arXiv:2601.12257v1 [cs.CV] 18 Jan 2026

软阴影扩散（SSD）：受物理启发的三维计算潜望成像学习

FadlullahRaji 和 John Murray Bruce

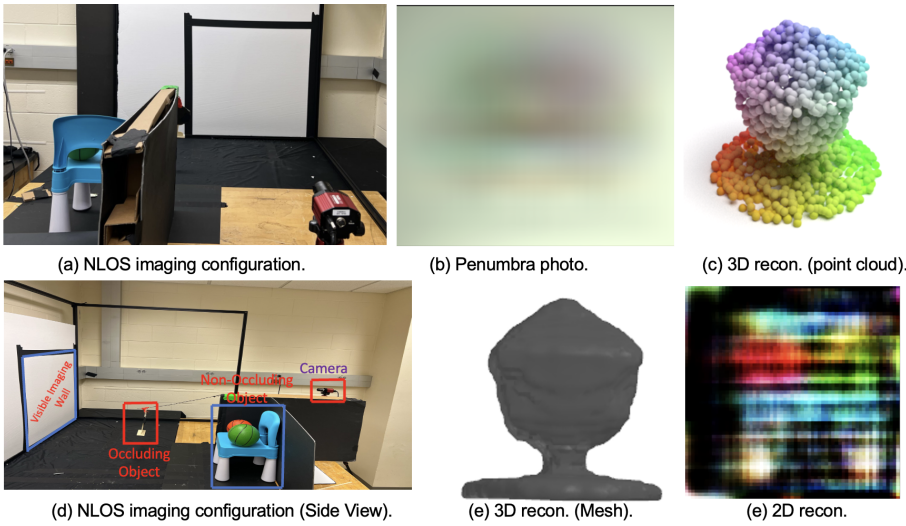
南佛罗里达大学计算机科学与工程系，4202 E. Fowler Avenue，坦帕，佛罗里达州，美国 33620。{fraji, murraybruce}@usf.edu

摘要。 传统成像需要视线才能准确呈现场景。然而，在某些情况下，获得合适的视线可能不切实际、危险，甚至不可能。非视线（NLOS）成像通过根据间接测量数据重建场景来应对这一挑战。近年来，利用隐藏场景投射到可见墙面上的微弱阴影的普通照片进行成像的被动非视线方法备受关注。这些方法目前仅限于一维或低分辨率二维彩色成像，或定位形状大致已知的隐藏物体。在此，我们推广了这类方法，并展示了如何从一张普通的非视线照片中重建隐藏场景的三维结构。为此，我们提出一种新颖的光传输模型重构方式，将隐藏场景便捷地分解为遮光 和非遮光 组件，从而得到一个可分离非线性最小二乘法（SNLLS）逆问题。我们提出两种解决方案：一种基于梯度的优化方法，以及一种受物理启发的神经网络方法，称为软阴影扩散（SSD）。尽管这里遇到的逆问题具有挑战性且条件不适定，我们的方法在真实实验场景中的众多三维场景上均表现有效。此外，SSD 虽在仿真中训练，却能很好地泛化到仿真中未见过的类别和真实世界的非视线场景。SSD 对噪声和环境照明也表现出出人意料的鲁棒性。

关键词： 计算成像 · 机器学习 · 三维生成模型 · 扩散模型 · 可分离非线性最小二乘

1 引言

自主导航中的避碰系统、军事行动中对隐藏敌人的探测、搜救行动中的危险评估，以及生物医学成像中的动脉阻塞评估，都是能够对观察者视线之外的场景成像后将受益的应用。由于潜在应用范围广泛，这一领域——非视域（NLOS）成像——持续快速发展。其根本目标是根据



(a) NLOS imaging configuration. (b) Penumbra photo. (c) 3D recon. (point cloud).

(d) NLOS imaging configuration (Side View). (e) 3D recon. (Mesh). (e) 2D recon.

图 1：非视域成像配置。非遮挡物体将光漫反射至可见墙面；反射光被隐藏场景中的遮挡物部分遮挡，从而在可见墙面上形成软阴影。

对信号的测量数据，例如光 [16, 27, 54, 55]，声音 [4, 30]，热量 [24, 39]，或其他模态 [1]，在非视域场景发生漫反射或发射后到达观察者，从而推断有关非视域场景的信息。

在本研究中，我们展示了如何根据到达观察者可见表面以及隐藏场景的光的测量数据，实现三维非视域成像。类似地依赖光测量数据的既有方法分为两类：主动方法，利用对隐藏场景的受控照明 [16, 18, 27, 31, 34–36, 41, 44, 45, 50, 56, 62, 65]；以及被动方法，利用已有光照，而不对隐藏场景的照明进行任何控制 [3, 5, 14, 17, 42, 55, 57, 58, 68, 73]。主动方法通过采用超快脉冲激光照明和时间分辨单光子探测，在实现高分辨率三维非视域成像方面取得了显著成果 [44, 45, 65]，。尽管取得了这些进展，主动方法所需设备的复杂性和成本仍可能成为多种应用的限制因素。相比之下，被动方法提供了一种更简单的替代方案，但目前仅限于对隐藏场景进行低分辨率的一维 [5, 57] 或二维 [54, 55, 58, 73] 重建。近期的一种方法 [13, 14] 利用门框的两条正交边缘，实现了对隐藏场景中非遮挡物体的出色三维成像。然而，该方法利用的阴影是由可见且已知的门框遮挡物投射的。在此，**我们将遮挡物辅助被动方法推广至对隐藏或其他未知遮挡物进行三维成像**。

贡献： 本文将计算潜望成像重新表述为一个可分离非线性最小二乘法（SNLLS）逆问题，从而能够根据单张半影照片进行三维重建，并提出了一种基于梯度的优化方法来求解该问题。此外，我们还引入了软阴影

扩散 （SSD），这是一种基于去噪概率扩散的全新生成模型，可根据物体软阴影的单张照片，以高分辨率三维点云的形式重建三维形状。最后，我们通过实验展示了在真实实验场景中首次联合重建三维遮挡物体和二维非遮挡物体。

1.1 相关工作

计算潜望成像： 图 1（a）展示了本文以及以往相关工作（例如 [3,54,55,73]）所研究的非视域成像场景照片。在该场景中，预先存在的环境光源照亮隐藏场景区域（以及观察者视线范围内的可见墙面）。隐藏场景的部分区域（此处为椅子和篮球）将光反射至可见墙面，但这些光被隐藏场景中的其他物体（此处为铲状形状）部分遮挡，从而在可见墙面上形成半影（或软阴影）。

通过简单的半影照片，Saunders 等人 [54,55] 和 Baradad 等人 [3] 重建隐藏场景中反光（非遮光）部分，其假设是遮挡光线的隐藏场景结构形状已知但位置未知，或者可从先验校准步骤获得遮挡物在可见墙面上的投影。此外，Yedidia 等人 [73]假设遮挡物近似为平面，并根据可见半影视频重建其二维投影，以及隐藏场景中非遮光部分的二维视频。本文提出一种新的计算策略，可根据半影的二维照片重建遮光物体的三维结构。因此，我们的方法所计算出的结果可以通过提供关于隐藏遮挡物的必要先验知识，补充这些既有方法。

与重建图像不同，直接推断隐藏场景的上下文属性可能对各种应用都很重要。沿着这一思路，Sharma 等人提出了基于学习的方法，可从半影的视频记录中准确判断人类活动，或识别未知房间中的人数 [59]。Medin 等人进一步证明，单个半影所包含的信息足以在没有视线的情况下对个体进行生物特征识别 [40]。尽管长期以来，人们一直在探索利用清晰的人体轮廓阴影的直接照片进行生物特征识别 [22,66], ，Medin 等人 [40] 考虑了一种更具挑战性的情况：观察到的阴影并非由遮挡光线的物体投射在隐藏个体面部所反射的光线形成。令人印象深刻的是，这两种方法 [40,59] 都展示了在不具备关于隐藏遮挡物或场景几何结构的先验知识时，仍能准确推断。

从S恢复形状： 从明暗恢复形状 [21,49] 和从阴影恢复形状 [7] 是计算机视觉领域长期以来广受关注的研究方向。从明暗恢复形状通过分析图像中亮度（明暗）的变化，推断物体表面各点的方向，从而根据二维图像重建三维形状。相比之下，从阴影恢复形状则根据三维形状

利用多个（二值）阴影图像重建三维形状。这些图像通过改变照明方向生成，并通过分析图像堆栈中最暗的部分来推断深度和表面方向。近期方法采用数据驱动技术，取得了令人瞩目的重建效果 [25, 32, 69], 即使在单目设置下 [9]也不例外。这些先前的研究利用了直接照明产生的明暗线索，以及在已知或可能未知的照明条件下由相互遮挡造成的清晰阴影。本文使用单幅图像，其中包含由隐藏的三维遮挡物体在附近表面上产生的软阴影；这些物体会部分遮挡复杂、漫射且不可见的照明源，而照明源与遮挡物体均位于图像视场之外。

扩散概率模型：生成式扩散概率模型[19, 61, 63]已展现出高质量图像生成能力 [52, 70], ，其表现可与生成对抗网络（GANs）媲美 [15]。这类模型的应用范围已扩展至时间序列预测 [51], 、音频生成 [28], 和文本转语音 [29, 47]等任务。一些研究探索了利用流模型 [71]或扩散先验 [6, 75]进行两阶段生成，以此重建点云。为降低生成延迟，Zhou 等人 [77] 提出了 PVD，这是一种用于生成点云的单阶段扩散先验。除了点云之外，Dream-Fields [23] 和 CLIP 引导的网格优化 [26]等方法也推动了文本条件三维生成的发展。其他方法则着重于利用基于扩散的先验，从多视角图像重建三维内容 [33, 46, 72]。

本文创新性地将扩散模型扩展到从二维软阴影照片生成三维点云。随后，我们使用训练好的回归模型生成投射阴影物体的三维网格表示。在对单个点进行扩散的既有工作基础上，我们的方法以软阴影图像为条件来引导扩散过程。

2 光传播模型

考虑图 1(a) 所示的配置，并令 $\mathbf{x}$ 表示隐藏场景体积中的任意一点。若隐藏场景中的某点到相机视场内可见墙面部分的光线路径在某些方向上被遮挡（或未被遮挡），但并非所有方向都如此，则该点适宜用于恢复。所有这类隐藏场景点的集合 $\mathcal{X}$ 构成了成像系统的计算视场（CFOV） [55]。令$f(\mathbf{x})$ 表示隐藏场景点$\mathbf{x}$处未知的辐射度分布，则可见墙面点 $i(\mathbf{p})$ 处的辐照度 $\mathbf{p}$ 为 [55]:

$$i(\mathbf{p}) = \int_{\mathbf{x}\in\mathcal{X}} \frac{g(\mathbf{p},\mathbf{x})}{\|\mathbf{x}-\mathbf{p}\|_2^2} v(\mathbf{x},\mathbf{p};\boldsymbol{\theta}_o) f(\mathbf{x})\,\mathrm{d}\mathbf{x} + b(\mathbf{p}), \tag{1}$$

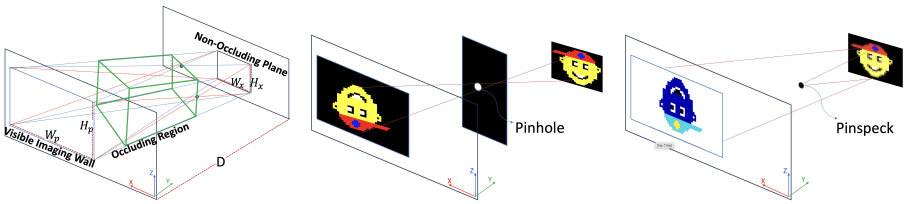
其中 $g(\mathbf{p}, \mathbf{x}) = \cos(\measuredangle(\mathbf{x} - \mathbf{p}, \mathbf{n}_\mathbf{x}))\cos(\measuredangle(\mathbf{p} - \mathbf{x}, \mathbf{n}_\mathbf{p}))$ 对朗伯反射和缩短效应进行建模，$\mathbf{n}_\mathbf{p}$ 和 $\mathbf{n}_\mathbf{x}$ 分别表示 $\mathbf{p}$ 和 $\mathbf{x}$ 处的表面法线。函数 $v(\mathbf{x}, \mathbf{p}; \boldsymbol{\theta}_o)$ 对未知遮挡物体在 $\mathbf{p}$ 和 $\mathbf{x}$ 之间的可见性进行建模，其形状由$\boldsymbol{\theta}_o$隐式参数化；当从 $\mathbf{x}$ 到 $\mathbf{p}$ 的射线未被遮挡时，$v(\mathbf{x}, \mathbf{p}; \boldsymbol{\theta}_o) = 1$，否则为 0。最后，$b(\mathbf{p})$ 对噪声以及其他照明源造成的影响进行建模——主要是任何现存的可见侧向照明源，以及照明

来自 CFOV 之外隐藏场景点的贡献，即$\int_{\mathbf{x}\notin\mathcal{X}} \frac{g}{\|(\mathbf{p},\mathbf{x})\,\mathbf{x}-\mathbf{p}\|_2^2} f(\mathbf{x})\,\mathrm{d}\mathbf{x}$——这些贡献不会形成半影。通常假设可恢复的隐藏场景 $\mathcal{X}$ 是位于已知深度 [2,55,73], 的单个平面， [3,54] 或是位于相对于可见墙面已知深度的若干平面。将测量平面和隐藏场景平面离散为像素，即可得到离散前向模型：

$$\mathbf{y} = \mathbf{A}(\boldsymbol{\theta}_{\mathrm{o}})\mathbf{f} + \mathbf{b}, \tag{2}$$

其中$\mathbf{y} \in \mathbb{R}^M$ 是实测照片的向量化表示，$\mathbf{f} \in \mathbb{R}^N$ 是包含隐藏场景像素未知辐射度的向量，而$\mathbf{A}(\boldsymbol{\theta}_{\mathrm{o}}) \in \mathbb{R}^{M\times N}$ 是给定的光传输矩阵 $\boldsymbol{\theta}_{\mathrm{o}}$（它隐式表示了遮挡物的形状和位置）。

Saunders 等人 [55]假设遮挡物形状已知，但使用 $\boldsymbol{\theta}_{\mathrm{o}}$ 表示其位置。本文研究遮挡物形状也未知的情况，并使用 $\boldsymbol{\theta}_{\mathrm{o}}$ 作为未知遮挡物 3D 形状的隐式表示。通过（2），估计未知量$\mathbf{f}$ 和$\mathbf{b}$ 是一个病态的可分离非线性逆问题。在第*4*节中，我们在模型基础上，为该逆问题提出两种解法，并明确将隐藏场景划分为两个组成部分：充当光线遮挡物的部分，以及仅充当发光体／反射体的部分。



(a) 计算视场 (b) 针孔遮挡物 (c) 针孔式遮挡物

图 2：计算视场和两个简单的遮挡物示例。(a) 相机视场投射到非遮挡场景平面上。遮挡区域是将测量视场的边界与所需隐藏场景平面的边界连接后形成的体积。(b) 理想针孔投射出场景图像。(c) 理想针孔式遮挡物投射出场景的负像。

2.1 隐藏场景分解与计算视场

我们的目标是同时重建隐藏遮挡物的三维图像，以及隐藏非遮挡物（即光的发光体/反射体）的二维图像。为构造一个可行的逆问题并便于准确重建，我们必须将重建体积限制为仅包含这样的隐藏场景部分：它们能够在可见墙面的拍摄区域内某处投射出半影。假设非遮挡物的二维图像是一个固定尺寸平面的辐射度，该平面 $W_{\mathrm{x}} \times H_{\mathrm{x}}$ 距可见墙面 D （见图2a）。接着，如图2a所示，遮挡体积定义为由以下视锥体

其边缘与连接相机视场中每个顶点和非遮挡隐藏场景平面中每个顶点所形成的直线重合。

对于我们的问题设定，三维遮挡体积和选定的二维非遮挡平面统称为计算视场（CFOV）。这代表了 CFOV 概念 [55] 从二维到三维的自然扩展。

3 三维计算潜望成像：模型与表示

为了利用我们新的重新表述进行三维重建，我们考虑三维结构的两种不同表示形式。第一种表示形式见第3.2节，是一种显式表示，采用对 CFOV 中遮挡视锥部分进行均匀体素化的方法。第二种表示形式见第3.4节，是一种基于隐式神经点云的表示形式，它为三维形状引入深度先验。每种表示形式都启发了一种相应的重建方法，该方法将在第4 节介绍，并建立在简单的针孔式遮挡物之上。

3.1 简单的针孔式遮挡物

针孔式遮挡物（或反针孔） [11, 64] 遮挡物是针孔的光学互补。因此，单个针孔式遮挡物位于$\mathbf{u} = (u_x, u_y, u_z)$ 会阻挡所有到达点$\mathbf{u}$的射线，使得可见墙面上的点$\mathbf{p} = (p_x, 0, p_z)$与非遮挡隐藏场景平面上的点$\mathbf{x} = (x, y, z)$之间的可见性为

$$v_{\mathrm{ps}}(\mathbf{p}, \mathbf{x}; \mathbf{u}) = 1 - \delta\left(p_x - m u_x - \frac{xy}{u_y}\right) \delta\left(p_z - m u_z - \frac{yz}{u_y}\right) \tag{3}$$

其中 $\delta(\cdot)$ 是狄拉克δ函数，且 $m = 1 - y/u_y$。施加光学互补性质 [11, 64]，后，可得到在可见平面上形成的（负）图像表达式如下：

$$i(\mathbf{p}) = \int_{\mathbf{x} \in \mathcal{X}} \frac{g(\mathbf{p}, \mathbf{x})}{\|\mathbf{x} - \mathbf{p}\|_2^2} \left(1 - v_{\mathrm{ph}}(\mathbf{p}, \mathbf{x}; \mathbf{u})\right) f(\mathbf{x}) \, \mathrm{d}\mathbf{x}, \tag{4}$$

其中，针孔可见性 $v_{\mathrm{ph}}(\mathbf{p}, \mathbf{x};\mathbf{u}) = 1 - v_{\mathrm{ps}}(\mathbf{p}, \mathbf{x};\mathbf{u})$。

3.2 UniformVoxels 表示

根据我们之前的工作 [48], ，假设光线遮挡视锥被均匀离散化为 K 个体素，其中心位于 $\{\mathbf{u}_k\}_{k=1}^K$处，则在高密度离散化的情况下，任意三维形状都可以通过选取由其索引标识的部分体素来进行良好近似。在此情形下， $v(\mathbf{x}, \mathbf{p}; \boldsymbol{\theta}_o) = 1 - \sum_{k=1}^K \alpha_k v_{\mathrm{ph}}(\mathbf{p}, \mathbf{x};\mathbf{u}_k)$，其中二值标量 α_k 表示在$\mathbf{u}_k$处是否存在针孔式遮挡物：存在时取1，不存在时取0；因此，可见墙面辐照度近似为：

$$i(\mathbf{p}) = \int_{\mathbf{x} \in \mathcal{X}} \frac{g(\mathbf{p}, \mathbf{x})}{\|\mathbf{x} - \mathbf{p}\|_2^2} \left(1 - \sum_{k=1}^K \alpha_k v_{\mathrm{ph}}(\mathbf{p}, \mathbf{x}; \mathbf{u}_k)\right) f(\mathbf{x}) \, \mathrm{d}\mathbf{x}. \tag{5}$$

我们将非遮挡平面离散化为 $N = (N_x \times N_z)$ 个像素，并用 f_n 表示 n号像素的未知辐射度，用$\mathbf{f} \in \mathbb{R}^N$表示所有$N$ 个辐射度的集合。此外，我们定义 $\{\mathbf{p}_m\}_{m=1}^M$ 为可见墙面上被 $M = (M_x \times M_z)$个相机像素捕获的像素位置。因此，

$$\mathbf{y} = \mathbf{A}\left(1 - \sum_{k=1}^{K} \alpha_k \mathbf{V}_k\right)\mathbf{f} = \left(\mathbf{A} - \sum_{k=1}^{K} \alpha_k \mathbf{A}\mathbf{V}_k\right)\mathbf{f}, \tag{6}$$

其中，$\mathbf{y} \in \mathbb{R}^M$ 是向量化的半影照片，$\mathbf{V}_k \in \mathbb{R}^{M \times N}$ 是对应于$\mathbf{u}_k$处针孔可见性的矩阵。因此，重建三维形状等价于从$\mathbf{y}$计算 $\boldsymbol{\theta}_o \overset{\text{def}}{=} \boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_K)^T$。

3.3 点云表示

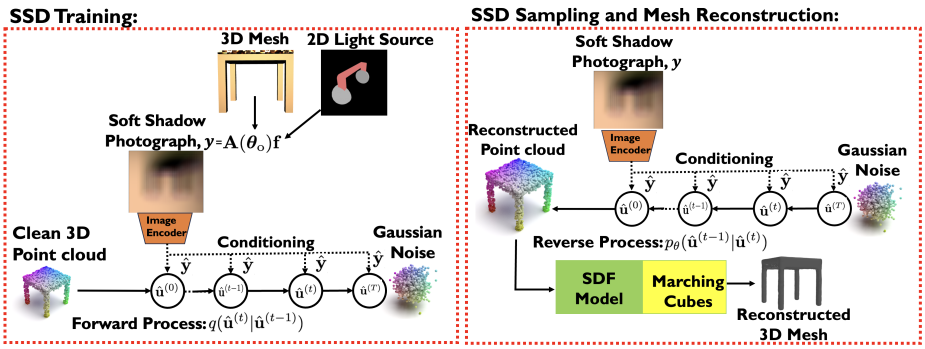
尽管采用均匀体素化并使用更精细的离散化有望实现准确的场景表示，但这会增加计算复杂度和内存需求，并导致条件数较差。此外，对 $\boldsymbol{\alpha}$ 施加二元约束会导致非凸优化。相比之下，点云表示不再采用基于均匀网格的方式，而是直接优化向量位置$\mathbf{u}_k$ ，而非二元向量 $\boldsymbol{\alpha}$。它还能以低维表示有效地捕捉各种三维形状，同时兼容我们的针孔式遮挡物模型（4）。因此，我们旨在求解$\mathbf{y} = \mathbf{A}(\boldsymbol{\theta}_o)\mathbf{f}$，其中 $\boldsymbol{\theta}_o \overset{\text{def}}{=} \{\mathbf{u}_k\}_{k=1}^K$。直接优化$K$ 点的位置颇具挑战。尽管如此，为了实现准确的高分辨率重建，我们开发并利用了一种学习得到的点云形状先验，称为软阴影扩散。

3.4 软阴影扩散模型

我们提出的软阴影扩散（SSD）模型采用了最先进的去噪扩散模型 [38, 60, 61, 76] ，用于条件点云生成。在此，去噪扩散过程以测得的软阴影照片 $\hat{\mathbf{y}}$ 的潜在表示 $\mathbf{y}$ 为条件。我们的动机是利用 SSD 作为强大的学习先验，生成能够解释半影照片中软阴影的高分辨率 3D 点云。

训练完成后，我们的 SSD 模型 $G_w(\mathbf{y})$ 通过迭代地逆转马尔可夫前向过程，生成去噪后的点云 $\{\mathbf{u}_k^0\}_{k=1}^K$ ，如图 3b所示。具体而言，以软阴影照片的潜在表示 $\hat{\mathbf{y}}$为条件，模型从初始点云 $\{\mathbf{u}_k^T\}_{k=1}^K$ 中迭代去除噪声；该点云是高斯噪声过程的一次随机实现。

训练过程（见图 3a）首先将 $\{\mathbf{u}_k^0\}_{k=1}^K$ 设为从无噪声点云数据集中随机采样得到的样本。然后，根据 $\{\mathbf{u}_k^0\}_{k=1}^K$，使用 $\mathbf{u}_k^t = \sqrt{\bar{\alpha}_t}\mathbf{u}_k^0 + \sqrt{1-\bar{\alpha}_t}\epsilon$，依次计算出一系列噪声逐渐增大的点云 $\{\mathbf{u}_k^t\}_{k=1}^K$，其中 $t = 1$、2、……、T；其中 $k = 1$、2、……、K，其中 ϵ 为零均值



(a) 训练流程 (b) 推理流程

图 3：软阴影扩散模型流程。(a) 训练时，模型通过前向扩散过程将干净的输入点云映射为噪声。(b) 推理时，模型从初始噪声开始，并通过以编码后的软阴影照片 $\hat{\mathbf{y}}$为条件数，将噪声转换为遮光结构的点云表示。随后，点云会被转换为遮挡结构的网格表示。

单位方差白高斯噪声，系数 α_t 会随着时间步 t 递增而逐渐减小。利用这些带噪样本，通过最小化以下损失函数来更新底层神经网络的权重 w [63]:

$$L = \mathbb{E}_{t,\{\mathbf{u}_k^0\}_{k=1}^K,\epsilon} \left\| \epsilon_\theta(\{\mathbf{u}_k^t\}_{k=1}^K, \hat{\mathbf{y}}, t) - \epsilon \right\|_2^2 .$$

我们在补充材料中提供了有关 SSD 模型训练和采样的更多细节。

4 3D计算潜望成像：反演

假设背景贡献对该项的影响$\mathbf{b}$ 很小，可以通过求解 Tikhonov正则化最小化问题来获得（2）的一个解

$$\underset{(\mathbf{f},\boldsymbol{\theta}_o)}{\arg\min} \|\mathbf{A}(\boldsymbol{\theta}_o)\mathbf{f} - \mathbf{y}\|_2^2 + \lambda \|\mathbf{f}\|_2^2, \tag{7}$$

其中 $\|\cdot\|_2$ 为二范数，$\lambda > 0$ 为正则化参数。对于最小化问题（7），一种交替优化方法为:

$$\begin{aligned} \widehat{\mathbf{f}}^k &= \underset{\mathbf{f}}{\arg\min} \left\| \mathbf{A}(\widehat{\boldsymbol{\theta}}_o^{k-1})\mathbf{f} - \mathbf{y} \right\|_2^2 + \lambda \|\mathbf{f}\|_2^2 \\ &= \left(\mathbf{A}^\mathsf{T}(\widehat{\boldsymbol{\theta}}_o^{k-1})\mathbf{A}(\widehat{\boldsymbol{\theta}}_o^{k-1}) + \lambda \mathbf{I} \right)^{-1} \mathbf{A}^\mathsf{T}(\widehat{\boldsymbol{\theta}}_o^{k-1})\mathbf{y} \qquad (8) \\ \widehat{\boldsymbol{\theta}}_o^k &= \underset{\boldsymbol{\theta}_o}{\arg\min} \left\| \mathbf{A}(\boldsymbol{\theta}_o)\widehat{\mathbf{f}}^k - \mathbf{y} \right\|_2^2 . \qquad (9) \end{aligned}$$

此外，将（7）中的 f 替换为其闭式估计，可得到以下最小化问题：

$$\widehat{\boldsymbol{\theta}}_{\mathrm{o}}=\arg\min_{\boldsymbol{\theta}_{\mathrm{o}}}\left\|\left[\mathbf{A}(\boldsymbol{\theta}_{\mathrm{o}})\left(\mathbf{A}^{\mathsf{T}}(\boldsymbol{\theta}_{\mathrm{o}})\mathbf{A}(\boldsymbol{\theta}_{\mathrm{o}})+\lambda\mathbf{I}\right)^{-1}\mathbf{A}^{\mathsf{T}}(\boldsymbol{\theta}_{\mathrm{o}})-\mathbf{I}\right]\mathbf{y}\right)\right\|_2^2. \tag{10}$$

原则上，无需显式估计 $\boldsymbol{\theta}_{\mathrm{o}}$ 半影测量值$\mathbf{y}$，即可重建$\mathbf{f}$，而不必将其作为中间步骤。我们在第4.2节概述的受物理启发的神经网络反演模型中运用了这一观察。

4.1 基于梯度的优化进行反演

使用均匀体素表示（见第3.2节），我们希望根据 $\boldsymbol{\theta}_{\mathrm{o}} \overset{\text{def}}{=} \{\alpha_k\}_{k=1}^{K}$ 和$\mathbf{f}$ 估计$\mathbf{y}$，使得（2）成立。为此，我们允许背景不可忽略，并求解等价的 Tikhonov 正则化最小化问题：

$$\arg\min_{(\mathbf{f},\boldsymbol{\theta}_{\mathrm{o}},\mathbf{b},\lambda)} \|\mathbf{A}(\boldsymbol{\theta}_{\mathrm{o}})\mathbf{f}+\mathbf{b}-\mathbf{y}\|_2^2+\lambda\|\mathbf{f}\|_2^2. \tag{11}$$

为施加凸性约束，我们使用 σ（$\mathbf{z}_k$），其中 $\sigma(\cdot)$ 是 Sigmoid 函数，用作二元值 α_k 的近似。因此，我们改为使用梯度下降优化 $\mathbf{z}_k$。该方法在模拟数据集上表现稳健。在真实实验中，我们发现，当背景和噪声的影响不可忽略时，加入正值约束效果更好。

算法1 总结了我们基于体素化优化的方法步骤。实证结果（见补充材料）表明，与假设背景可忽略相比，使用算法$\mathbf{f}$、$\mathbf{z}$、$\mathbf{b}$和 λ进行交替优化，具有更优的收敛特性。1。

算法1 交替最小化方法

需要： $\mathbf{A}, [\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3, \ldots\ldots, \mathbf{V}_K]$、$\mathbf{y}$、numIter

1：随机初始化 $\mathbf{z}_0$、$\mathbf{b}_0=\mathbf{0}$和 $\lambda_0=1$
2：初始化步长 η_b、η_z和 η_λ
3：**对于** $i=1$至numIter **执行**
4：　$\mathbf{A}_v = A\odot(1-\frac{1}{K}\sum_{k=1}^{K}\mathbf{V}_k\cdot\sigma(\mathbf{z}^{i_k}))$　　施加二元约束
5：　$\mathbf{f}^* = (\mathbf{A}_v{}^{\mathsf{T}}\mathbf{A}_v+\lambda_{i-1}I)^{-1}\mathbf{A}_v{}^{\mathsf{T}}\mathbf{y}$　　估计二维非遮挡场景
6：　$\mathbf{f}_i = \text{最大}(\mathbf{f}^*+\mathbf{b}_{i-1},\ 0)$
7：　$\mathcal{L}(\mathbf{y},\mathbf{A}_v\mathbf{f}_i) = \frac{1}{M}\sum_{j=1}^{M}\|\mathbf{y}^j-(\mathbf{A}_v\mathbf{f}_i)_j\|_2^2$
8：　$\mathbf{b}_i \leftarrow \mathbf{b}_{i-1}-\eta_b\frac{\partial\mathcal{L}(\mathbf{y},\mathbf{A}_v\mathbf{f}_i)}{\partial\mathbf{b}_{i-1}}$　　更新背景贡献估计
9：　$\mathbf{z}_i \leftarrow \mathbf{z}_{i-1}-\eta_z\frac{\partial\mathcal{L}(\mathbf{y},\mathbf{A}_v\mathbf{f}_i)}{\partial\mathbf{z}_{i-1}}$　　更新遮挡物估计
10：　$\lambda_i \leftarrow \lambda_{i-1}-\eta_\lambda\frac{\partial\mathcal{L}(\mathbf{y},\mathbf{A}_v\mathbf{f}_i)}{\partial\lambda_{i-1}}$　　更新正则化常数的估计
11：**结束循环**
12：**返回** $\widehat{\mathbf{f}}=\mathbf{f}_i, \widehat{\boldsymbol{\alpha}}=\sigma(\mathbf{z}_i)>0.5$

实现细节：交替最小化算法（算法1）使用 PyTorch 实现。遮挡物区域被离散化为 $N_x \times N_y \times N_z$个均匀体素。遮挡物需要较高的离散化精度。即使采用 $10 \times 5 \times 10$这类粗略离散化，所需的计算量和内存仍然很大。

为降低预先计算矩阵 $\mathbf{V}_k$ 所需的高复杂度和内存开销（该矩阵涵盖所有可能的针孔贡献），我们采用稀疏矩阵表示来表示 $\mathbf{V}_k$。这种方法仅存储每个针孔所贡献的特定像素位置。所需的梯度更新（针对所有针孔代理 $\{\mathbf{z}_k\}_k$）均在这种新表示中计算。（更多详情和模拟示例请参见补充材料。）

4.2 通过受物理启发的神经网络进行反演

在此，我们提出一种重建方法，将基于物理的可分离模型与作为学习式形状先验的 SSD 相结合。具体而言，由于 SSD 只能生成合理的三维形状点云，而无法确定它们在非视域场景中的位置，因此我们在估计流程中加入了对 [55] 中位置估计算法的改编。我们完整的重建流程（包括对 $\mathbf{f}$ 的估计）如下：

– **阶段 1：三维形状的条件生成**。给定软阴影照片 $\mathbf{y}$，从条件扩散模型中采样得到三维形状。

$$\{\widehat{\mathbf{u}}_k^0\}_{k=1}^K = G_w(\mathbf{y}), \tag{12}$$

– 阶段 2：生成三维形状的定位。通过求得三维平移向量 $\boldsymbol{\delta}_\mathbf{u}$，估计遮挡物体的空间位置：$\widehat{\boldsymbol{\delta}_\mathbf{u}} = \arg\max_{\boldsymbol{\delta}_\mathbf{u}} \left\|\mathrm{H}(\{\widehat{\mathrm{u}}_k^0 + \boldsymbol{\delta}_\mathbf{u}\}_{k=1}^K)\mathrm{y}\right\|_2^2$，其中 $\mathrm{H}(\cdot) = \mathrm{A}(\cdot)\left(\mathrm{A}(\cdot)^T\mathrm{A}(\cdot)\right)^{-1}\mathrm{A}(\cdot)^T$。– 阶段 3：非遮挡部分的重建。我们将经过适当平移的遮挡物点云 $\{\widehat{\mathrm{u}}_k^0 + \widehat{\boldsymbol{\delta}_\mathbf{u}}\}_{k=1}^K$ 转换为网格表示，并计算其网格表示以获得 $\boldsymbol{\theta}_\mathrm{o}$。然后，我们建立并求解以下全变分正则化问题：$\widehat{\mathrm{f}} = \arg\min ||\mathrm{y} - \mathrm{A}(\widehat{\boldsymbol{\theta}_\mathrm{o}})\mathrm{f}||_2^2 + \lambda|\mathrm{f}|_{\mathrm{TV}}^2$。 f

实现细节：我们使用 PyTorch 实现了受物理启发的神经网络模型，其中包括 SSD 网络。
数据集。我们使用数千个三维模型训练 SSD模型。我们使用 Blender 将 ShapeNet [8] 数据集中的所有三维模型从 3 个随机相机视角渲染为 RGB图像，[12]。这些图像呈现的是位于二维平面上的非遮挡物体，而三维模型则是遮光结构。我们的 Blender 脚本会将每个三维模型归一化到单位边界立方体，配置标准照明设置，并使用 Blender 内置的实时渲染引擎导出 RGB图像。随后，我们对所有可能的渲染图像与三维模型配对应用（2）中的渲染方程，并随机选取配对，以模拟多样化的训练样本。为了生成三维模型的点云，我们首先沿物体表面采样 50,000 个点，为每个物体构建稠密点云。然后对这些分布不均匀的稠密点云进行重采样，

生成由 2,048 个均匀分布点组成的点云。我们的数据集包含约 260,000 个不同物体类别的模拟实例。

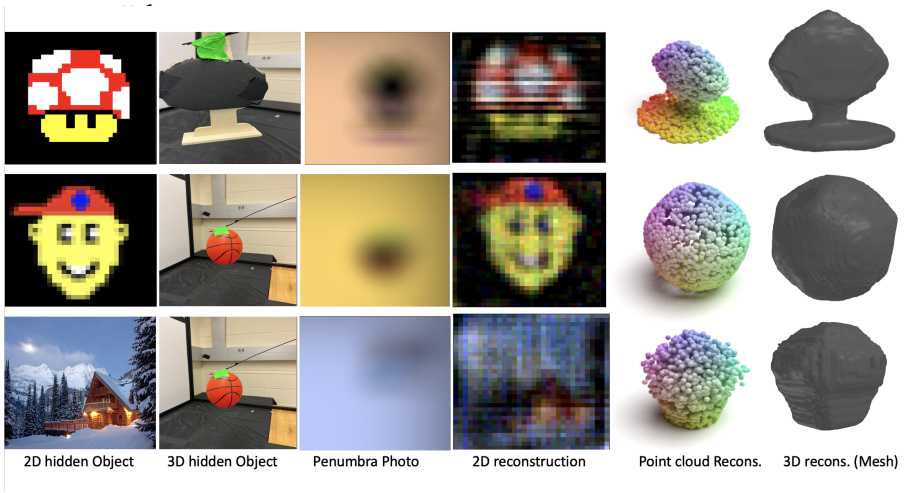
软阴影图像编码器。 为了将软阴影照片编码为潜在表示，$\hat{}\mathbf{y}$我们采用了 CrossFormer 架构 [67]并将其与扩散过程联合训练。CrossFormer 引入了对处理视觉输入至关重要的跨尺度注意力机制。此外，与传统的视觉变换器不同，CrossFormer 能够有效捕捉不同尺度特征之间的相关性，这对于理解复杂的视觉输入（如我们的半影照片）至关重要。

模型实现。 不同于以往通常采用专用三维架构的点云重建工作 [77] 或基于变换器的 [43] 模型，我们采用了一种更简单但有效的基于 U-Net 的架构 [53]，使用一维卷积。我们的模型支持快速采样出比训练时更多的点，并旨在根据输入预测噪声的均值 μ（3.4）中的噪声，输入包括$\hat{}\mathbf{y}$、 t和 $\{\mathbf{u}_k^t\}_{k=1}^K$，即软阴影二维照片的编码表示、扩散过程中的当前时间步 t 以及中间噪声点云。整个模型架构如图3 所示，更多细节请参见补充文档。

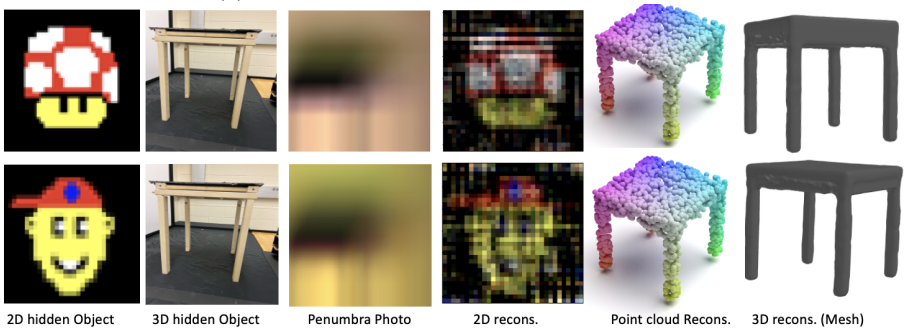
三维网格表示。重建隐藏场景中不被遮挡的部分$\mathbf{f}$ 需要准确渲染估计出的遮挡物所产生的软阴影图像。稀疏点云可能导致误差，从而无法准确近似软阴影。因此，我们将软阴影扩散生成的点云转换为网格。我们使用整个 ShapeNet [8] 数据集训练一个有符号距离场（SDF）神经网络模型，用于预测重建点云的 SDF。该 SDF 模型采用自动编码器架构 [10]。随后，我们将移动立方体算法 [37] 应用于计算得到的 SDF，以生成遮光物体的网格表示。网格表示有助于更高保真地渲染软阴影。

5 实验

我们通过真实数据和合成数据的实验结果，展示所提出方法的有效性。由于以往方法针对不同的采集场景（例如可获取视频帧 [73, 74], 预先标定的遮挡物可见性函数 [3], 已知的遮挡物形状 [55], 甚至完全已知的遮挡物 [17, 68]），因此无法进行直接且公平的比较。这些已有技术仅关注重建隐藏场景中的发光部分，而我们的方法首次进一步寻求对隐藏遮挡物进行三维重建。我们考虑的是一个难得多的问题，因此，预期所能实现的二维重建不会优于现有方法。



(a) 基于 SSD 的简单三维物体重建。

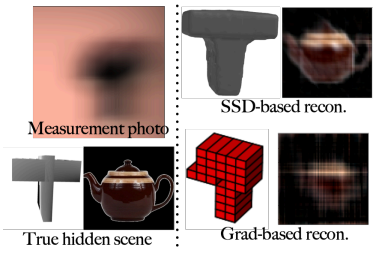


(b) 基于 SSD 的大型三维物体重建。

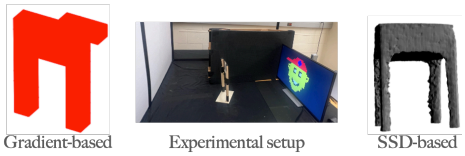
图4：基于软阴影扩散模型的真实实验重建结果。该模型以软阴影照片（第 3 列）为输入，用于条件化扩散过程，并生成隐藏区域内三维遮光结构（第 2 列）的三维网格表示（第 6 列）。随后，将该网格表示转换为发光物体的二维RGB图像（第 4 列），以实现重建。

5.1 定性结果

真实实验示例展示了 软阴影扩散，如图 1 和 4所示；基于梯度的 方法的真实实验示例如图 5所示。图 4 和 5展示了隐藏场景中的非遮挡部分，该部分显示在距离可见墙面 1.08 米的显示器上，二者之间放置了三维遮挡物。我们通过使用 400 万像素相机拍摄可见墙面，获得了测量数据。测量数据被输入我们的受物理启发的神经网络反演方法，或基于梯度的反演方法中的算法 1。在 SSD 的最终阶段，我们使用全变分正则化重建非遮挡部分的二维图像的每个通道。重建结果与真实值在定性上具有良好的一致性



(a) 模拟数据。



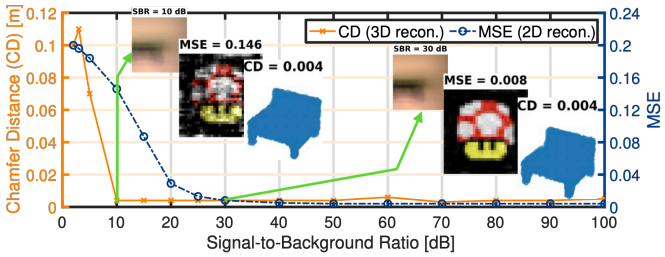
(b) 真实数据。

图 5： 比较 SSD 与基于梯度的优化方法所得到的重建结果。SSD 在三维物体上进行训练，并为规划器遮挡物输出一个（三维）四腿桌。基于梯度的重建使用了遮挡区域的 $10\times5\times10$ 离散化。

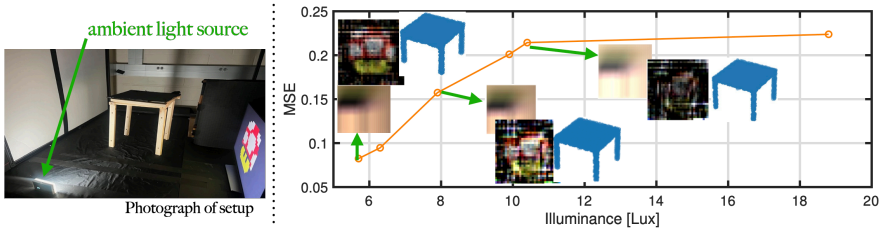
隐藏场景。然而，对于需要更精细离散化的复杂遮挡物形状，基于梯度的方法难以良好扩展（如补充文档所示）。

即使非遮挡的隐藏场景部分会反射光（而非自发光），并且覆盖较大的深度范围（而不是局限于一个平面），重建得到的遮挡物三维图像和非遮挡物二维图像仍然非常准确，如图1所示。这一实验进一步突显了所提出的基于 SSD 方法令人惊讶的泛化能力；该方法仅使用模拟数据进行训练。

补充材料中提供了更多真实实验和仿真的定性结果。



(a) 基于模拟数据的重建。



(b) 基于真实实验数据的重建。

图 6：基于 SSD 的方法对背景照明的鲁棒性。 这些结果展示了每种背景光照水平下的实测照片和重建结果。在显示器亮度固定的情况下，照度值越大，表示 (b) 中环境光源的贡献越大。

5.2 定量结果

在此，我们研究了随着背景照明强度增加，我们的方法的鲁棒性（图6）；在不同测量噪声水平下获得的重建结果则见补充文档（补充章节 4.1）。在真实实验结果中，我们引入了一个额外的外部光源，它提高了拍摄照片的整体亮度，但不会投射半影。我们报告二维重建的均方误差（MSE），以及三维遮挡物重建的倒角距离（CD）。在仿真实验中，当信号背景比处于合理范围时，我们的方法性能会缓慢下降；即使信号背景比约为 10 dB，三维重建仍然准确，而当信号背景比低于 20 dB 时，二维重建则逐渐退化。真实数据重建也呈现出类似趋势。即使在信号背景比极低时，三维重建仍表现出令人惊讶的稳定性，这证明了所提出的 SSD 网络所施加强形状先验的实用价值。

SSD 对测量噪声表现出类似的鲁棒性，即使信噪比低至 10dB，三维重建质量也相对不受影响。补充材料第 4.2 节所示的其他实验表明，SSD 在重建未见过的三维形状类别方面也具有良好的泛化能力。

6 讨论与结论

我们提出了一种计算潜望成像方法，使三维重建成为可能。该方法启发我们设计了一种基于物理的算法和一种受物理启发的神经网络，用于计算光遮挡物的三维重建，以及隐藏场景中光反射结构的二维重建。这标志着相较于先前的遮挡物辅助被动 NLOS 成像技术的一项进步；这类技术依赖于场景中光遮挡物的先验已知信息。我们受物理启发的神经网络模型的核心是一种新颖的软阴影扩散（SSD）网络，它能够从软阴影中重建点云，并将去噪扩散模型拓展至软阴影的解读。尽管 SSD 是在仿真中训练的，但这两种方法在真实实验中均取得了成功。研究还表明，面对噪声和额外背景时，SSD 的性能会平稳退化。

尽管如此，我们的工作仍有一些方面值得进一步研究：首先，由于更高离散化程度所需的计算量较大，我们的基于梯度的方法仅限于重建简单形状。尽管稀疏矩阵方法在仿真中缓解了这一限制，但在真实场景中效果较差（如补充材料所示）。其次，我们的SSD模型能够预测视图一致的三维点云，却无法提供NLOS配置中各点真实位置的信息。为便于恢复非遮挡物体，我们将遮挡物空间划分为多个尺寸与遮挡物相同的三维边界框。接下来，我们在这些边界框内进行网格搜索，以便在隐藏区域中准确定位遮挡物。然而，在真实的非受控场景中，遮挡物体的尺寸通常未知

。这些不足为未来研究提供了方向。

参考文献

1. Adib, F., Katabi, D.: 用 Wi-Fi 穿透墙壁！载于：ACM SIGCOMM 会议论文集。第 75–86 页（2013）2 2. Aittala, M., Sharma, P., Murmann, L., Yedidia, A., Wornell, G., Freeman, B., Durand, F.: 计算镜：通过深度矩阵分解盲逆光传输。神经信息处理系统进展（NeurIPS）32（2019）5 3. Baradad, M., Ye, V., Yedidia, A.B., Durand, F., Freeman, W.T., Wornell, G.W., Torralba, A.: 从阴影推断光场。载于：IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集。第 6267–6275 页（2018）2, 3, 5, 11 4. Boger-Lombard, J., Slobodkin, Y., Katz, O.: 利用非受控随机噪声源实现拐角处被动非视距声学定位。Scientific Reports 13(2)，4952（2023）5. Bouman, K.L., Ye, V., Yedidia, A.B., Durand, F., Wornell, G.W., Torralba, A., Freeman, W.T.: 将拐角变成相机：原理与方法。载于：IEEE/CVF国际计算机视觉会议（ICCV）（2017年10月）2 6. Cai, R., Lai, H., Chami, I., Guibas, L.J.: 学习用于形状生成的梯度场。载于：欧洲计算机视觉会议（ECCV）论文集。第 10751–10760 页（2020）4 7. Cavanagh, P., Leclerc, Y.G.: 由阴影恢复形状。《实验心理学杂志：人类感知与表现》15(3)，813（1989）8. Chang, A.X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., Su, H., Xiao, J., Yi, L., Yu, F.: ShapeNet：一个信息丰富的 3D 模型库（2015）3 10, 11 9. Chaudhury, A.N., Keselman, L., Atkeson, C.G.: 面向机器人操作的从明暗恢复形状。载于：IEEE/CVF冬季计算机视觉应用会议（WACV）论文集。第 8389–8398 页。IEEE（2024）4 10. Chou, G., Bahat, Y., Heide, F.: 扩散 SDF：有符号距离函数的条件生成建模。载于：IEEE/CVF国际计算机视觉会议（ICCV）（2023）11 11. Cohen, A.L.: 反针孔成像。《光学学报》29(6)，63–67（1982）3 12. Community, B.O.: Blender——一个 3D 建模与渲染软件包。Blender基金会、Stichting Blender基金会，阿姆斯特丹（2018），http://www.blender.org 10 13. Czajkowski, R., Murray-Bruce, J.: 双边缘分辨三维非视距成像：费希尔信息均衡离散化。载于：IEEE声学、语音与信号处理国际会议（ICASSP）论文集。第 2535–2539 页（2024）。https://doi.org/10.1109/ICASSP48485.2024.10446406 2 14. Czajkowski, R., Murray-Bruce, J.: 使用普通相机进行双边缘分辨三维非视距成像。《自然·通讯》15(2)，1162（2024）15. Dhariwal, P., Nichol, A.: 扩散模型在图像合成方面优于生成对抗网络。载于：神经信息处理系统进展（NeurIPS）（2021）4 16. Faccio, D., Velten, A., Wetzstein, G.: 非视域成像。《自然综述：物理》2(6)，318–327（2020）2

17. Geng, R., Hu, Y., Lu, Z., Yu, C., Li, H., Zhang, H., Chen, Y.: 使用最优传输进行被动非视线成像。《IEEE图像处理汇刊》31，110–124（2022）2, 11 18. Heide, F., O'Toole, M., Zang, K., Lindell, D.B., Diamond, S., Wetzstein, G.: 使用部分遮挡物和表面法线进行非视线成像。《ACM图形学汇刊》（2019）2 19. Ho, J., Jain, A., Abbeel, P.: 去噪扩散概率模型。载于：神经信息处理系统进展（NeurIPS）。第 33 卷，第 6840–6851 页（2020）4 20. Ho, J., Jain, A., Abbeel, P.: 去噪扩散概率模型。《神经信息处理系统进展》（NeurIPS）33，6840–6851（2020）26, 27 21. Horn, B.K.P., Brooks, M.J.（编）：从明暗恢复形状。麻省理工学院出版社，美国马萨诸塞州剑桥（1989）3 22. Iwashita, Y., Stoica, A., Kurazume, R.: 使用阴影生物特征识别步态。《模式识别快报》33(16)，2148–2155（2012）3 23. Jain, A., Mildenhall, B., Barron, J.T., Abbeel, P., Poole, B.: 利用梦境场进行零样本文本引导物体生成。载于：IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集（2021年12月）4 24. Kaga, M., Kushida, T., Takatani, T., Tanaka, K., Funatomi, T., Mukaigawa, Y.: 基于镜面反射与漫反射的热非视域成像。《日本信息处理学会计算机视觉与应用汇刊》11（2019年12月）2 25. Karnieli, A., Fried, O., Hel-Or, Y.: Deepshadow：基于阴影的神经形状重建。载于：欧洲计算机视觉会议（ECCV）论文集。第 415–430 页（2022）4 26. Khalid, N.M., Xie, T., Belilovsky, E., Tiberiu, P.: CLIP-Mesh：使用预训练图文模型从文本生成带纹理网格。载于：SIGGRAPH Asia 2022会议论文集（2022年12月）4 27. Kirmani, A., Jeelani, H., Montazerhodjat, V., Goyal, V.K.: 漫射成像：利用非聚焦时间分辨照明与传感创建光学图像。《IEEE信号处理快报》19(2)，31–34（2012）28. Kong, Z., Ping, W., Huang, J., Zhao, K., Catanzaro, B.: DiffWave：一种适用于音频合成的多功能扩散模型。载于：国际学习表征会议（ICLR）论文集（2021）4 29. gil Lee, S., Kim, H., Shin, C., Tan, X., Liu, C., Meng, Q., Qin, T., Chen, W., Yoon, S.H., Liu, T.Y.: PriorGrad：通过数据依赖自适应先验改进条件去噪扩散模型。载于：国际学习表征会议（ICLR）论文集（2022）4 30. Lindell, D.B., Wetzstein, G., Koltun, V.: 声学非视线成像。载于：IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集。第 6773–6782 页（2019）2 31. Lindell, D.B., Wetzstein, G., O'Toole, M.: 使用快速FK迁移的基于波的非视线成像。《ACM图形学汇刊》38(4)，1–13（2019）2 32. Ling, J., Wang, Z., Xu, F.: ShadowNeUS：通过阴影射线监督进行神经SDF重建。载于：IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集。第 175–185 页（2023）33. Liu, R., Wu, R., Hoorick, B.V., Tokmakov, P., Zakharov, S., Vondrick, C.: Zero-1-to-3：零样本单图像转3D物体。arXiv预印本 arXiv:2309.16653（2023）4 34. Liu, X., Bauer, S., Velten, A.: 基于相量场衍射的快速非视线成像系统重建。《自然·通讯》11(1)，1645（2020）2

35. Liu, X., Guillén, I., La Manna, M., Nam, J.H., Reza, S.A., Huu Le, T., Jarabo, A., Gutierrez, D., Velten, A.：基于相量场虚拟波动光学的非视域成像。《自然》572 2 36 7771 620 623 2019 ( )，– ( )。Liu, X., Wang, J., Xiao, L., Fu, X., Qiu, L., Shi, Z.：利用信号表面协同正则化进行少样本非视域成像。载于：IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集。第13303 2 37 13312页，2022 3– ( )。Lorensen, W.E., Cline, H.E.：行进立方体：一种高分辨率三维表面构建算法。载于：SIGGRAPH。第163 11 38 169页，1987 3–。ACM ( )。Luo, S., Hu, W.：用于三维点云生成的扩散概率模型。载于：IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集（2021）7 39. Maeda, T., Wang, Y., Raskar, R., Kadambi, A.：热非视域成像。载于：IEEE计算摄影国际会议（ICCP）论文集。第1–11页。IEEE（2019 2 40）。Medin, S.C., Weiss, A., Durand, F., Freeman, W.T., Wornell, G.W.：阴影能否揭示生物特征信息？载于：IEEE/CVF冬季计算机视觉应用会议（WACV）论文集。第869 3 41 879页，2023 –（1月）。Metzler, C.A., Heide, F., Rangarajan, P., Balaji, M.M., Viswanath, A., Veeraraghavan, A., Baraniuk, R.G.：深度逆相关成像：迈向实时高分辨率非视域成像。《光学》7 2 42 1 63 71 2020 ( )，– ( )。Murray-Bruce, J., Saunders, C., Goyal, V.K.：利用消费级相机进行基于遮挡的计算潜望成像。载于：小波与稀疏性 XVIII。第11138卷，第286 2 43 297页，2019 –。SPIE ( )。Nichol, A., Jun, H., Dhariwal, P., Mishkin, P., Chen, M.：Point-e：一种根据复杂提示生成三维点云的系统（2022），https://arxiv.org/abs/2212.08751 11 44. O'Toole, M., Lindell, D.B., Wetzstein, G.：基于光锥变换的共焦非视域成像。《自然》555 2 45 7696 338 341 2018 ( )，– ( )。Pawlikowska, A.M., Halimi, A., Lamb, R.A., Buller, G.S.：在最远10公里距离进行单光子三维成像。《光学快报》25(10)，11919–11931 2 46 2017( )。Poole, B., Jain, A., Barron, J.T., Mildenhall, B.：DreamFusion：使用二维扩散进行文本到三维生成。arXiv预印本 arXiv：2 4 47 2309 16653 2022。（）。Popov, V., Vovk, I., Gogoryan, V., Sadekova, T., Kudinov, M.：Grad-TTS：一种文本转语音扩散概率模型。载于：ICML（2021 4 48 3）。Raji, F., Murray-Bruce, J.：利用普通相机进行三维计算潜望成像：一种可分离非线性最小二乘公式。载于：IEEE声学、语音与信号处理国际会议（ICASSP）论文集。第7475–7479页（2024）。https://doi.org/10 10446230 6 49 1109 48485 2024 . /ICASSP . . . Ramachandran, V.S.：明暗形状知觉。《自然》331，163–166（1988 3 50）。Rapp, J., Saunders, C., Tachella, J., Murray-Bruce, J., Altmann, Y., Tourneret, J.Y., McLaughlin, S., Dawson, R.M., Wong, F.N., Goyal, V.K.：利用边缘分辨瞬态成像实现绕角成像。《自然·通讯》11(1)，5929（2020）2 51. Rasul, K., Seward, C., Schuster, I., Vollgraf, R.：用于多变量概率时间序列预测的自回归去噪扩散模型。载于：机器学习国际会议（ICML）论文集（2021 4 52）。Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.：利用潜在扩散模型进行高分辨率图像合成。载于：IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集。第10684–10695页（2022年6月）4

53. Ronneberger, O., Fischer, P., Brox, T.：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（MICCAI）。第234–241页。计算机科学讲义，施普林格国际出版（2015 11 54）。Saunders, C., Bose, R., Murray-Bruce, J., Goyal, V.K.：利用普通相机进行多深度计算潜望成像。载于：IEEE声学、语音与信号处理国际会议（ICASSP）论文集。第9299 2, 3, 5 55 9305页，2020– ( )。Saunders, C., Murray-Bruce, J., Goyal, V.K.：利用普通数码相机进行计算潜望成像。《自然》565 2, 3, 4, 5, 6, 10, 11 56 7740 472 475 2019 ( )，– ( )。Seidel, S., Rueda-Chacón, H., Cusini, I., Villa, F., Zappa, F., Yu, C., Goyal, V.K.：主动式角落相机的非视线快照与背景映射。《自然·通讯》14 2 57 1 3677 2023 ( )，( )。Seidel, S.W., Ma, Y., Murray-Bruce, J., Saunders, C., Freeman, W.T., Christopher, C.Y., Goyal, V.K.：角落遮挡物计算潜望成像：从单张照片估计隐藏场景。载于：IEEE计算摄影国际会议（ICCP）论文集。第1 2 58 9页，2019 –。IEEE ( )。Seidel, S.W., Murray-Bruce, J., Ma, Y., Yu, C., Freeman, W.T., Goyal, V.K.：利用单个边缘遮挡物进行二维非视域场景估计。《IEEE计算成像汇刊》7，58 2 59 72 2021– ( )。Sharma, P., Aittala, M., Schechner, Y.Y., Torralba, A., Wornell, G.W., Freeman, W.T., Durand, F.：凝视空白墙面能让你学到什么。载于：IEEE/CVF国际计算机视觉会议（ICCV）论文集。第2330–2339页（2021年10月）3 60. Shim, J., Kang, C., Joo, K.：用于三维形状生成的基于扩散的有符号距离场。载于：IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集。第20887 7 61 20897页，2023 –（6月）。Sohl-Dickstein, J., Weiss, E.A., Maheswaranathan, N., Ganguli, S.：利用非平衡热力学进行深度无监督学习。载于：机器学习国际会议（ICML）论文集。第2256页，4, 7 62 2265 2015– ( )。Somasundaram, S., Dave, A., Henley, C., Veeraraghavan, A., Raskar, R.：瞬态信号在双反射非视线成像中的作用。载于：IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集。第9192 2 63 9201页，2023– ( )。Song, Y., Ermon, S.：通过估计数据分布的梯度进行生成建模。《神经信息处理系统进展》32 4, 8 64 2019( )。Torralba, A., Freeman, W.T.：意外针孔与针孔式遮挡物相机：揭示画面之外的场景。《国际计算机视觉杂志》110 6 65 2 92 112 2014 ( )，– ( )。Velten, A., Willwacher, T., Gupta, O., Veeraraghavan, A., Bawendi, M.G., Raskar, R.：利用超快飞行时间成像恢复角落后方的三维形状。《自然·通讯》3 2 66 1 745 2012 ( )，( )。Verlekar, T.T., Soares, L.D., Correia, P.L.：利用阴影轮廓进行野外步态识别。《图像与视觉计算》76，1 3 67 13 2018– ( )。Wang, W., Yao, L., Chen, L., Lin, B., Cai, D., He, X., Liu, W.：CrossFormer：一种基于跨尺度注意力的多功能视觉变换器。载于：学习表征国际会议（ICLR）论文集（2022 11, 28 68）。Wang, Y., Zhang, Y., Huang, M., Chen, Z., Jia, Y., Weng, Y., Xiao, L., Xiang, X.：准确但脆弱的被动式非视线识别。《通讯·物理》4 2, 11 69 1 88 2021 ( )，（12月）。Yamashita, Y., Sakaue, F., Sato, J.：根据非平面阴影恢复三维形状和光源位置。载于：IEEE/CVF国际模式识别会议（ICPR）论文集。第1775–1779页（2010）4

70. Yang, B., Gu, S., Zhang, B., Zhang, T., Chen, X., Sun, X., Chen, D., Wen, F：示例绘画：使用扩散模型进行基于范例的图像编辑。载于：IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集。第18381–18391页4 712023() 。Yang, G., Huang, Y., Hao, Z., Liu, M., Belongie, S., Hariharan, B：Pointflow：使用连续归一化流进行三维点云生成。载于：IEEE/CVF国际计算机视觉会议（ICCV）论文集（20194 72) 。Ye, J., Wang, P., Li, K., Shi, Y., Wang, H：Consistent-1-to-3：通过几何感知扩散模型实现一致的图像到三维视图合成。arXiv预印本 arXiv:23094 73 16653 2024. ( ) 。Yedidia, A.B., Baradad, M., Thrampoulidis, C., Freeman, W.T., Wornell, G.W：利用未知遮挡物恢复隐藏场景。载于：IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集。第12231–12239页（2019）2, 3,5, 11 74. Yedidia, A.B., Baradad, M., Thrampoulidis, C., Freeman, W.T., Wornell, G.W：利用未知遮挡物恢复隐藏场景。载于：IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集。第1223111 75 12239 2019– ( ) 。Zeng, X., Vahdat, A., Williams, F., Gojcic, Z., Litany, O., Fidler, S., Kreis, K：Lion：用于三维形状生成的潜在点扩散模型。载于：神经信息处理系统进展（NeurIPS）（20224 76) 。Zeng, X., Vahdat, A., Williams, F., Gojcic, Z., Litany, O., Fidler, S., Kreis, K：Lion：用于三维形状生成的潜在点扩散模型。载于：神经信息处理系统进展（NeurIPS）。第100217 77 10039 2022 3 – ( ) 。Zhou, L., Du, Y., Wu, J：通过点体素扩散进行三维形状生成与补全。载于：IEEE/CVF国际计算机视觉会议（ICCV）论文集。第5806–5815页（2021）4,11

《软阴影扩散（SSD）：受物理启发的三维计算潜望成像学习》的补充材料

FadlullahRaji John Murray Bruce

南佛罗里达大学。美国佛罗里达州坦帕市 E. Fowler Avenue 4202 号，邮编 33620。
fraji@usf.edu murraybruce@usf.edu

本文的补充文档提供了有关纯物理方法复杂度的更多信息（第 A 节），并介绍了该方法对未知环境照明的鲁棒性（第 B 节）。第 C 节详细介绍了物理启发式软阴影扩散模型的实现与训练。最后，第 D 节展示了更多真实实验结果和合成实验结果，并探讨了 SSD模型在噪声和背景照明下的稳定性，同时展示了更多三维重建结果。

一种基于物理的三维计算潜望成像：迈向更高分辨率的体素化

正文中的图5(b)展示了使用 $10 \times 5 \times 10$ 体素离散化遮挡区域所得的真实实验数据重建结果。（请注意，我们所说的隐藏场景中的“遮挡区域”在正文图2中有示意。）

A.1 基于物理的重建复杂度分析

这种 $10 \times 5 \times 10$ 针孔散斑体素化较为粗糙，无法描述种类繁多的三维物体。增加针孔散斑的数量会提高计算复杂度和空间需求。主文稿中的图5(a)展示了增大网格尺寸的示例，我们使用 $16 \times 5 \times 16$ 对遮挡区域进行离散化来重建该场景。给定一张 128×128的相机照片，以及需要以 32×32分辨率估计的非遮挡场景，使用 Float32 精度时，内存需求为 $16384 \times 1024 \times 1280 = 80$GB。由于我们的算法需要计算梯度并更新待估计变量，内存需求会翻倍。因此，为了实现主文稿图5(a)中的模拟重建，我们使用了3块各具备80GB内存的 A100 GPU，并将前向模型分布到这些 GPU 上进行计算。接下来将介绍该前向模型的一种稀疏表示。

A.2 稀疏矩阵表示

隐藏场景中的点光源所产生的隐藏场景针孔式遮挡物的贡献（阴影），可以用$\mathbf{V}_k$的稀疏结构来解释。如下图1所示，该贡献在少数位置为零，在其他所有位置均为一。更准确地说，该贡献的按位取反是稀疏的。因此，为了实现更高分辨率，无需保存大型矩阵$\mathbf{V}_k$（其对应于每个以 k为索引的针孔散斑位置），即可计算主文稿公式（10）中的离散前向模型。相反，可以使用稀疏数据结构来提高效率。

采用稀疏表示在更高重建分辨率下评估针孔式遮挡物前向模型的结果如图1所示。尽管我们能够在仿真中正确完成重建，但准确的真实世界重建仍然具有挑战性。这部分是由于针孔式遮挡物近似所致。尽管存在这一不足，这种基于物理的表述仍提供了一个灵活的框架，使我们能够采用受物理启发的神经网络方法。正如主论文（图5）和补充图12（此处所示）中的结果所展示的，我们的SSD模型即使面对真实实验数据集，也能处理更高的离散化。

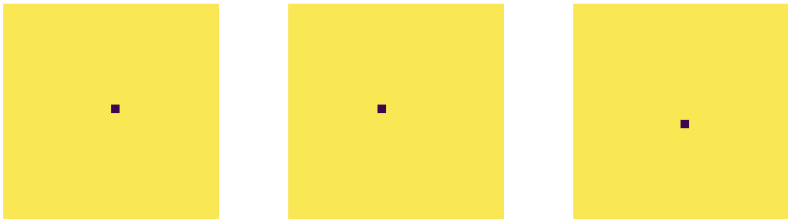


图1： 每个面板展示了对于隐藏场景发光二维平面中的某个面片和隐藏场景针状散射体体素，可见墙面/相机视场范围内可见性函数的变化。可见性函数取二值，其中1（黄色）表示隐藏场景发光二维平面中的面片相对于可见墙面上的这些面片未被遮挡，0（深蓝色）表示被针状散射体体素遮挡。

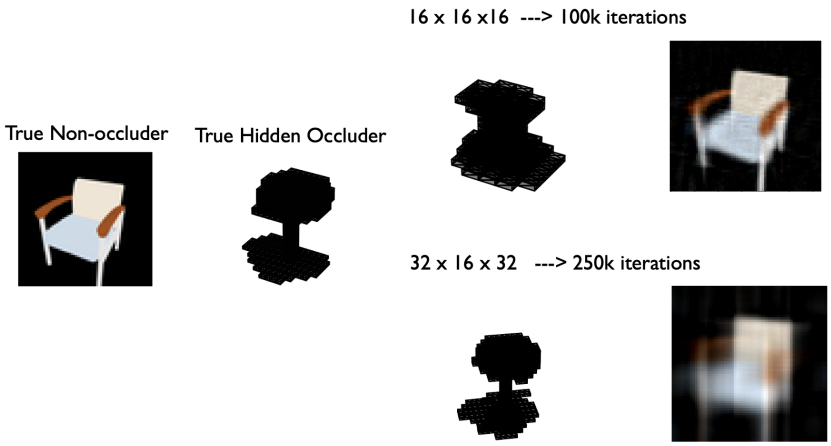


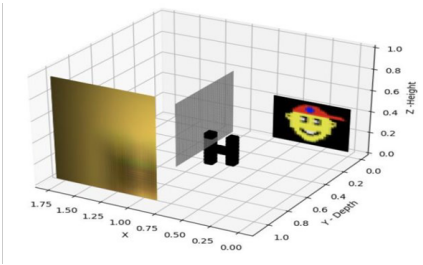
图2： 使用稀疏数据结构进行仿真的高分辨率重建。

B 基于梯度的反演：忽略背景与建模背景贡献

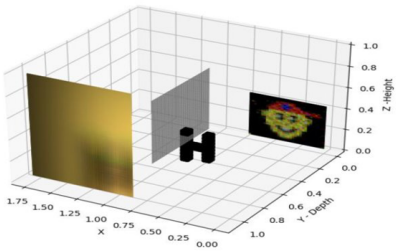
我们在图3 和4中比较了通过交替最小化进行的基于梯度的反演，以及一种将背景贡献优化纳入其中的变体。结果表明，未学习噪声和背景贡献时得到的解较为模糊。优化背景噪声效率更高，并能生成清晰的非遮挡结构图像。

算法 2 忽略背景的交替最小化方法

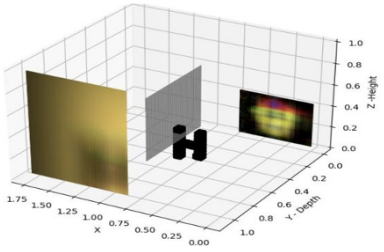
要求：A, $[V_1, V_2, V3, ..., V_K]$, y,numIter 1：随机初始化 z_0 ，以及 $\lambda_0 = 12$：初始化步长 η_z, 以及 $\eta_\lambda 3$：对于 $i = 1$ 至 numIter 执行 4：$A_v = 1 - \frac{1}{K}\sum_{k=1}^{K} V_k \cdot \sigma(z^{ik})$ 5：$f^* = (A_v{}^T A_v + \lambda_{i-1} I)^{-1} A_v{}^T y$ 6：$f_i = \max(f^*, 0)$ 7：$\mathcal{L}(y, A_v f_i) = \frac{1}{M}\sum_{j=1}^{M} \|y^j - (A_v f_i)_j\|_2^2 8$：$z_i \leftarrow z_{i-1} - \eta_z \frac{\partial \mathcal{L}(\mathbf{y}, \mathbf{A}_v \mathbf{f}_i)}{\partial z_{i-1}} 9$：$\lambda_i \leftarrow \lambda_{i-1} - \eta_\lambda \frac{\partial \mathcal{L}(\mathbf{y}, \mathbf{A}_v \mathbf{f}_i)}{\partial \lambda_{i-1}} 10$：结束循环 11：返回 $\hat{f} = f_i, \hat{\alpha} = \sigma(z_i) > 0.5$



(a) 真实场景示意图。



(b) 联合估计背景贡献。



(c) 不估计背景。

图 3：通过交替最小化进行基于梯度的反演所得的重建。面板 (b) 和 (c) 展示了三维遮挡结构和二维发光场景的重建，其中 (b) 纳入了背景建模与估计，而 (c) 未纳入。

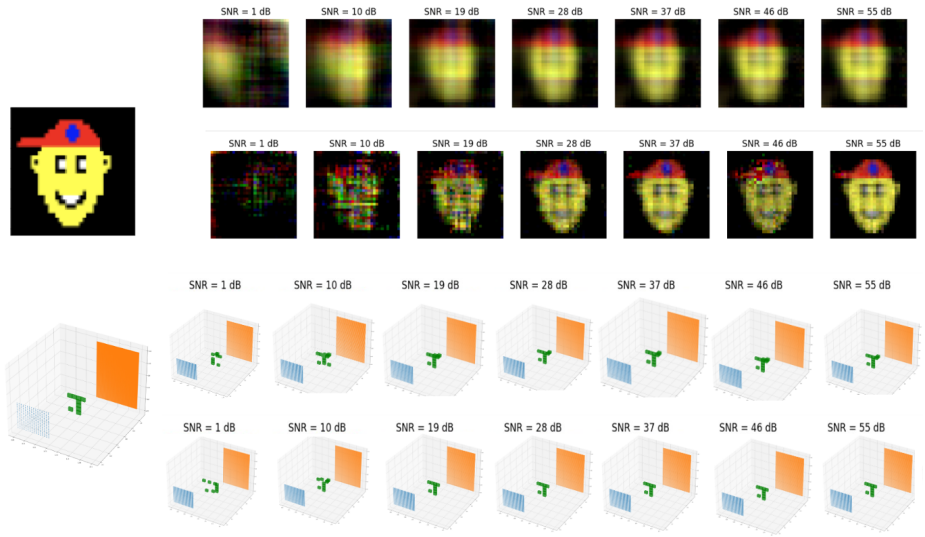


图 4：比较基于梯度的隐藏场景重建算法的两种变体。第 1 行：忽略背景贡献 **b** 时发光隐藏场景组件的二维重建。第 2 行：对背景贡献 **b** 进行建模和估计时发光隐藏场景组件的二维重建。第 3 行：忽略背景贡献 **b** 时遮光隐藏场景组件的三维重建。第 4 行：对背景贡献 **b** 进行建模和估计时遮光隐藏场景组件的三维重建。

C SoftShadow 扩散模型

将针孔点云记为 $\{\mathbf{u}_k\}_{k=1}^K = \hat{\mathbf{u}}$。联合分布 $p_\theta(\hat{\mathbf{u}}^{(0:T)})$ 称为反向过程。该过程可描述为一个马尔可夫链，其起始于 $p(\hat{\mathbf{u}}^{(T)}) = \mathcal{N}(\hat{\mathbf{u}}^{(T)}; \mathbf{0}, \mathbf{I})$，并具有经学习得到的高斯转移：

$$p_\theta(\hat{\mathbf{u}}^{(0:T)}) := p(\hat{\mathbf{u}}^{(T)}) \prod_{t=1}^{T} p_\theta(\hat{\mathbf{u}}^{(t-1)}|\hat{\mathbf{u}}^{(t)}), \tag{1}$$

$$p_\theta(\hat{\mathbf{u}}^{(t-1)}|\hat{\mathbf{u}}^{(t)}) := \mathcal{N}(\hat{\mathbf{u}}^{(t-1)}; \mu_\theta(\hat{\mathbf{u}}^{(t)}, t), \sigma^2\mathbf{I}). \tag{2}$$

近似后验分布 $q(\hat{\mathbf{u}}^{1:T}|\hat{\mathbf{u}}^0)$（称为前向过程）定义为一个马尔可夫链，它按照预先设定的方差调度 $\beta_1, \ldots, \beta_T$，逐步向数据中加入高斯噪声：

$$q(\hat{\mathbf{u}}^{(1:T)}|\hat{\mathbf{u}}^0) := \prod_{t=1}^{T} q(\hat{\mathbf{u}}^{(t)}|\hat{\mathbf{u}}^{(t-1)}), \tag{3}$$

$$q(\hat{\mathbf{u}}^{(t)}|\hat{\mathbf{u}}^{(t-1)}) := \mathcal{N}(\hat{\mathbf{u}}^{(t)}); \sqrt{1-\beta_t}\hat{\mathbf{u}}^{(t-1)}, \beta_t\mathbf{I}) \tag{4}$$

令 $\hat{\mathbf{u}} = \{\mathbf{u}_k\}_{k=1}^K$，则训练目标为：

$$\begin{aligned} L(\theta,\phi) = \mathbb{E}_q \Bigg[& \sum_{t=2}^{T}\sum_{k=1}^{K} \underbrace{\mathrm{D}_{\mathrm{KL}}\big(q(\mathbf{u}_k^{(t-1)}|\mathbf{u}_k^{(t)}, \mathbf{u}_k^{(0)}) \| p_\theta(\mathbf{u}_k^{(t-1)}|\mathbf{u}_k^{(t)}, \hat{\mathbf{y}})\big)}_{L_k^{(t-1)}} \\ & - \sum_{k=1}^{K} \underbrace{\log p_\theta(\mathbf{u}_k^{(0)}|\mathbf{u}_k^{(1)}, \hat{\mathbf{y}})}_{L_k^{(0)}} \\ & + \underbrace{\mathrm{D}_{\mathrm{KL}}\big(q_\phi(\hat{\mathbf{y}}|\mathbf{u}^{(0)}) \| p(\hat{\mathbf{y}})\big)}_{L_{\hat{\mathbf{y}}}} \Bigg]. \end{aligned} \tag{5}$$

C.1 简化训练算法

我们采用 [20]中提出的简化算法来训练模型。为评估 $L_k^{(t-1)}$，我们需要从 $\mathbf{u}_k^{(t)}$ 中采样 $q(\hat{\mathbf{u}}^{(t)}|\hat{\mathbf{u}}^{(0)})$。原则上，可以通过沿马尔可夫链迭代采样来实现。不过，[20] 指出 $q(\hat{\mathbf{u}}^{(t)}|\hat{\mathbf{u}}^{(0)})$ 服从高斯分布，因此我们可以高效地采样 $\hat{\mathbf{u}}^{(t)}$，无需迭代采样。为说明这一点，请注意：

$$q(\hat{\mathbf{u}}^{(t)}|\hat{\mathbf{u}}^{(0)}) = \mathcal{N}(\hat{\mathbf{u}}^{(t)}|\sqrt{\bar{\alpha}_t}\hat{\mathbf{u}}^{(0)}, (1-\bar{\alpha}_t)I). \tag{6}$$

由于 $q(\mathbf{u}_k^{(t-1)}|\mathbf{u}_k^{(t)},\mathbf{u}_k^{(0)})$ 和 $p_\theta(\mathbf{u}_k^{(t-1)}|\mathbf{u}_k^{(t)},\hat{\mathbf{y}})$ 都服从高斯分布，因此项 $L_k^{(t-1)}$ 可以展开为：

$$L_k^{(t-1)} = \mathbb{E}_{\mathbf{u}_k^{(0)},\mathbf{u}_k^{(t)},\hat{\mathbf{y}}}\left[\frac{1}{2\beta_t}\left\|\frac{\sqrt{\bar{\alpha}_{t-1}\beta_t}}{1-\bar{\alpha}_t}\mathbf{u}_k^{(0)} + \frac{\sqrt{\alpha_t(1-\bar{\alpha}_{t-1})}}{1-\bar{\alpha}_t}\mathbf{u}_k^{(t)} - \mu_\theta(\mathbf{u}_k^{(t)},t,\hat{\mathbf{y}})\right\|^2\right] + C. \tag{7}$$

使用上述高斯分布，$\mathbf{u}_k^{(t)} = \sqrt{\bar{\alpha}_t}\mathbf{u}_k^{(0)} + \sqrt{1-\bar{\alpha}_t}\epsilon$，其中 $\epsilon \sim \mathcal{N}(\mathbf{0}, I)$：

$$L_k^{(t-1)} = \mathbb{E}_{\mathbf{u}_k^{(0)},\epsilon,\hat{\mathbf{y}}}\left[\frac{1}{2\beta_t}\left\|\frac{1}{\sqrt{\alpha_t}}\left(\mathbf{u}_k^{(t)} - \beta_t\sqrt{1-\bar{\alpha}_t}\epsilon\right) - \mu_\theta(\mathbf{u}_k^{(t)},t,\hat{\mathbf{y}})\right\|^2\right] + C. \tag{8}$$

如上所示，$\mu_\theta(\mathbf{u}_k^{(t)},t,\hat{\mathbf{y}})$ 必须根据 $\mathbf{u}_k^{(t)}$ 预测 $\frac{1}{\sqrt{\alpha_t}}\left(\mathbf{u}_k^{(t)} - \beta_t\sqrt{1-\bar{\alpha}_t}\epsilon\right)$。因此，$\mu_\theta(\mathbf{u}_k^{(t)},t,\hat{\mathbf{y}})$ 可参数化为：

$$\mu_\theta(\mathbf{u}_k^{(t)},t,\hat{\mathbf{y}}) = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{u}_k^{(t)} - \beta_t\sqrt{1-\bar{\alpha}_t}\epsilon_\theta(\mathbf{u}_k^{(t)},t,\hat{\mathbf{y}})\right), \tag{9}$$

其中，$\epsilon_\theta(\mathbf{u}_k^{(t)},t,\hat{\mathbf{y}})$ 是受物理启发的神经网络，旨在根据 $\mathbf{u}_k^{(t)}$ 预测 ϵ。最后，$L_k^{(t-1)}$ 可简化为

$$L_k^{(t-1)} = \mathbb{E}_{\mathbf{u}_k^{(0)},\epsilon,\hat{\mathbf{y}}}\left[\frac{\beta_t^2}{2\beta_t\alpha_t(1-\bar{\alpha}_t)}\left\|\epsilon - \epsilon_\theta(\sqrt{\bar{\alpha}_t}\mathbf{u}_k^{(0)} + \sqrt{1-\bar{\alpha}_t}\epsilon, t, \hat{\mathbf{y}})\right\|^2\right] + C. \tag{10}$$

为了最小化 $L_i^{(t-1)}$，我们只需最小化 $\mathbb{E}[\|\epsilon - \epsilon_\theta\|^2]$，因为系数 $\frac{\beta_t^2}{2\beta_t\alpha_t(1-\bar{\alpha}_t)}$ 是常数。

简化算法 [20] 提出在每一步 $\left\{\sum_{k=1}^K L_k^{(t-1)}\right\}_{t=1}^T$ 训练过程中随机从集合中选择一个项进行优化。

随后，训练算法如算法 3所示。

算法3 SSD训练算法

repeat
  Sample the occluding components $\{\mathbf{u}_k^{(0)}\}_{k=1}^K \sim p_{\text{data}}(\mathbf{u}^{(0)})$
  Sample the non-occluding components $\mathbf{f} \in \mathbb{R}^N \sim p_{\text{data}}(\mathbf{f})$
  Generate the measurement $\mathbf{y} = \mathbf{A}\left(1 - \sum_{k=1}^K \mathbf{u}_k^{(0)}\right)\mathbf{f}$
  Sample $\hat{\mathbf{y}} \sim q_\phi(\mathbf{y}|\hat{\mathbf{u}}^{(0)})$
  Sample $t \sim \text{Uniform}(\{1,\ldots,T\})$
  Sample $\epsilon \sim \mathcal{N}(0, I)$
  Compute $\nabla\left[\sum_{k=1}^K \left\|\epsilon - \epsilon_\theta\left(\sqrt{\bar{\alpha}_t}\mathbf{u}_k^{(0)} + \sqrt{1-\bar{\alpha}_t}\epsilon, t, \hat{\mathbf{y}}\right)\right\|^2\right]$;
  Perform gradient descent.
until converged

C.2 推理算法

给定$\mathbf{y}$，根据算法4使用 SSD 进行推理。

算法4 SSD 推理算法

1：给定软阴影照片 $\mathbf{y}$
2：对照片进行编码：$\hat{}\mathbf{y} \leftarrow F_{\alpha}(\mathbf{y})$
3：从高斯分布中采样噪声 $\{\mathbf{u}_k^{(T)}\}_{k=1}^K \leftarrow \hat{}\mathbf{u}^{(T)} \sim \mathcal{N}(0, I)$
4: **对于** $t = T \ldots, 1$ **do**
5:　　去噪
6:　　采样 $\{\mathbf{u}_k^{(t-1)}\}_{k=1}^K \sim p_\theta(\{\mathbf{u}_k^{(t-1)}\}_{k=1}^K | \{\mathbf{u}_k^{(t)}\}_{k=1}^K, \hat{}\mathbf{y})$
7: </style}结束循环
8: **返回** $\{\mathbf{u}_k^{(0)}\}_{k=1}^K$

C.3 实现细节

软阴影图像编码器。我们的编码器架构遵循CrossFormer的设计 [67]，包含数量分别为（2、4、8、2）的Transformer模块，如表 1所示。

数据集准备。我们模拟了262,000条测量数据，并采样了相应的点云。从中随机选取3,000个样本作为测试集，其余259,000个样本用于训练集。

扩散过程。扩散过程的步数为 $T = 256$。我们采用余弦调度设置方差调度 β_t 。

训练参数如表2所示。

基于***U-Net***的扩散模型。模型架构基于U-Net设计，如图 5所示。该架构特别适合处理点云数据，因为它能够高效地处理空间层次结构和特征表示。在我们的实现中，模型在形状为（2048，3）的点云张量上进行训练，表示三维空间中的一组2048个点。

为了增强模型重建精细点云的能力，我们采用一种能够生成更密集点云的采样策略。具体而言，在重建阶段，我们采样更多点，目标数量为 15,000 个。这种方法能够生成更精细、更密集的重建遮挡结构表示，捕捉对于高保真网格表示至关重要的细微差异和结构。

表1：用于软阴影图像编码的CrossFormer模型（CrossFormer-T）

阶段	输出尺寸	层名称	CrossFormer-T
阶段1	32×32	交叉嵌入 SDA/LDA/多层感知机	卷积核大小：$\begin{bmatrix} 4\times4 \\ 8\times8 \\ 16\times16 \\ 32\times32 \\ Stride=4 \end{bmatrix}$ $\begin{bmatrix} D1=64 \\ H1=2 \\ G1=7 \\ I1=8 \end{bmatrix} \times 2$
阶段2	16×16	交叉嵌入 SDA/LDA/多层感知机	卷积核大小：4×4，步幅=4（S1 = 56） $\begin{bmatrix} D2=128 \\ H2=4 \\ G2=7 \\ I2=4 \end{bmatrix} \times 4$
阶段3	8×8	交叉嵌入 SDA/LDA/多层感知机	卷积核大小：2×2，步幅=2(S2 = 28) $\begin{bmatrix} D3=256 \\ H3=8 \\ G3=7 \\ I3=2 \end{bmatrix} \times 8$
阶段4	4×4	交叉嵌入 SDA/LDA/多层感知机	卷积核大小: 2×2, 步幅=2(S3 = 14) $\begin{bmatrix} D4=512 \\ H4=16 \\ G4=7 \\ I4=1 \end{bmatrix} \times 2$
		交叉嵌入	卷积核大小: 2×2, 步幅=2(S4 = 7)
Head	1×1	平均池化卷积核大小: 4×4	
		线性潜变量：512	

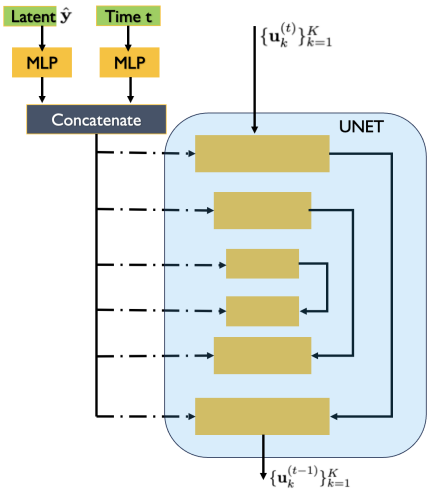


图 5： SSD UNet模块

表 2： SSD训练详情

参数	值
基础通道数	128
优化器	Adam
通道倍数	1, 2, 4,8
学习率	1×10^{-4}
每个分辨率的模块数	2
批量大小	192
注意力分辨率	256, 128
EMA	0.9999
注意力头数	8
随机失活	0.0
条件嵌入维度	512
训练硬件	3 × A100(80G)
条件嵌入多层感知机层数	1
时间嵌入多层感知机层数	1
训练迭代次数	600000
扩散噪声调度	余弦
权重衰减	0.01
采样时间步	256
Loss	均方误差（L2）

D 结果

D.1 背景光与失真

为了研究背景和噪声对我们方法的影响，我们在实验装置中加入了一个受控光源。定量结果展示于主论文的图 6(b) 中。图 6 展示了该实验的定性结果。我们还在实际测量上模拟了一个噪声过程，结果展示于图7中。模拟示例见图 8 和9

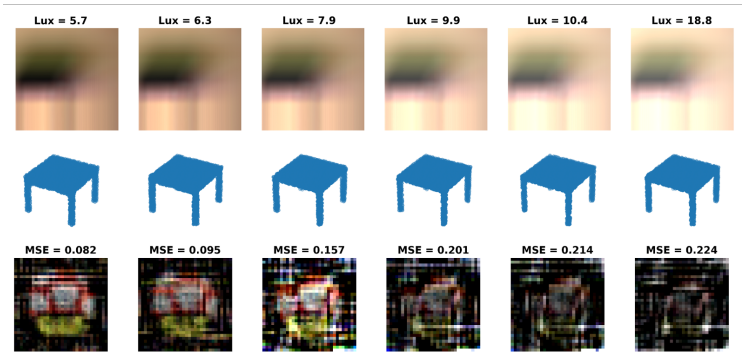


图 6： 变化的背景光下的实验结果

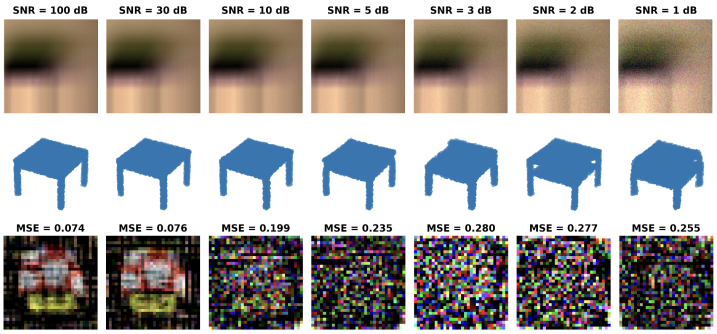


图 7： 不同信噪比（SNR）下的实验结果

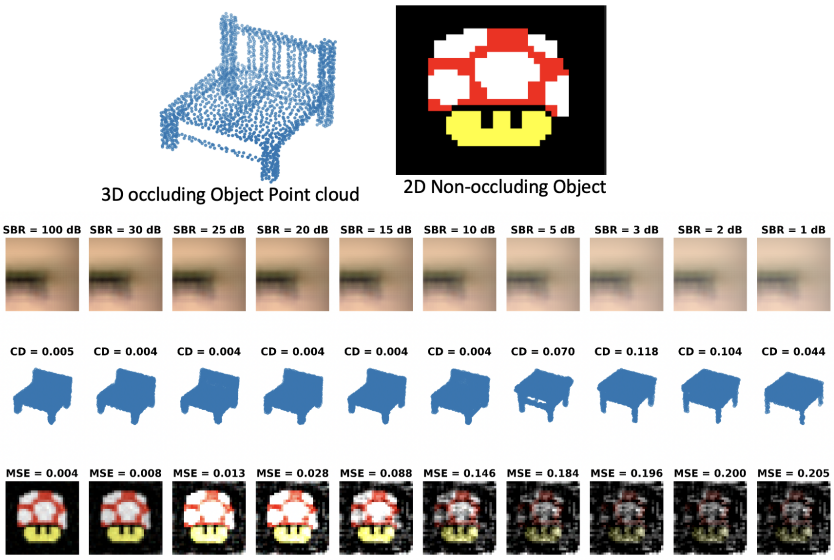


图 8： 信号背景比（SBR）变化时的模拟结果

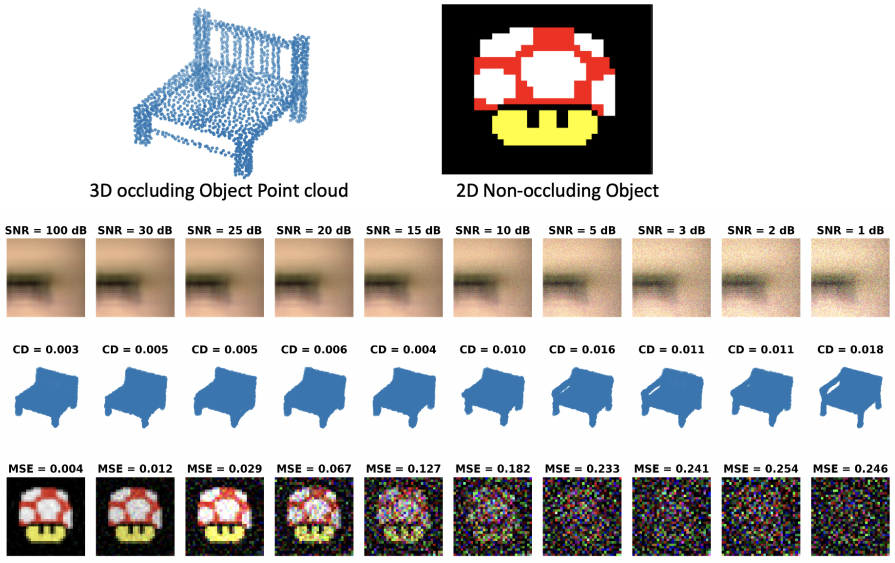


图 9： 信噪比（SNR）变化时的模拟结果

D.2 形状重建

此处提供了更多结果，以展示所提出的 SSD 模型在多种形状类别上的表现。我们从训练数据集之外采样了 64 个网格，并使用随机的二维非遮挡物体仿真半影照片。该实验结果如图10所示。

我们再次说明，该模型是在仿真环境中训练的。仿真实验数据的结果见图11和13 至15，基于实际测量数据的更多结果见图12。

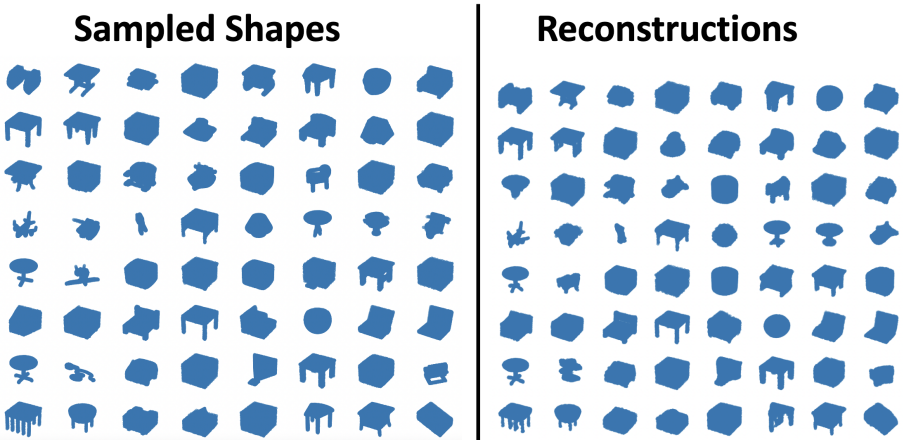


图 10：训练数据集之外重建的形状。

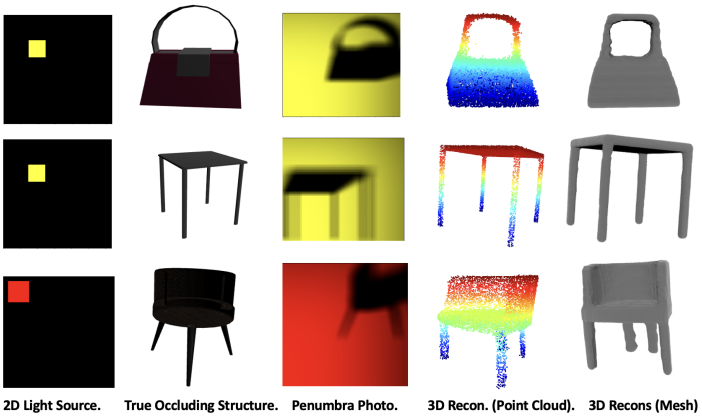


图 11：基于硬阴影的 SSD 重建。

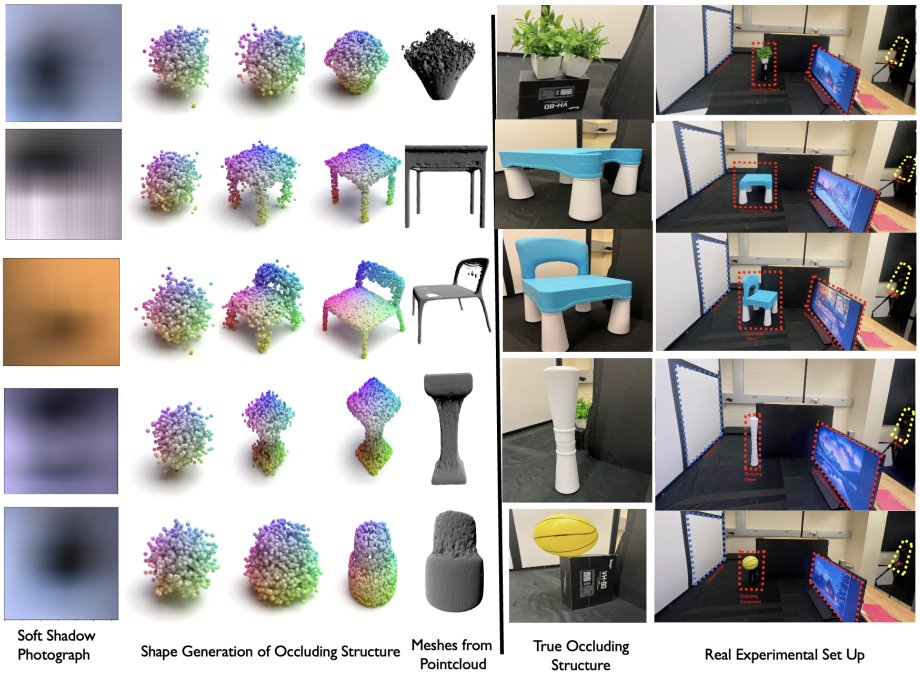


图12：SSD 在真实实验设置中的结果。

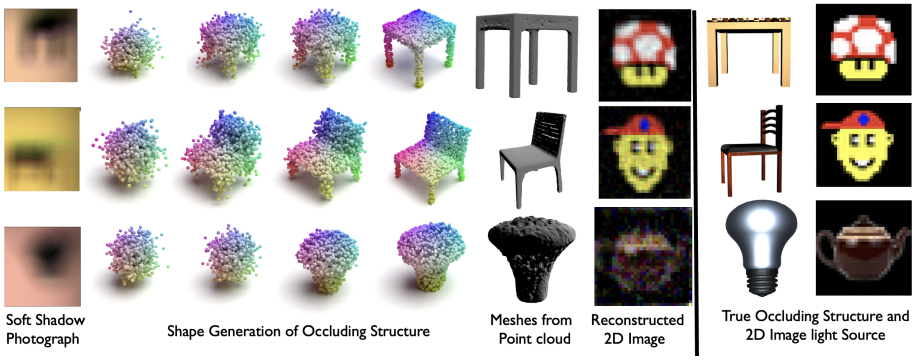


图13：SSD 在仿真中的结果。

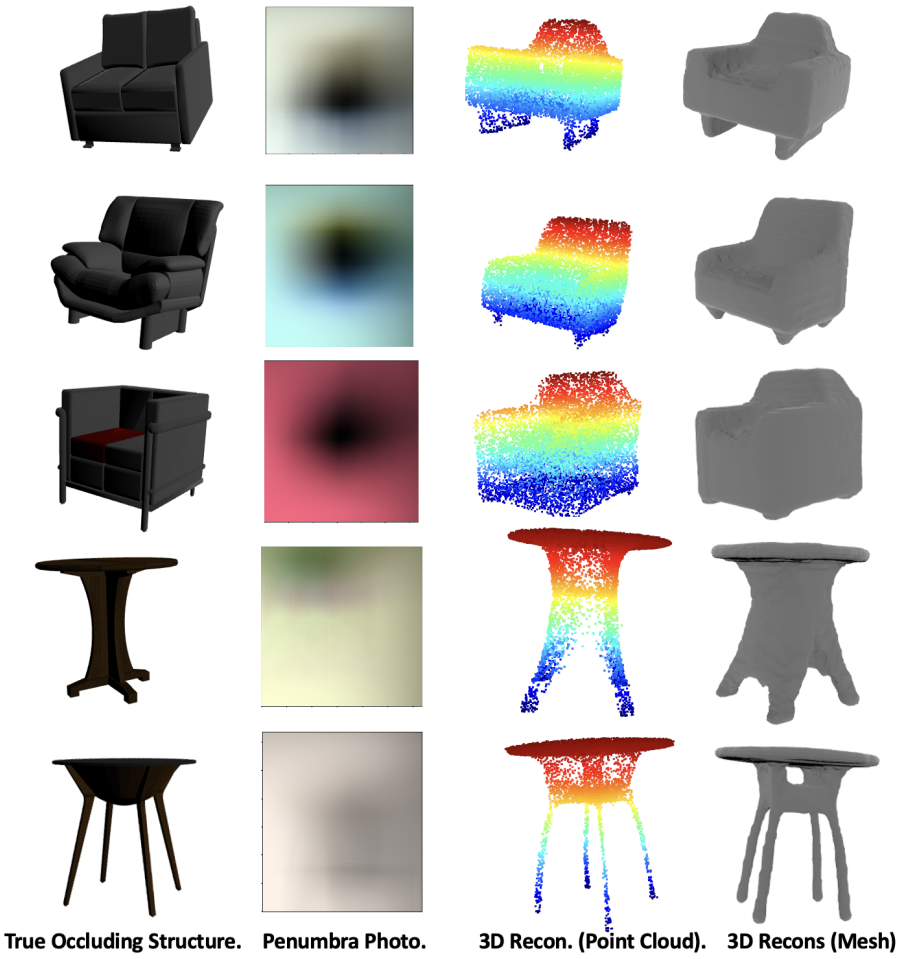


图14： SSD 在仿真中的结果。

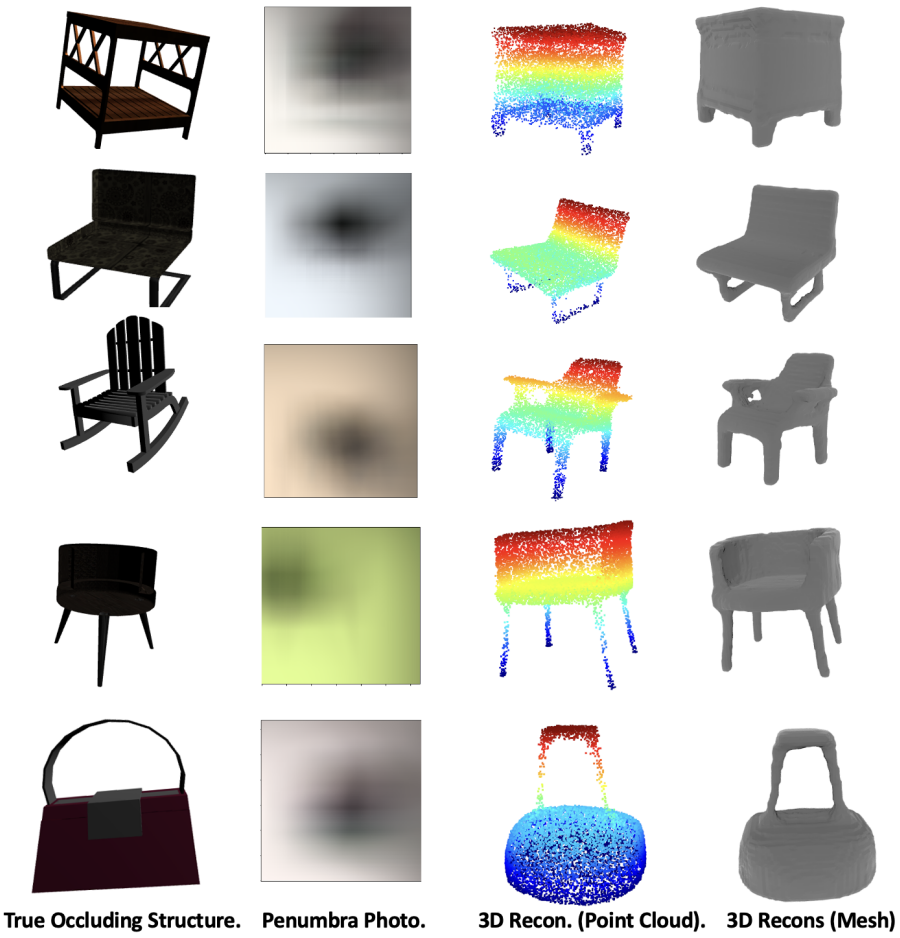

True Occluding Structure.
Penumbra Photo.
3D Recon. (Point Cloud).
3D Recons (Mesh)

图15： SSD 仿真结果。