A3: 通过我们的方法估计的范围空间内容的示例。

图 A2 显示了 PhlatCam 的原始地面实况  $x$  和相应的模拟范围空间内容  $A^\dagger Ax$  的视觉示例。结果表明, 空间内容的范围  $A^\dagger Ax$  保持了原始场景的低频结构, 并在无透镜成像过程中丢失了一些高频细节。

图 A3 显示了通过我们的方法估计的范围空间内容的可视化示例。与 FlatNet 的输出相比，SVDeconv 的输出表现出与真实情况的高度视觉一致性，如绿色框中突出显示的那样。FlatNet 改变了原始场景的内容并引入了不正确的高频细节。相反，我们的方法仅保留范围空间中的低频内容，与原始地面实况紧密对齐。

#### 零空间扩散模型的 C 训练目标

在第 6 节中，我们确定方程中阐明的目标。(7) 直接对应于方程中描述的优化目标。(8) 用于扩散训练。我们提供了详细的推导来支持这种等价性。

扩散模型[32]可以通过由两个序列( $\alpha_t$ ) $Tt=1$  和( $\sigma_t$ ) $Tt=1$  定义的信噪比来表征。给定数据样本 $\mathbf{x}_0$ ，前向扩散过程 $q$ 可表示为： $q(\mathbf{x}_t | \mathbf{x}_0) = N(\mathbf{x}_t | \alpha_t \mathbf{x}_0, \sigma_t^2 I)$ , (A3)

其中  $N$  表示正态分布。

扩散的马尔可夫性质给出为：

$$q(\mathbf{x}_t | \mathbf{x}_s) = N(\mathbf{x}_t | \alpha_{t|s} \mathbf{x}_s, \sigma_{t|s}^2 \mathbb{I}), \quad (\text{A4})$$

for  $s < t$ , with:

$$\alpha_{t|s} = \frac{\alpha_t}{\alpha_s} \quad \text{and} \quad \sigma_{t|s}^2 = \sigma_t^2 - \alpha_{t|s}^2 \sigma_s^2. \quad (\text{A5})$$

给定  $\mathbf{x}_t = \alpha_t \mathbf{x}_0 + \sigma_t \epsilon$  和范围空间条件  $\mathbf{c} = \mathbf{A}^\dagger \mathbf{A} \mathbf{x}$ ，扩散输出  $\mathbf{x}$  可以通过  $\mathbf{x} = \mathbf{x}_0 \theta(\mathbf{x}_t, \mathbf{c}, t)$  建模。确保重建一致性，使得方程中  $\mathbf{A}(\mathbf{x} - \mathbf{x}_0) = 0$ 。 (7)、优化目标定义为：

$$\min_{\theta} \mathbb{E}_{\mathbf{x}_0, t, \mathbf{c}, \epsilon \sim N(0, 1)} [\|\mathbf{A}(\mathbf{x}_\theta(\mathbf{x}_t, \mathbf{c}, t) - \mathbf{x}_0)\|_2^2]. \quad (\text{A6})$$

继之前的工作[12]之后，我们采用重建项的重新参数化作为去噪目标：

$$\epsilon_\theta(\mathbf{x}_t, \mathbf{c}, t) = \frac{\mathbf{x}_t - \alpha_t \mathbf{x}_\theta(\mathbf{x}_t, \mathbf{c}, t)}{\sigma_t}, \quad (\text{A7})$$

最小化：  $\min$

$$\min_{\theta} \mathcal{L}_{\text{null}} = \mathbb{E}_{\mathbf{x}_0, t, \mathbf{c}, \epsilon \sim N(0, 1)} [\|\mathbf{A}(\epsilon_\theta(\mathbf{x}_t, \mathbf{c}, t) - \epsilon)\|_2^2], \quad (\text{A8})$$

这正是等式。(8)。

由于与模拟传递矩阵  $\mathbf{A}$  相关的大量计算成本，在实践中可以采用替代优化目标。具体来说，我们可以近似方程式中描述的目标。(8) 使用以下优化公式：

$$\min_{\theta} \mathcal{L}_{\text{null}_{\text{app}}} = \mathbb{E}_{\mathbf{x}_0, t, \mathbf{c}, \epsilon \sim N(0, 1)} [\|(\epsilon_\theta(\mathbf{x}_t, \mathbf{c}, t) - \epsilon)\|_2^2]. \quad (\text{A9})$$

#### D 定性结果

我们在图 2 中提供了额外的定性结果，证明了我们的方法的优越性能。A4 为 PhlatCam 数据集，图 A5 为 DiffuserCam 数据集。

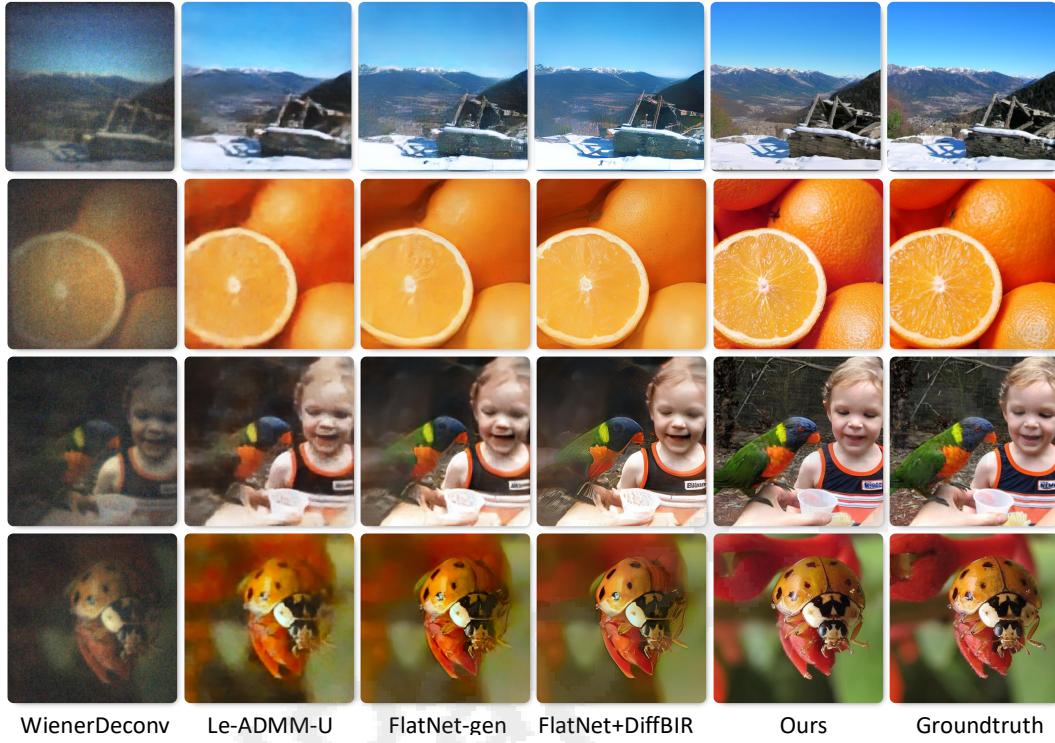


图 A4: 我们的方法与 PhlatCam 数据集上的其他方法之间的定性比较。

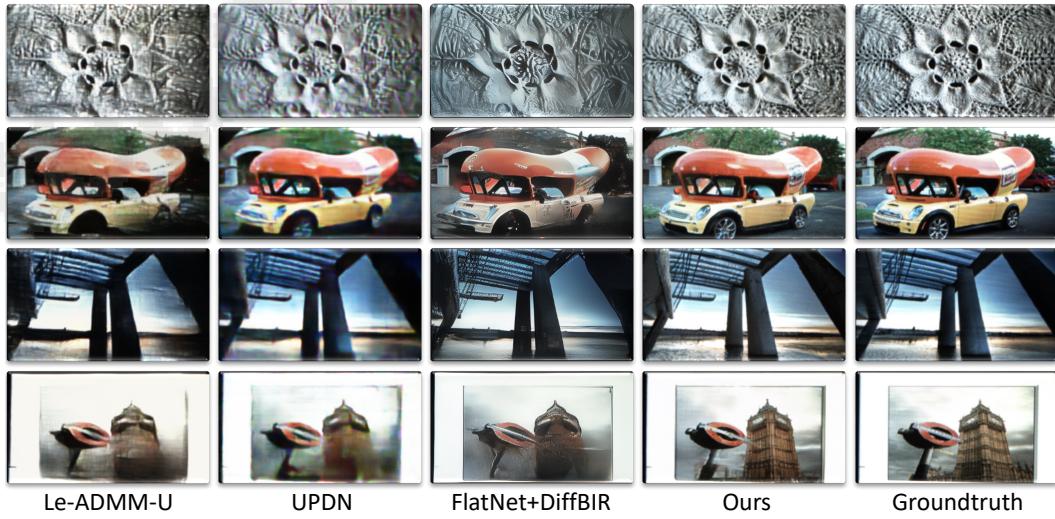


图 A5: 我们的方法与其他方法在 DiffuserCam 数据集上的定性比较。

## E 更广泛的影响

无镜头成像算法代表了相机技术的重大进步，为社会带来了光明的好处和潜在的缺点。从积极的一面来看，这些算法有助于开发更小、更轻、更节能的相机。这可以极大地扩展它们在各个领域的适用性，特别是在成像设备的小型化至关重要的医疗环境中。例如，微型无镜头相机可以集成到内窥镜工具中，以实现侵入性较小的手术，从而改善患者的治疗效果并缩短恢复时间。此外，这种紧凑型相机可用于可穿戴技术，以持续监测健康状况，为患者和医疗保健提供者提供实时数据。

然而，小型无镜头相机的普及也带来了潜在的负面社会影响，特别是在隐私方面。这些隐蔽的摄像机可以轻松地融入日常环境，这增加了滥用的可能性，例如未经授权的监视或秘密摄影。这种担忧在个人期望高度隐私的私人空间中尤其严重。因此，虽然无镜头成像技术有望在医疗保健等领域取得重大进步，但它也需要严格的法规和道德准则来防止侵入性使用并保护个人隐私权。

1. 权利要求问题：摘要和引言中提出的主要权利要求是否准确反映了论文的贡献和范围？答案：[是]理由：摘要和引言准确地反映了论文的贡献和范围。指导原则：  
 · 答案 NA 表示该论文的贡献和范围。  
 · 摘要和引言不包括论文中提出的主张。  
 · 摘要和/或引言应清楚地说明所提出的主张，包括论文中所做的贡献以及重要的假设和限制。  
 对这个问题的“否”或“不适用”答案不会被审稿人很好地理解。  
 · 所提出的主张应与理论和实验结果相匹配，并反映可以在多大程度上推广到其他环境的结果。  
 · 可以将理想目标包括为只要明确论文没有达到这些目标，就可以激发动机。

2. 限制问题：论文是否讨论了作者所做工作的局限性？答案：[是]理由：限制在第 6 节中讨论。指南：  
 · 答案 NA 表示论文没有限制，而答案 No 表示论文没有限制有局限性，但论文中没有讨论这些局限性。  
 · 鼓励作者在论文中创建单独的“局限性”部分。  
 · 论文应指出任何强有力假设以及结果对违反这些假设的稳健性有多大（例如，独立性假设、无噪声设置、模型良好规范、仅在局部保持的渐近近似）。  
 作者应该反思这些假设在实践中可能如何被违反以及会产生什么影响。  
 作者应该反思所提出的主张的范围，例如，如果该方法仅在几个数据集或几次运行中进行了测试。一般来说，实证结果通常依赖于隐含的假设，这些假设应该被阐明。  
 · 作者应该反思影响该方法性能的因素。例如，当图像分辨率较低或在低光照条件下拍摄图像时，面部识别算法可能表现不佳。或者，语音转文本系统可能无法可靠地为在线讲座提供隐藏式字幕，因为它无法处理技术术语。

作者应讨论所提出算法的计算效率以及它们如何随数据集大小进行扩展。  
 · 如果适用，作者应该讨论他们解决隐私和公平问题的方法可能存在的局限性。  
 · 虽然作者可能担心审稿人可能会利用完全诚实的限制作为拒绝的理由，但更糟糕的结果可能是审稿人发现了未被承认的局限性在报纸上。作者应该运用他们的最佳判断，并认识到有利于透明度的个人行动在制定维护社区完整性的规范方面发挥着重要作用。将特别指示审稿人不要惩罚有关限制的诚实行为。

3. 理论假设和证明问题：对于每个理论结果，论文是否提供了全套假设和完整（且正确）的证明？答案：[NA]理由：论文不包括理论结果。指南：

- 答案NA意味着论文不包括理论结果。
- 论文中的所有定理、公式和证明都应编号并交叉引用。
- 所有假设都应在任何定理的陈述中清楚地陈述或引用。
- 证明可以出现在主论文或补充材料中，但如果出现在补充材料中，则鼓励作者提供简短的证明草图以提供直觉。
- 相反，核心中提供的任何非正式证明论文的内容应辅以附录或补充材料中提供的正式证明。
- 应正确引用证明所依赖的定理和引理。

#### 4. 实验结果重现性

问题：论文是否充分披露了重现论文主要实验结果所需的所有信息，以影响论文的主要主张和/或结论（无论是否提供代码和数据）？

Answer: [Yes]

理由：有关重现实验结果的详细信息，请参阅第5节。

Guidelines:

- 答案NA表示论文不包含实验。
- 如果论文包含实验，审稿人不会很好地理解此问题的“否”答案：使论文可重现非常重要，无论是否提供了代码和数据或不是。
- 如果贡献是数据集和/或模型，作者应描述为使其结果可重复或可验证而采取的步骤。
- 根据贡献，可通过多种方式实现可重复性。例如，如果贡献是一种新颖的架构，完全描述该架构可能就足够了，或者如果贡献是一个特定的模型和经验评估，则可能有必要使其他人能够使用相同的数据集复制该模型，或者提供对模型。一般来说。发布代码和数据通常是实现此目的的一种好方法，但是也可以通过如何复制结果、访问托管模型（例如，在大型语言模型的情况下）、发布模型检查点的详细说明来提供可重复性，或适合所进行研究的其他方式。
- 虽然NeurIPS不要求发布代码，但会议确实要求所有提交内容提供一些合理的再现性途径，这可能取决于贡献的性质。例如(a)如果贡献主要是一个新算法，论文应该明确如何重现该算法。
- (b)如果贡献主要是一个新的模型架构，论文应该清楚、完整地描述该架构。
- (c)如果贡献的是一个新模型（例如，一个大型语言模型），那么应该有一种方法来访问该模型以重现结果，或者一种重现模型的方法（例如，使用开源数据集或如何进行操作的说明）来构建(d)我们认识到，在某些情况下，再现性可能很棘手，在这种情况下，欢迎作者描述他们提供再现性的特定方式。在闭源模型的情况下，对模型的访问可能会受到某种方式的限制（例如，注册用户），但其他研究人员应该有可能有某种途径来复制或验证结果。

#### 5. 开放数据和代码的访问

问题：论文是否提供对数据和代码的开放访问，并提供足够的说明来忠实地重现主要实验结果，如补充材料中所述？

Answer: [Yes]

理由：代码和数据将在项目网站上提供。

Guidelines:

- 答案NA表示论文不包含需要代码的实验。

· 请参阅 NeurIPS 代码和数据提交指南 (<https://nips.cc/public/guides/CodeSubmissionPolicy>) 了解更多详细信息。 · 虽然我们鼓励发布代码和数据，但我们知道这可能是不可能的，因此“不”是一个可以接受的答案。论文不能仅仅因为不包含代码而被拒绝，除非这是贡献的核心（例如，对于新的开源基准）。 · 说明应包含运行以重现结果所需的确切命令和环境。有关更多详细信息，请参阅 NeurIPS 代码和数据提交指南 (<https://nips.cc/public/guides/CodeSubmissionPolicy>)。 · 作者应提供有关数据访问和准备的说明，包括如何访问原始数据、预处理数据、中间数据数据和生成的数据等。 · 作者应提供脚本来重现新提出的方法和基线的所有实验结果。如果只有一部分实验是可重复的，他们应该说明脚本中省略了哪些实验以及原因。 · 在提交时，为了保持匿名，作者应发布匿名版本（如果适用）。 · 在补充材料中提供尽可能多的信息（附在论文中）是推荐的，但允许包含数据和代码的 URL。

6. 实验设置/细节问题：论文是否指定了理解结果所需的所有训练和测试细节（例如，数据分割、超参数、如何选择它们、优化器的类型等）？答案：[是]理由：实施实验的详细信息在第 5.2 节中提供。指导原则： · 答案 NA 表示论文不包含实验。 · 实验设置应在论文的核心部分详细介绍，以符合以下要求：是欣赏结果并理解它们的意义所必需的。 · 完整的详细信息可以与代码、附录或补充材料一起提供。

7. 实验统计显着性问题：论文是否适当且正确地报告了误差线或有关实验统计显着性的其他适当信息？答案：[否]理由：我们使用无透镜成像研究中的标准惯例报告实验结果，与之前的研究一致。指南： · 答案 NA 表示论文不包含实验。 · 如果结果附有误差线、置信区间或统计显着性检验（至少对于以下实验），作者应回答“是”支持论文的主要主张。 · 应明确说明误差线捕获的变异因素（例如，训练/测试分割、初始化、随机绘制某些参数或在给定实验条件下的总体运行）。 · 方法应解释计算误差线的方法（封闭式公式、调用库函数、引导程序等）。应给出所做的假设（例如，正态分布误差）。应清楚误差线是否为标准偏差或标准均值误差。 · 报告 1-sigma 误差线是可以的，但应该说明这一点。如果误差正态性假设未得到验证，作者最好报告 2-sigma 误差条，而不是声明其具有 96% CI。 · 对于非对称分布，作者应注意不要在表格或图中显示对称误差条：会产生超出范围的结果（例如负错误率）。 · 如果在表格或图中报告了误差线，作者应在文本中解释它们是如何计算的，并在文本中引用相应的图形或表格。

#### 8. 实验计算资源

问题：对于每个实验，论文是否提供了重现实验所需的计算机资源（计算工作者的类型、内存、执行时间）的足够信息？

Answer: [Yes]

理由：使用配备 8 个 NVIDIA RTX 3090 GPU、一个 64 核 CPU、256GB RAM 和 1TB 存储的服务器，Evrey 实验可以在两天内完成。

Guidelines:

- 答案 NA 表示论文不包含实验。
- 论文应指出计算工作线程 CPU 或 GPU、内部集群或云提供商的类型，包括相关内存和存储。
- 论文应提供以下内容所需的计算量：每个单独的实验运行以及估计总计算量。论文应披露整个研究项目是否需要比论文中报告的实验更多的计算量（例如，未纳入论文的初步或失败的实验）。

#### 9. 道德准则

问题：论文中进行的研究是否在各个方面都符合 NeurIPS 道德规范 <https://neurips.cc/public/EthicsGuidelines>？

Answer: [Yes]

理由：我们严格遵守 NeurIPS 道德准则。

Guidelines:

- 答案 NA 表示作者尚未审查 NeurIPS 道德准则。
- 如果作者回答否，他们应解释需要偏离道德准则的特殊情况。
- 作者应确保保持匿名（例如，如果由于其管辖范围内的法律或法规有特殊考虑）。

#### 10. 更广泛的影响

问题：本文是否讨论了所开展工作的潜在积极社会影响和消极社会影响？

Answer: [Yes]

理由：附录 E 解释了我们工作的更广泛影响。

Guidelines:

- 答案“不适用”意味着所做的工作没有社会影响。
- 如果作者回答“不适用”或“否”，他们应该解释为什么他们的工作没有社会影响或为什么论文没有解决社会影响。
- 负面社会影响的示例包括潜在的恶意或非故意用途（例如，虚假信息、生成虚假个人资料、监视）、公平性考虑（例如，部署可能做出不公平地影响特定群体的决策的技术），隐私考虑和安全考虑。
- 会议预计许多论文将是基础研究，不与特定应用程序相关，更不用说部署了。然而，如果存在任何负面应用的直接路径，作者应该指出。例如，可以合理地指出，生成模型质量的改进可用于生成虚假信息的深度伪造品。另一方面，无需指出用于优化神经网络的通用算法可以使人们能够更快地训练生成 Deepfakes 的模型。
- 作者应该考虑当该技术按预期使用并正常运行时可能出现的危害，当技术按预期使用但给出不正确的结果时可能出现的危害，以及（有意或无意）滥用技术而造成的危害。

· 如果存在负面影响，作者还可以讨论可能的缓解策略  
(例如，模型的门控发布、除了攻击之外还提供防御、监控滥用的机制、监控系统如何从反馈中学习的机制、提高效率) 和机器学习的可访问性) .11。保障措施问题：本文是否描述了为负责任地发布滥用风险较高的数据或模型(例如，预训练语言模型、图像生成器或抓取的数据集)而采取的保障措施？答案：[不适用]理由：我们会不发布误用高风险的数据或模型。指南：  
· 答案 NA 表示论文不构成此类风险。  
· 发布误用或双重用途高风险的模型发布时应附带必要的保障措施，以允许模型的受控使用，例如要求用户遵守使用指南或访问模型的限制或实施安全过滤器。  
· 从互联网上抓取的数据集可能会带来安全风险。作者应该描述他们如何避免发布不安全的图像。  
· 我们认识到提供有效的保障措施具有挑战性，许多论文并不要求这样做，但我们鼓励作者考虑到这一点并尽最大努力。12。现有资产的许可问题：本文中使用的资产(例如代码、数据、模型)的创建者或原始所有者是否得到了适当的认可，许可和使用条款是否明确提及并得到适当尊重？答案：[是]理由：我们引用所有原始论文并确保我们的使用是合法的。指南：  
· 答案 NA 意味着该论文不使用现有资产。  
· 作者应引用生成代码包或数据集的原始论文。  
· 作者应说明使用的资产版本，如果可能，请包含 URL。  
· 每个资产应包含许可证名称(例如 CC-BY 4.0)。  
· 对于从特定来源(例如，CC-BY 4.0)抓取的数据、网站)，应提供该来源的版权和服务条款。  
· 如果发布资产，应提供包中的许可证、版权信息和使用条款。对于流行的数据集，paperswithcode.com/datasets 已为某些数据集策划了许可证。他们的许可指南可以帮助确定数据集的许可。  
· 对于重新打包的现有数据集，应提供原始许可和衍生资产的许可(如果已更改)。  
· 如果无法在线获取此信息，鼓励作者联系资产的创建者。13。新资产问题：本文中引入的新资产是否有详细记录，并且文档是否与资产一起提供？答案：[NA]理由：我们不引入新资产。指导原则：  
· 答案 NA 表示本文未发布新资产。  
· 研究人员应通过结构化模板传达数据集/代码/模型的详细信息，作为其提交任务的一部分。这包括有关培训、许可、限制等的详细信息。  
· 文件应讨论是否以及如何获得资产被使用者的同意。

· 提交时，请记住对您的资产进行匿名化（如果适用）。您可以创建匿名 URL 或包含匿名 zip 文件。14.众包和人类受试者研究问题：对于众包实验和人类受试者研究，论文是否包括向参与者提供的说明全文和屏幕截图（如果适用）以及有关补偿的详细信息（如果有）？答案：[不适用]理由：我们的工作不涉及众包或人类受试者研究。指南：· 答案 NA 表示该论文不涉及众包或人类受试者研究。· 在补充材料中包含此信息很好，但是如果论文的主要贡献涉及人类受试者，那么主论文中应包含尽可能多的细节。· 根据 NeurIPS 道德准则，参与数据收集、管理或其他劳动的工人应至少获得最低工资在数据收集者的国家。15。人体受试者研究的机构审查委员会 (IRB) 批准或同等批准

问题：论文是否描述了研究参与者产生的潜在风险，是否向受试者披露了此类风险，以及是否获得了机构审查委员会 (IRB) 批准（或根据您所在国家或机构的要求进行同等批准/审查）？

答案：[NA]理由：我们的工作不涉及人类受试者的研究。指南：· 答案 NA 表示论文不涉及众包或人类受试者的研究。· 根据进行研究的国家/地区，IRB 批准任何人类受试者研究可能需要（或同等内容）。如果您获得了 IRB 批准，则应在论文中明确说明这一点。· 我们认识到，不同机构和地点的程序可能存在很大差异，我们希望作者遵守 NeurIPS 道德准则及其机构指南。· 对于初始提交的材料，请勿包含任何会破坏匿名性的信息（如果适用），例如进行审查的机构。