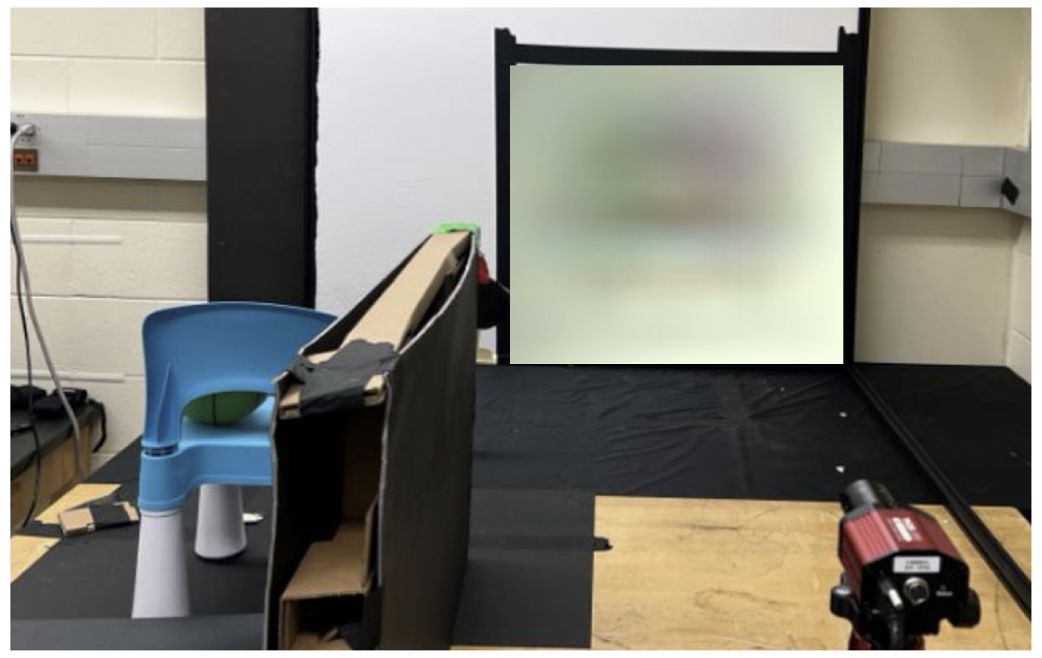
Soft Shadow Diffusion (SSD): Physics-inspired Learning for 3D Computational Periscopy

Fadlullah Raji, and John Murray-Bruce

University of South Florida

Can we see (in 3D) by looking at soft shadows?



We demonstrate **passive 3D non-line-of-sight** (NLOS) imaging—or **3D Computational Pericopy**—from photographs of soft shadows cast on a visible surface.

Prior approaches to 3D NLOS imaging:

- *Active methods:* Rely on time-resolved measurements of the hidden scene in response to ultrashort laser pulses. They have long acquisition times and use expensive and complex optical hardware setups.
- *Passive methods:* Rely on shadows of *known* simple occluders (e.g., a doorway).

Our contributions:

- **Reformulation** of NLOS imaging to enable 3D reconstructions for hidden occluders.
- **Physics-based** approach based on alternating minimization.
- **Physics-inspired** *Soft Shadow Diffusion* model which facilitates high-resolution 3D reconstructions from a soft shadow photograph.

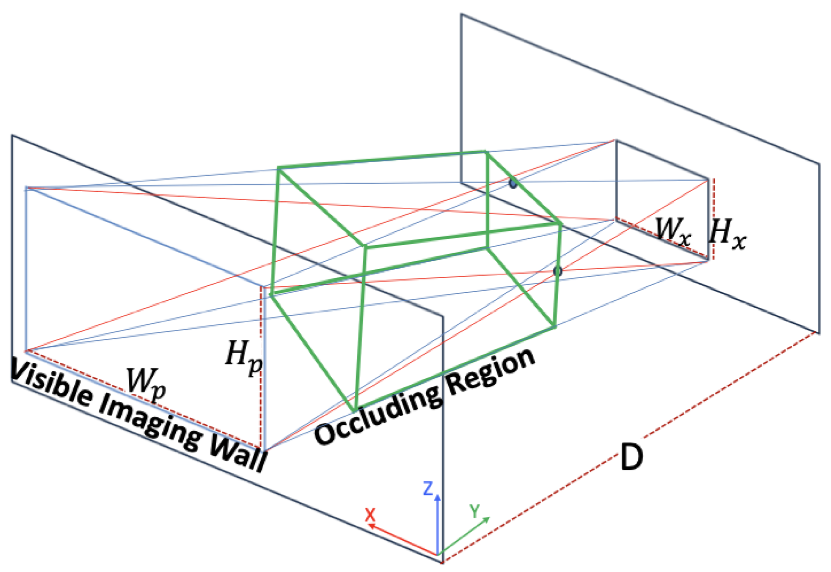
Separable Non-Linear Least Square Formulation

Generally, the brightness $i(\boldsymbol{p})$ of a patch $\boldsymbol{p}$ on the visible wall surface is:

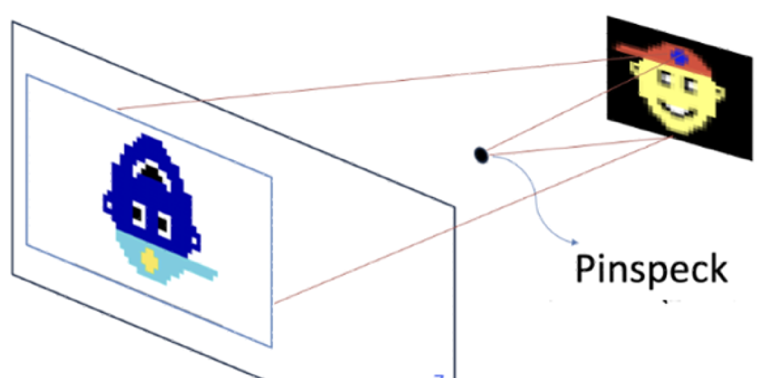
$$i(\boldsymbol{p}) = \int_{\boldsymbol{x}\in\mathcal{X}} \frac{g(\boldsymbol{p},\boldsymbol{x})}{\|\boldsymbol{x}-\boldsymbol{p}\|_2^2} v(\boldsymbol{x},\boldsymbol{p};\boldsymbol{\theta}_{\text{o}}) f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} + b(\boldsymbol{p}),$$

where $f(\boldsymbol{x})$ is the brightness of some *non-occluding* hidden scene patch $\boldsymbol{x}$, $v(\cdot)$ models the visibility between $\boldsymbol{x}$ and $\boldsymbol{p}$ due to an unknown hidden 3D occluder represented by $\boldsymbol{\theta}_{\text{o}}$, and $b(\cdot)$ ambient/visible side illumination.

Computational field of view: Hidden region that can be reliably reconstructed, for a measurement FOV



Field of View: projecting the **edges/vertices of the non-occluding plane** toward the edges of the FOV on the visible wall.



Pinspeck Occluder: model the properties of the occluding region as a pinspeck Occluder

Hidden Region = Non-occluding sub-region (f) & Occluding sub-region ($\boldsymbol{\theta}_{\text{o}}$)

Discretizing the integral yields a **separable non-linear least square** problem:

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

where $\mathbf{y}$ is a soft shadow photograph, $\mathbf{f}$ is a (discrete) representation of the *non-occluding sub-region's* brightness, $\boldsymbol{\theta}_{\text{o}}$ is representation of hidden 3D occluder, and $\mathbf{b}$ is background.

Pure Physics with Voxelized 3D Occupancy Grid

By representing the occluding sub-region as being composed of discrete voxels (pinpecks/anti-pinholes), and the non-occluding sub-region as discrete pixel patches, we can formulate a regularized LS problem for reconstructing $(\mathbf{f}, \boldsymbol{\theta}_{\text{o}})$:

$$\underset{(\mathbf{f},\boldsymbol{\theta}_{\text{o}})}{\arg\min} \|\mathbf{A}(\boldsymbol{\theta}_{\text{o}})\mathbf{f} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{f}\|_2^2.$$

An **alternating minimization** solution proceeds as:

$$\widehat{\mathbf{f}}^k = \underset{\mathbf{f}}{\arg\min} \left\|\mathbf{A}(\widehat{\boldsymbol{\theta}}_{\text{o}}^{k-1})\mathbf{f} - \mathbf{y}\right\|_2^2 + \lambda\|\mathbf{f}\|_2^2 = \left(\mathbf{A}^\top(\widehat{\boldsymbol{\theta}}_{\text{o}}^{k-1})\mathbf{A}(\widehat{\boldsymbol{\theta}}_{\text{o}}^{k-1}) + \lambda\mathbf{I}\right)^{-1}\mathbf{A}^\top(\widehat{\boldsymbol{\theta}}_{\text{o}}^{k-1})\mathbf{y}$$

$$\widehat{\boldsymbol{\theta}}_{\text{o}}^k = \underset{\boldsymbol{\theta}_{\text{o}}}{\arg\min} \left\|\mathbf{A}(\boldsymbol{\theta}_{\text{o}})\widehat{\mathbf{f}}^k - \mathbf{y}\right\|_2^2.$$

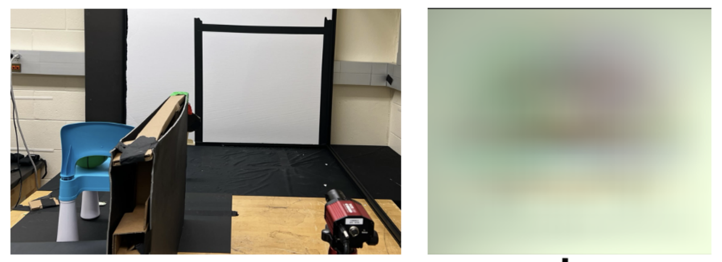
Drawback: 3D voxel-based representation is limited by voxel grid density (highly redundant representation), yet a dense grid increases computational complexity.

Soft Shadow Diffusion (SSD) Model

Promotes computational tractability as a learned lower dimensional representation for $\boldsymbol{\theta}_{\text{o}}$ and acts as an implicit prior for 3D structures. By conditioning on a soft shadow photograph, we train our novel soft shadow diffusion model to reconstruct a 3D point cloud of an occluding hidden scene object.

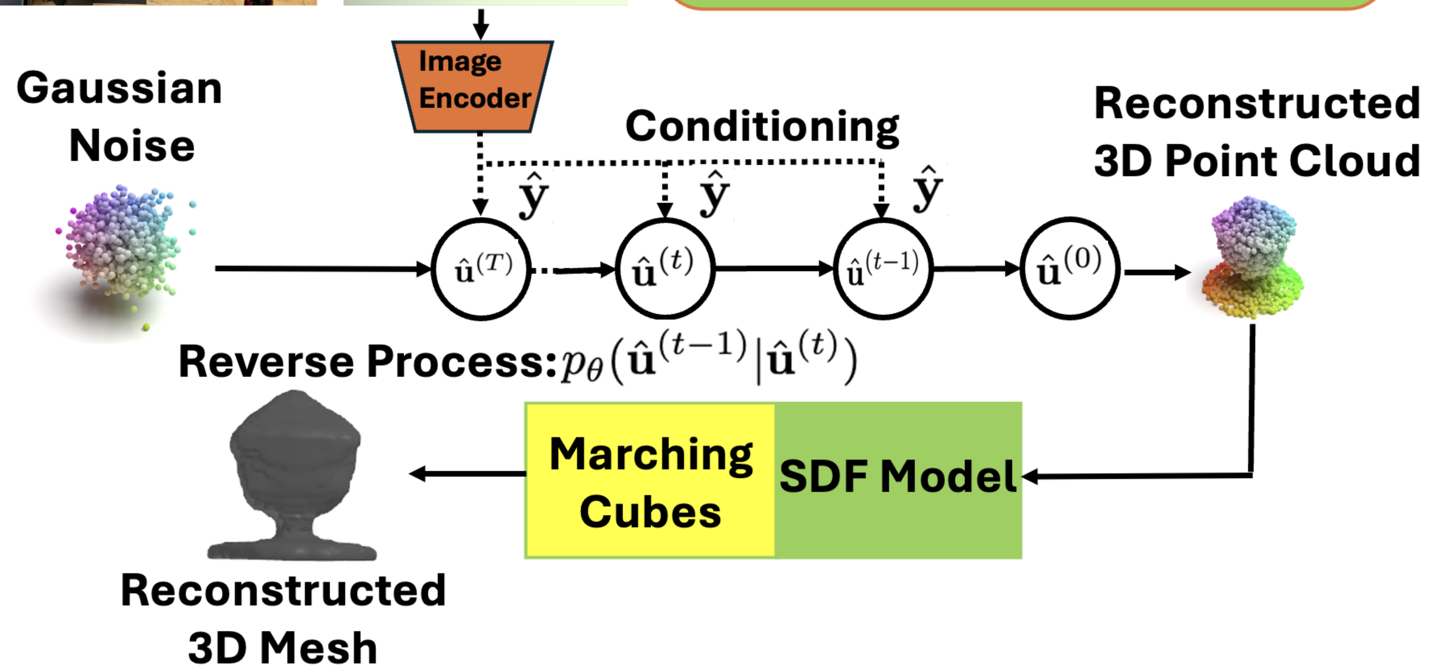
Intuition: From the LS-regularized problem above, we can reconstruct $\boldsymbol{\theta}_{\text{o}}$ without knowing $\mathbf{f}$, as $\widehat{\boldsymbol{\theta}}_{\text{o}} = \arg\min_{\boldsymbol{\theta}_{\text{o}}} \left\|\left[\mathbf{A}^\top(\boldsymbol{\theta}_{\text{o}})\left(\mathbf{A}^\top(\boldsymbol{\theta}_{\text{o}})\mathbf{A}(\boldsymbol{\theta}_{\text{o}})+\lambda\mathbf{I}\right)^{-1}\mathbf{A}^\top(\boldsymbol{\theta}_{\text{o}})-\mathbf{I}\right]\mathbf{y}\right\|_2^2.$

1. *Reconstruct shape* of $\boldsymbol{\theta}$ from $\mathbf{y}$ via SSD:

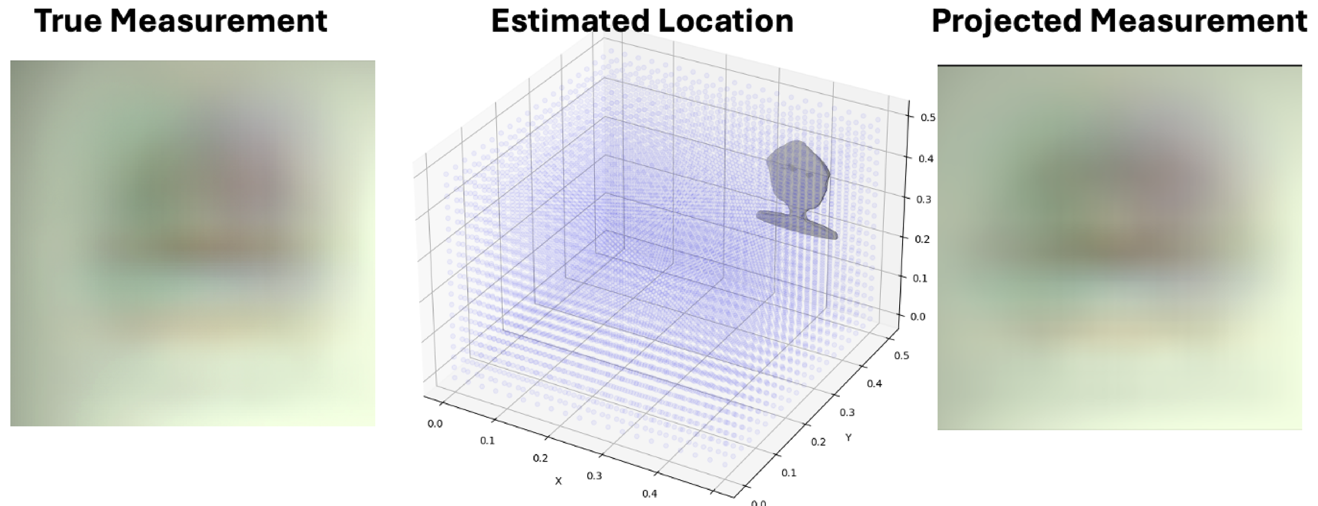


Training Details

- 250K Simulated Soft Shadow + Occluder Point cloud
- CrossFormer Image Encoder
- Multi-Layers of 1D Convolutions + Self Attention
- 95 Million Parameters
- 600k Training Iterations
- 256 Diffusion Timestep with DDPM sampling



2. *Estimate location* of the reconstructed occluder shape by grid search (over hidden region):

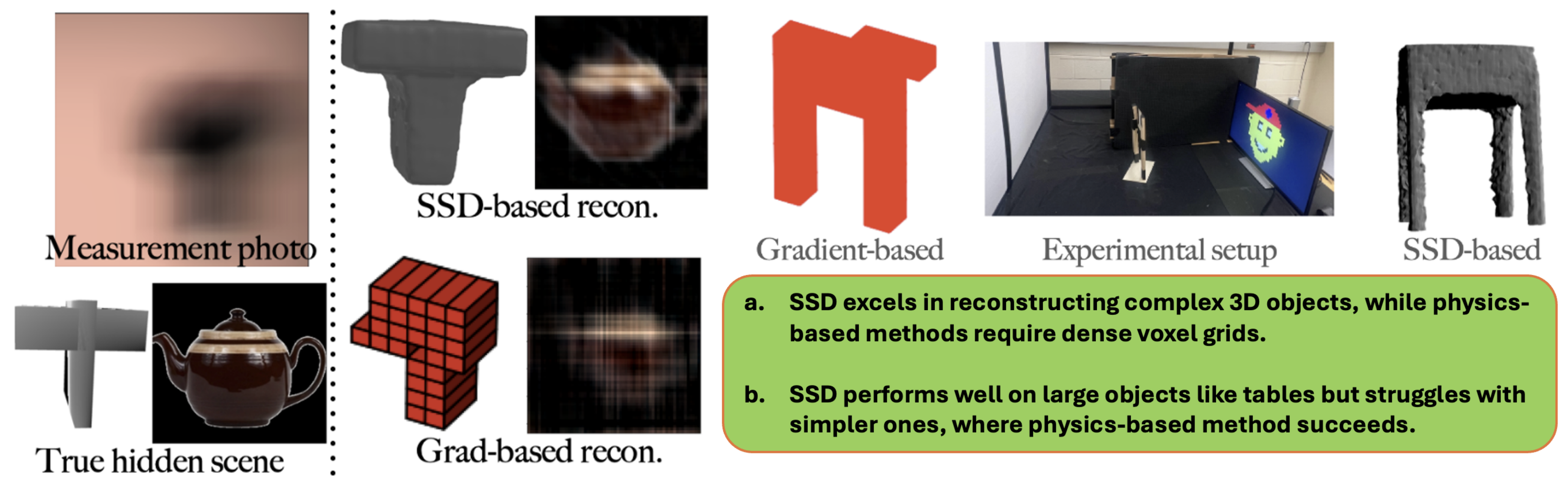


3. *Reconstruct non-occluding hidden scene* in 2D f by solving a TV-regularized inverse problem:

$$\widehat{\mathbf{f}} = \underset{\mathbf{f}}{\arg\min} \|\mathbf{y} - \mathbf{A}(\widehat{\boldsymbol{\theta}}_{\text{o}})\mathbf{f}\|_2^2 + \lambda\|\mathbf{f}\|_{\text{TV}}^2.$$

Results

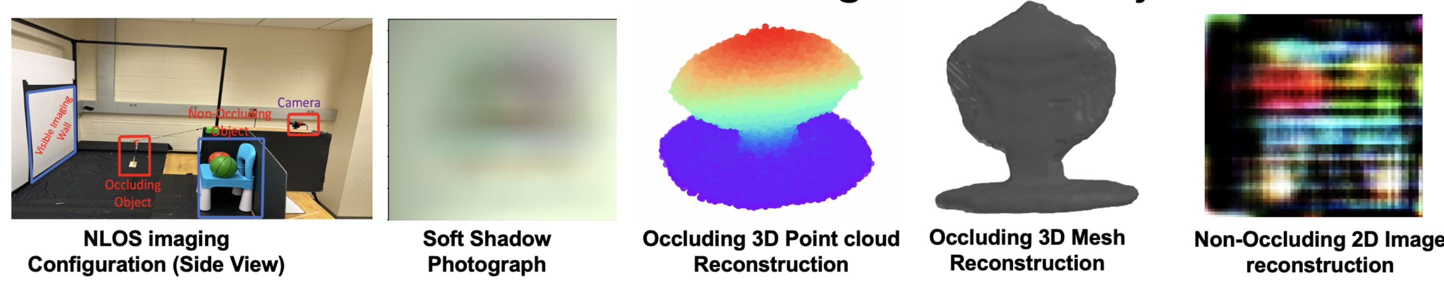
Soft Shadow Diffusion Vs Physics-Based Reconstruction



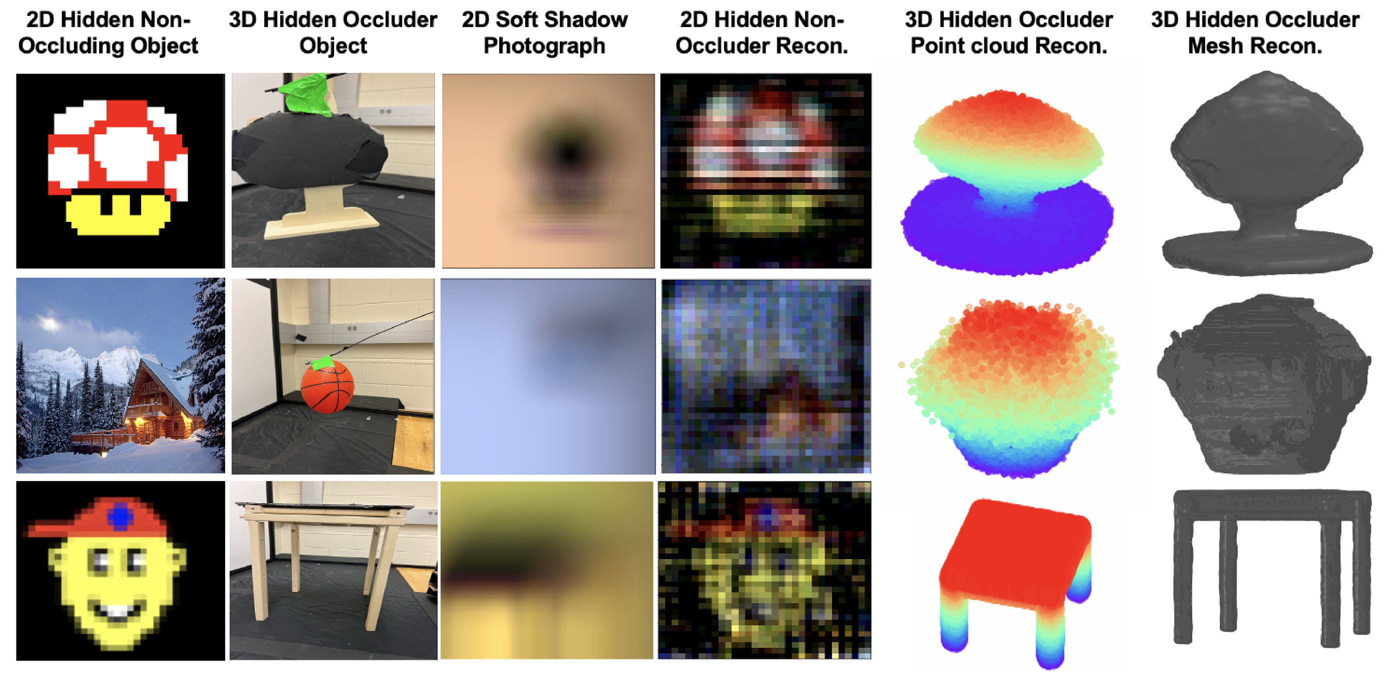
a. SSD excels in reconstructing complex 3D objects, while physics-based methods require dense voxel grids.

b. SSD performs well on large objects like tables but struggles with simpler ones, where physics-based method succeeds.

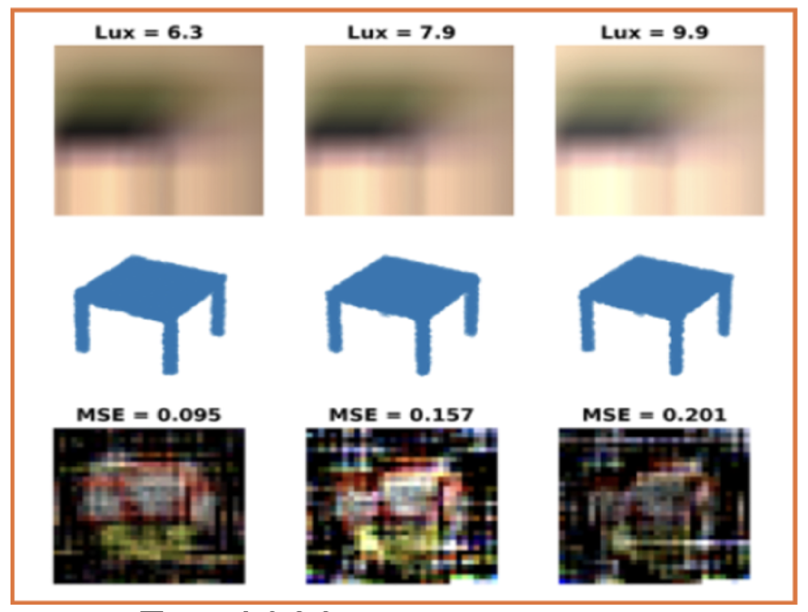
Soft Shadow Diffusion on real 3D light reflective object



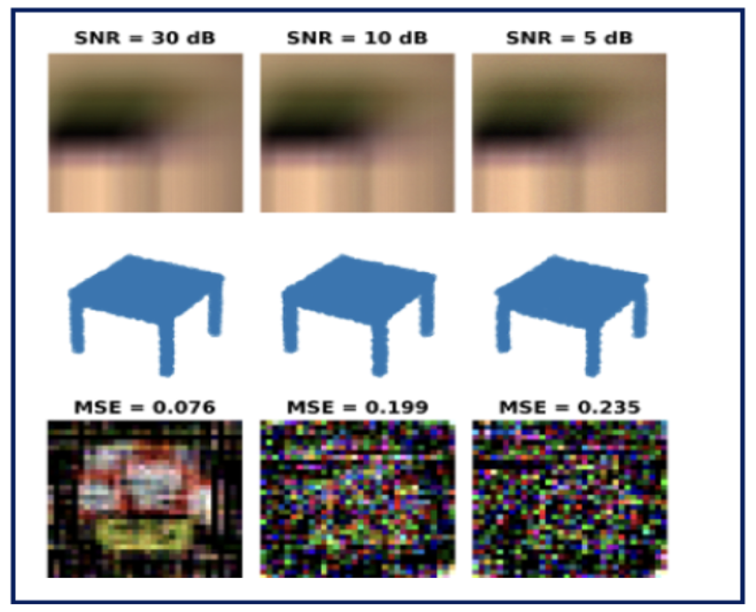
SSD-based reconstruction



Reconstructions for increasing ambient light levels



Reconstructions for increasing noise levels



True hidden scene setup

a. SSD is robust to background light, allowing reconstruction of non-occluded objects at different light levels.

b. SSD shows robustness to noise, as it degrades gracefully with increasing noise levels