CSC2457 3D & Geometric Deep Learning

Neural Reflectance Fields for Appearance Acquisition

SAI BI*, UC San Diego
ZEXIANG XU*, UC San Diego, Adobe Research
PRATUL SRINIVASAN and BEN MILDENHALL, UC Berkeley
KALYAN SUNKAVALLI, MILOŠ HAŠAN, and YANNICK HOLD-GEOFFROY, Adobe Research
DAVID KRIEGMAN and RAVI RAMAMOORTHI, UC San Diego

Feb 23, 2021

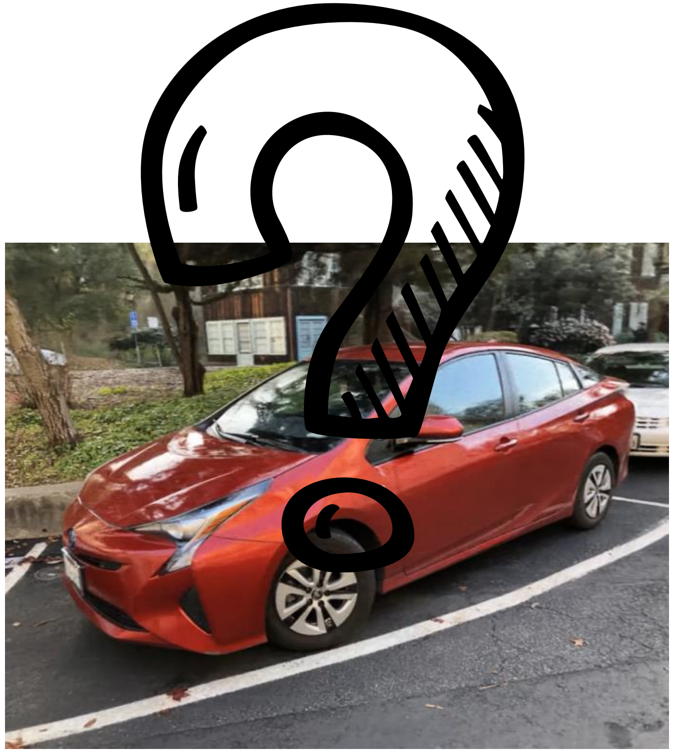
Presenter: Zian Wang

Instructor: Animesh Garg

UNIVERSITY OF TORONTO

Motivation

- NeRF is impressive on capturing appearance.
- But is that all we need?



Motivation

- NeRF is impressive on capturing appearance.
- But is that all we need?

- NeRF captures radiance but not material!

View synthesis ✓   Relighting ✗

Motivation

- Games / VR / AR – We need materials!
  - Game Contents
  - VR/AR applications such as object insertion

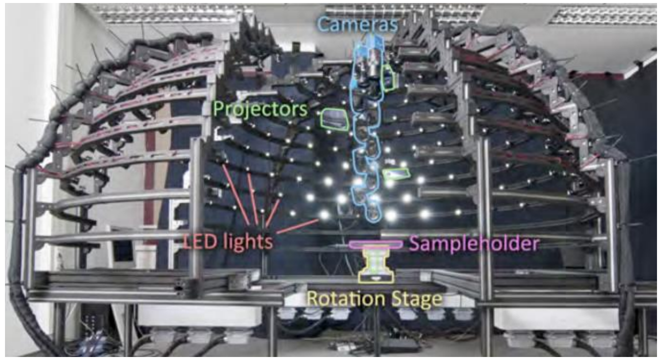
[Source: https://www.businessinsider.com/uncharted-4-photos-2016-5]

Motivation

- Why is it challenging?
  - Ill-posed inverse problem.
  - Appearance is correlated to both material, lighting and geometry.
    - Same appearance leads to multiple solutions.

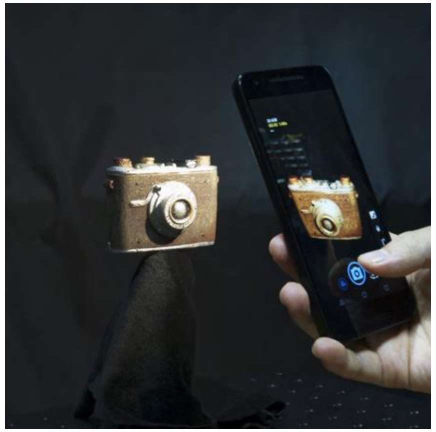
Prior Works

- Material Capturing
  - Light stage settings: accurate but bulky.
  - Portable capturing with cellphone camera and flash.
    - Mesh and Voxel representation cannot handle fine details.

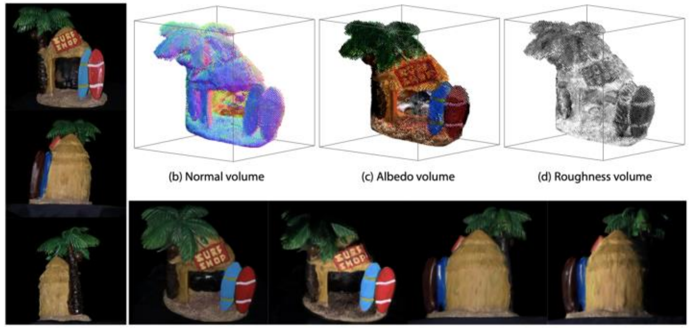


[Schwartz et al., 2013]

Portable



[Nam et al., 2018]



[Bi et al., 2020]

Prior Works

- Material Capturing
  - Light stage settings: accurate but bulky.
  - Portable capturing with cellphone camera and flash.
    - Mesh and Voxel representation cannot handle fine details.
- NeRF
  - Impressive results and can capture fine details.
  - Do not handle material properties.

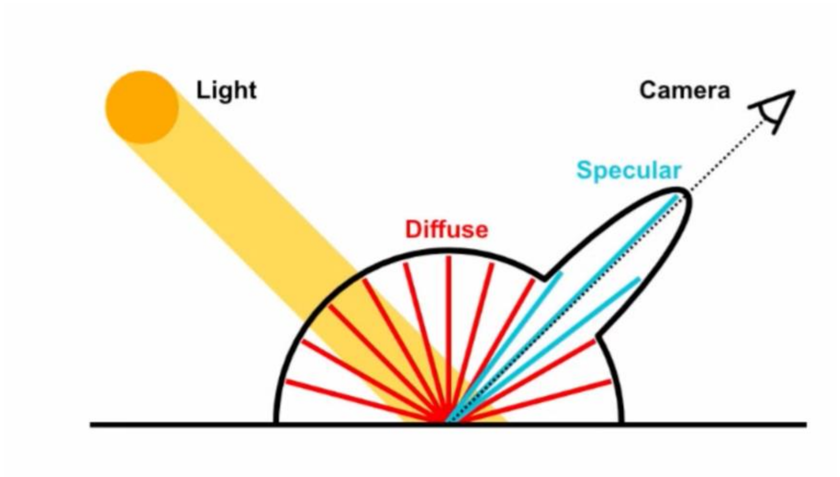
Extend NeRF to capture materials!

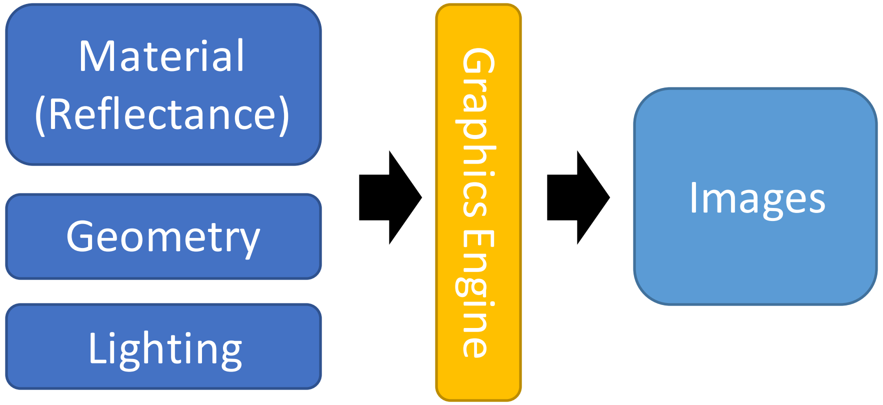
Contributions

- Task
  - The ill-posed problem of jointly capturing material and geometry from multi-view images.
  - Portable content capturing is important for Game / VR / AR.
  - Prior works either do not handle material (NeRF) or cannot capture details (Mesh, Voxel).
- Key Insight
  - Following prior works, using controlled lighting condition to constrain ambiguity.
  - Extend NeRF to predict material properties and optimize with photometric loss.
  - Adapt NeRF's ray marching to render radiance with geometry, lighting and material.
- Result
  - Given cellphone captured videos (under controlled lighting condition),
  - We get relightable high-quality (fine details) implicit function representation of objects.

Background – Reflectance

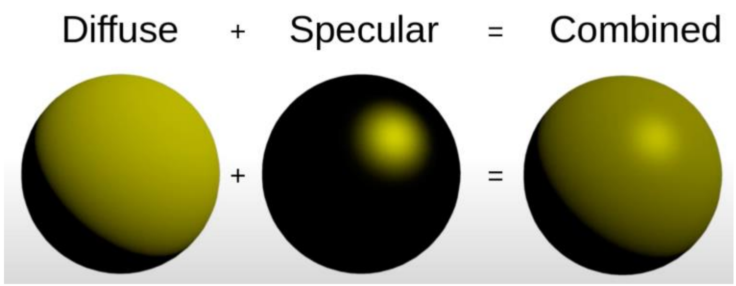
- Reflectance
  - We see appearance because surfaces reflect light.

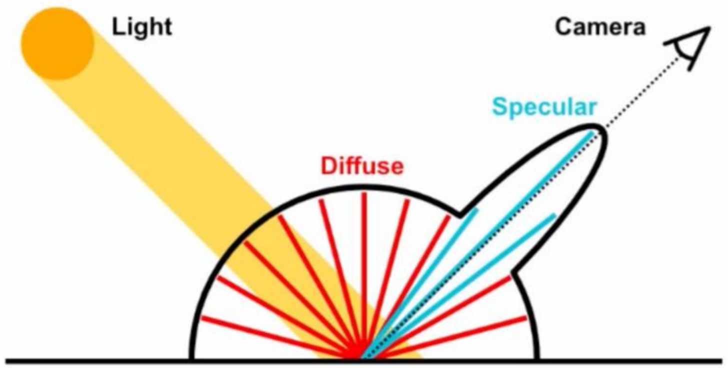




Background – Reflectance

- Diffuse (Lambertian)
  - Reflects light uniformly in all directions.
  - E.g. the wall.
- Specular
  - Reflected light depends on viewing direction.
  - E.g. the mirror.

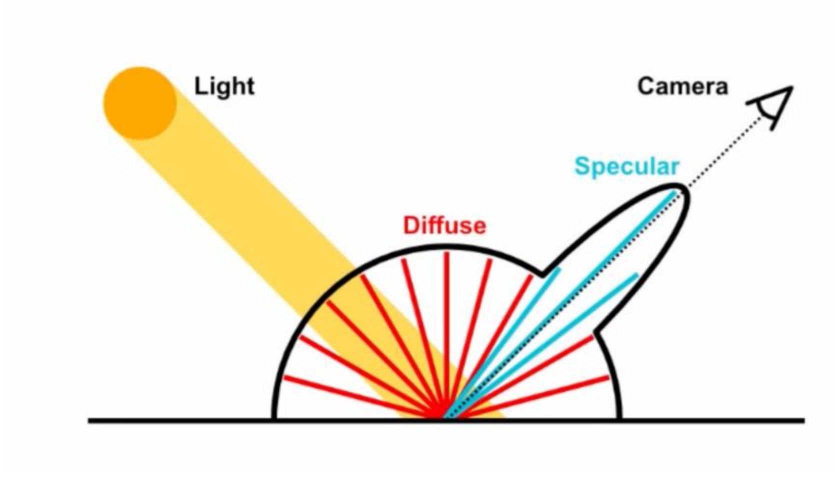




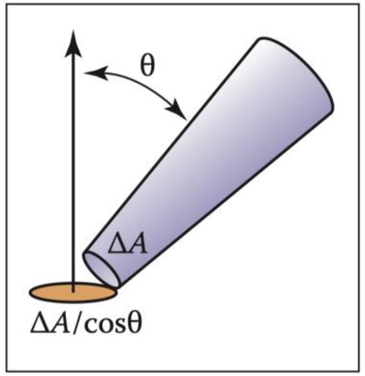
Background – Reflectance

- Bidirectional reflectance distribution function (BRDF)
  - BRDF is a surface material property describing how light reflects

$$f_{\mathrm{r}}(\mathbf{x}, \omega_{\mathrm{i}}, \omega_o)$$



Background – Rendering Equation



Rendering equation defines how light scatters in a scene

$$L_o(\mathbf{x}, \omega_o, \lambda, t) = L_e(\mathbf{x}, \omega_o, \lambda, t) + \underbrace{\int_\Omega f_r(\mathbf{x}, \omega_i, \omega_o, \lambda, t) L_i(\mathbf{x}, \omega_i, \lambda, t)(\omega_i \cdot \mathbf{n})\, \mathrm{d}\,\omega_i}_{\text{Reflected Light}}$$

Outgoing light

Incoming light

Emitted light

Hemisphere on point

BRDF

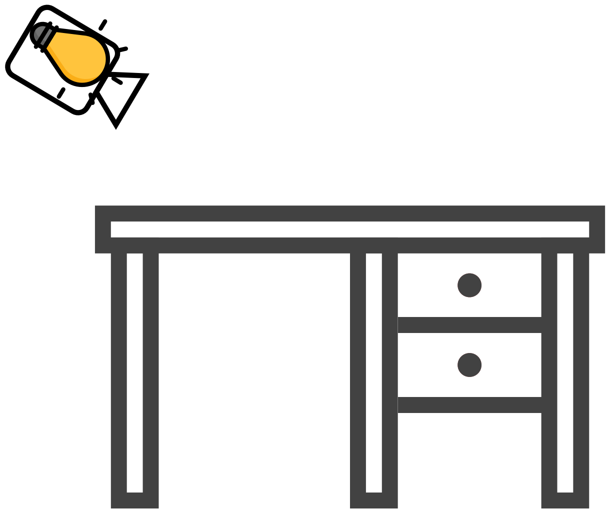
Irradiance factor

Task Definition

- One step further than NeRF - Capturing material
  - Input is multi-view images with collocated camera-light setup.
  - Output is a Neural Reflectance Field.

Task Definition

- One step further than NeRF - Capturing material
  - Input is multi-view images with collocated camera-light setup.
  - Output is a Neural Reflectance Field.



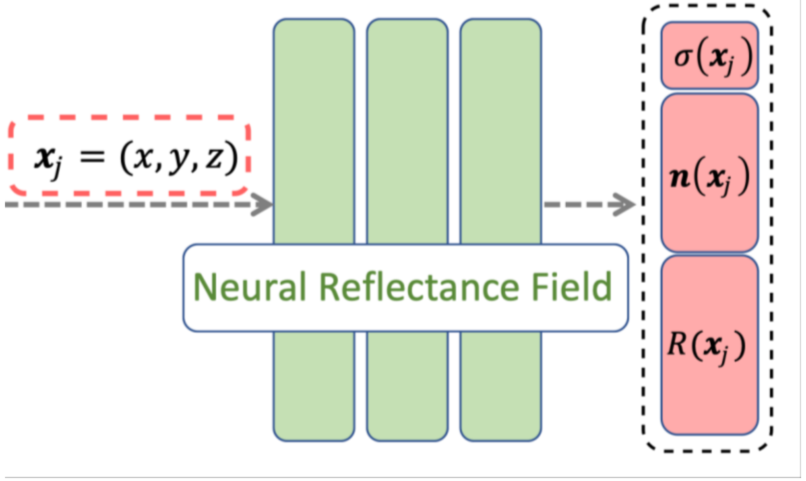
Task Definition

- Input image samples
  - Robot arm
  - Galaxy Note 8

Task Definition

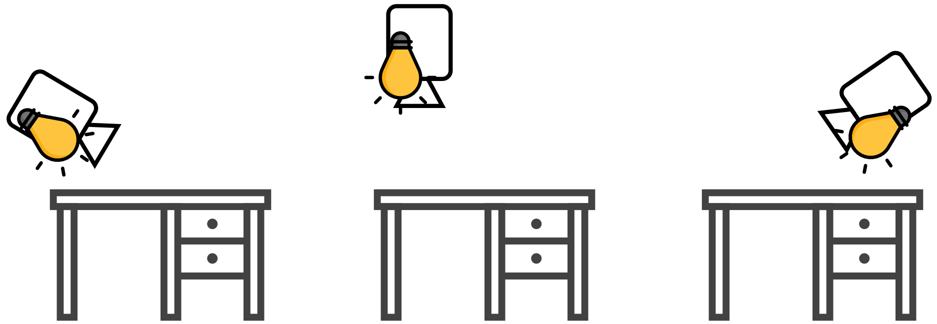
- One step further than NeRF - Capturing material
  - Input is multi-view images with collocated camera-light setup.
  - Output is a Neural Reflectance Field.

An MLP to predict density, normal, reflectance at every 3D location.

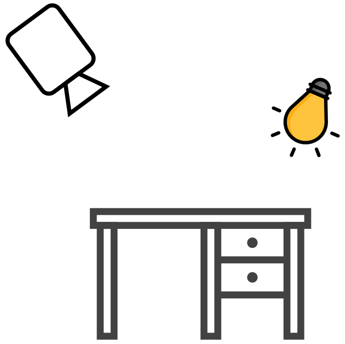


Task Definition

- Goal of Neural Reflectance Field
  - Render with novel view and light



Training



Testing

Task Definition

- Why collocated camera-light setup?

Task Definition

- Ill-posedness
  - Same appearance leads to multiple solutions.

$$L_o(\mathbf{x}, \omega_o, \lambda, t) = \cancel{L_e(\mathbf{x}, \omega_o, \lambda, t)} + \int_\Omega f_r(\mathbf{x}, \omega_i, \omega_o, \lambda, t) L_i(\mathbf{x}, \omega_i, \lambda, t)(\omega_i \cdot \mathbf{n})\, \mathrm{d}\,\omega_i$$

Outgoing light Hemisphere BRDF Incoming light Irradiance factor

Task Definition

- Why collocated camera-light setup?
  - Known single light source – Removes the integral
  - Use point light to approximate the cellphone flash

$$L_o(\mathbf{x},\omega_o,\lambda,t) = \cancel{L_e(\mathbf{x},\omega_o,\lambda,t)} + \cancel{\int_\Omega} f_r(\mathbf{x},\omega_i,\omega_o,\lambda,t)\,L_i(\mathbf{x},\omega_i,\lambda,t)\,(\omega_i\cdot\mathbf{n})\,\cancel{d\omega_i}$$

Outgoing light

BRDF

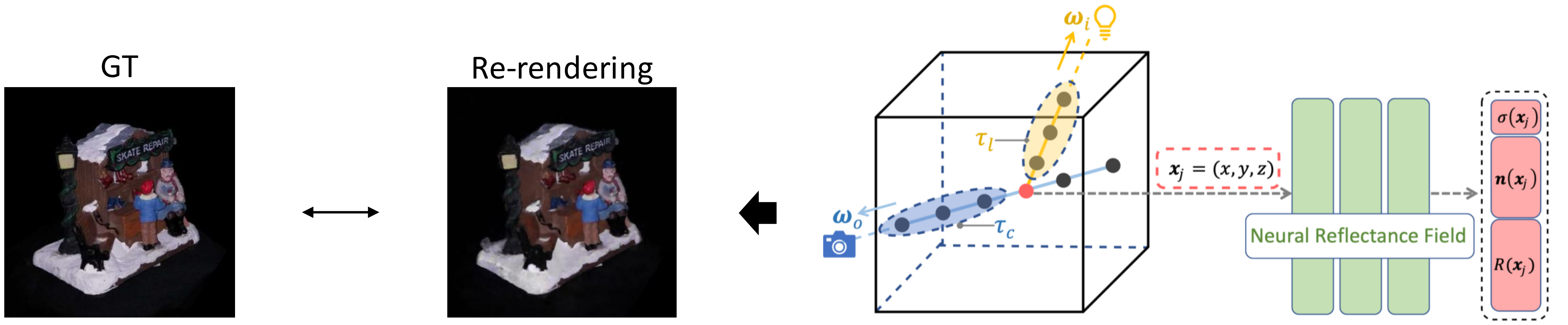
Incoming light

Irradiance factor

$$L_{\mathrm{o}}(\mathbf{x},\omega_{\mathrm{o}}) = f_{\mathrm{r}}(\mathbf{x},\omega_{\mathrm{i}},\omega_{\mathrm{o}})\,L_{\mathrm{i}}(\mathbf{x},\omega_{\mathrm{i}})$$

Method

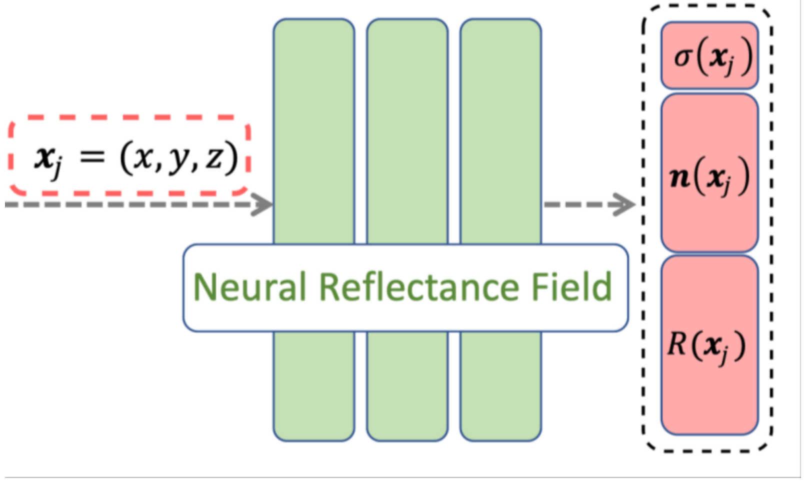
- General Idea
  - Jointly optimize material and geometry with re-render loss.



Supervision **Differentiable Rendering** **3D Representation**

Method

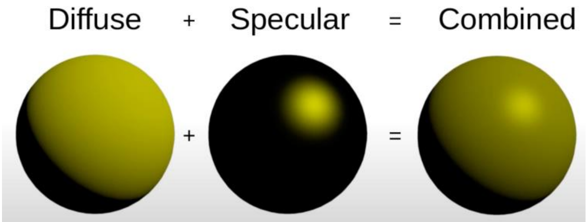
- Neural Reflectance Fields
  - At every 3D location, this MLP predicts
    - Volume density (1-channel)
    - Surface normal (3-channel)
    - Reflectance (4-channel)



Method

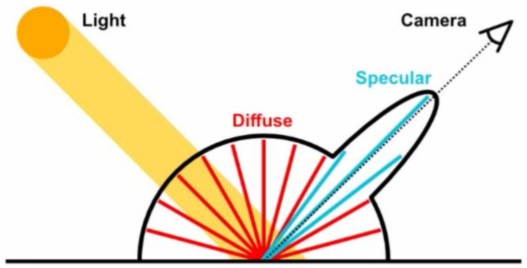
- Material / BRDF Parameters
  - The BRDF used in this paper is a simplified *microfacet model*.
  - Use diffuse albedo (3-channel) and roughness (1-channel) to describe reflectance

$$f_\mathrm{r}(\omega_\mathrm{i}, \omega_\mathrm{o}; A, R, N) = f_\mathrm{d}(\omega_\mathrm{i}, \omega_\mathrm{o}; A, N) + f_\mathrm{s}(\omega_\mathrm{i}, \omega_\mathrm{o}; R, N)$$



Diffuse: $$f_\mathrm{d}(\omega_\mathrm{i}, \omega_\mathrm{o}; A, N) = \frac{A}{\pi}$$

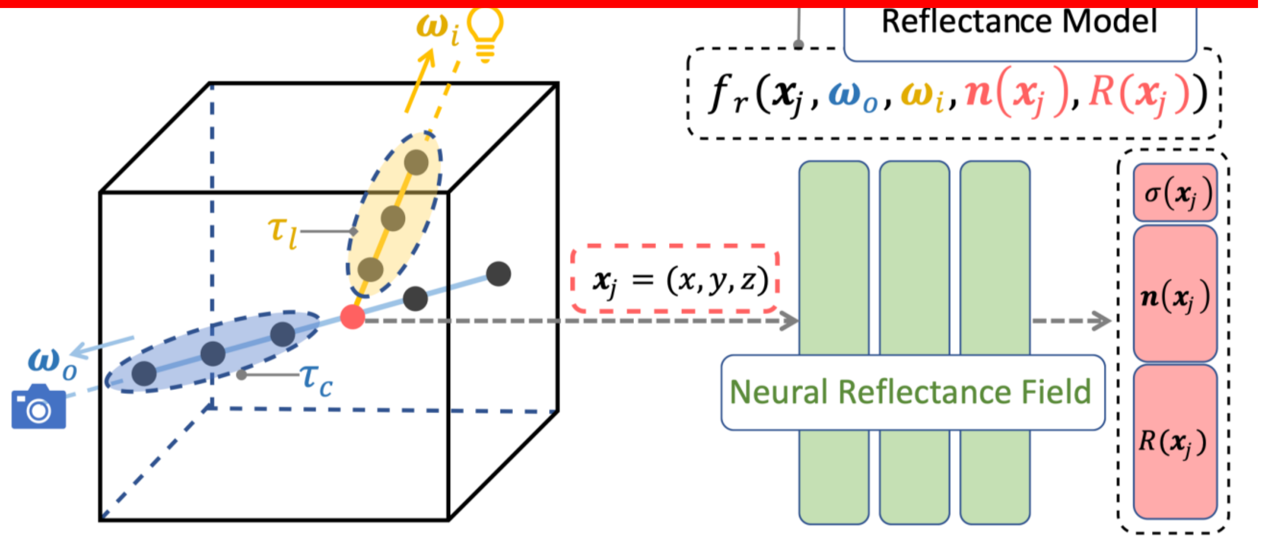
Specular: $$f_\mathrm{s}(\omega_\mathrm{i}, \omega_\mathrm{o}; R, N) = \frac{\mathcal{D}(R, h)\mathcal{F}(\omega_\mathrm{o}, h)\mathcal{G}(R, N, \omega_\mathrm{i}, \omega_\mathrm{o})}{4(N \cdot \omega_\mathrm{i})(N \cdot \omega_\mathrm{o})}$$



Method

- Reflectance-Aware Ray Marching
  - At any sampled point on the ray, use material and lighting to render the current location.
  - Use “alpha compositing” for sampled points along a ray.

$$\sum_j \tau_c(\boldsymbol{x}_j)\,\tau_l(\boldsymbol{x}_j)\,\left(1-\exp(-\sigma(\boldsymbol{x}_j)\,\Delta t_j)\right)\boxed{f_r(\boldsymbol{x}_j)}\,L_l(\boldsymbol{x}_j)$$



Method

- Reflectance-Aware Ray Marching
  - At any sampled point on the ray, use material and lighting to render the current location.
  - Use “alpha compositing” for sampled points along a ray.

(1) Consider rendering one sampled location:

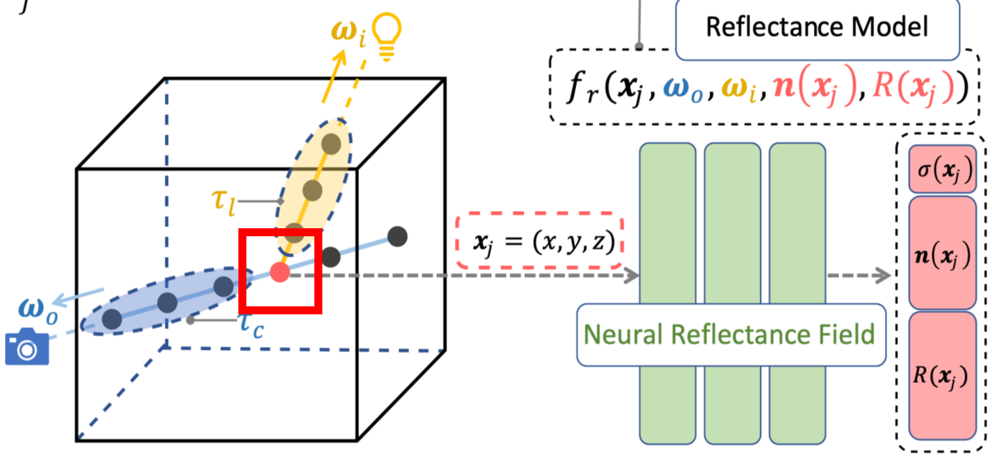
$$f_r(\boldsymbol{x}_j, \boldsymbol{\omega}_o, \boldsymbol{\omega}_i, \boldsymbol{n}(\boldsymbol{x}_j), R(\boldsymbol{x}_j)) \qquad \tau_l(\boldsymbol{x}_j) \qquad L_l(\boldsymbol{x}_j)$$

BRDF

Transmittance (Visibility)

Point Light Intensity

$$\sum_j \tau_c(\boldsymbol{x}_j)\, \tau_l(\boldsymbol{x}_j)\, (1 - \exp(-\sigma(\boldsymbol{x}_j)\, \Delta t_j))\, f_r(\boldsymbol{x}_j)\, L_l(\boldsymbol{x}_j)$$



Method

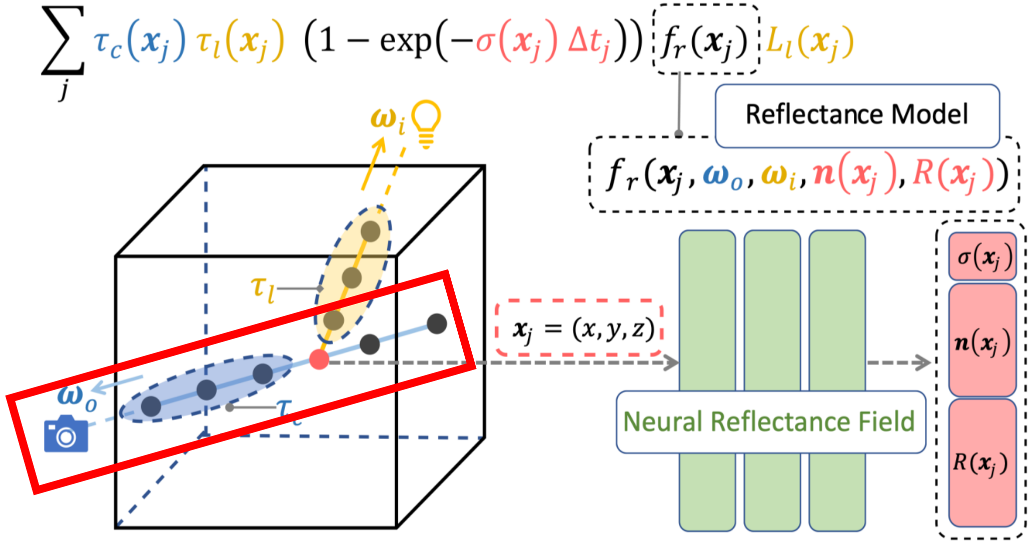
- Reflectance-Aware Ray Marching
  - At any sampled point on the ray, use material and lighting to render the current location.
  - Use “alpha compositing” for sampled points along a ray.

(2) Composite along a ray:

$$\sum_j \tau_c(\boldsymbol{x}_j)\,\underbrace{(1-\exp(-\sigma(\boldsymbol{x}_j)\,\Delta t_j))}_{\text{Density of the sampled point}}\,\underbrace{f_r(\boldsymbol{x}_j)\,\tau_l(\boldsymbol{x}_j)\,L_l(\boldsymbol{x}_j)}_{\text{Radiance of one sampled point}}$$

Transmittance (Visibility) of the point

$$\tau_c(\boldsymbol{x}_j) = \exp\left(-\sum_{k=0}^{j}\sigma(\boldsymbol{x}_k)\Delta t_k\right)$$



Method

GT

Re-rendering

- Supervision
  - Re-render L2 Loss
  - Regularization on transmittance (either 0 or 1)

$$\sum_q \|L^q_{\text{coarse}} - \tilde{L}^q\|^2 + \|L^q_{\text{fine}} - \tilde{L}^q\|^2 + \beta[\log(\tau_{\boldsymbol{c}}^q) + \log(1 - \tau_{\boldsymbol{c}}^q)]$$

Method

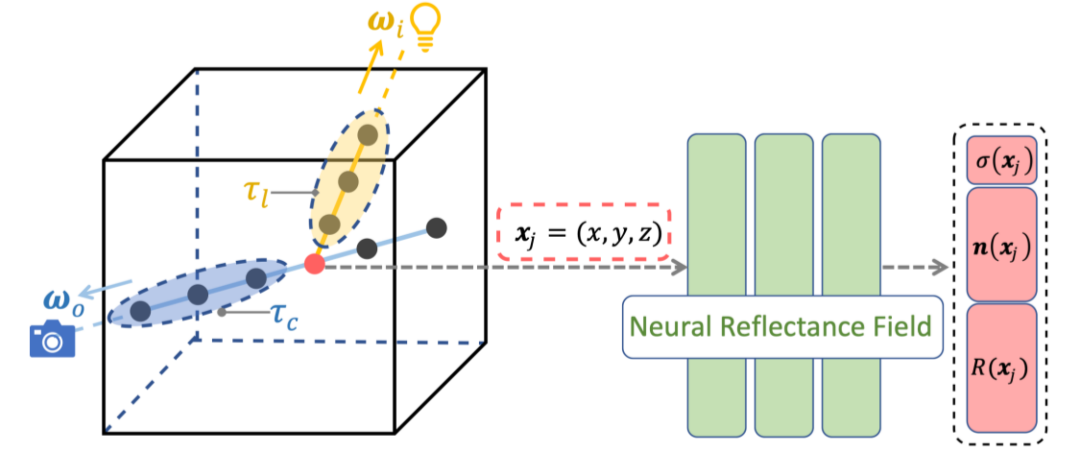
- Quick recap

$$\sum_q \|L^q_{\text{coarse}} - \tilde{L}^q\|^2 + \|L^q_{\text{fine}} - \tilde{L}^q\|^2 + \beta[\log(\tau_{\boldsymbol{c}}^q) + \log(1-\tau_{\boldsymbol{c}}^q)]$$

$$\sum_j \tau_c(\boldsymbol{x}_j)\,(1-\exp(-\sigma(\boldsymbol{x}_j)\,\Delta t_j))\,f_r(\boldsymbol{x}_j)\,\tau_l(\boldsymbol{x}_j)\,L_l(\boldsymbol{x}_j)$$

GT

Re-rendering



Supervision

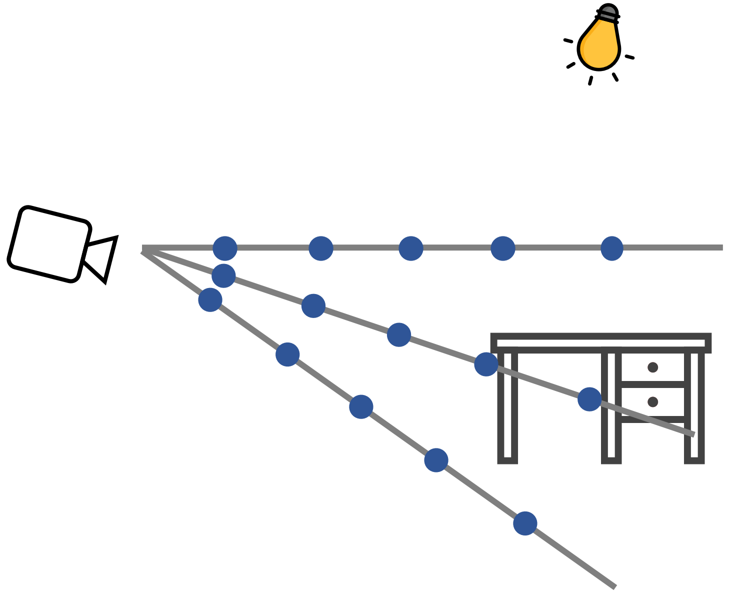
Differentiable Rendering

3D Representation

Method

- Efficiency
  - How many queries do we need to render an N_{pixel} = HxW image?

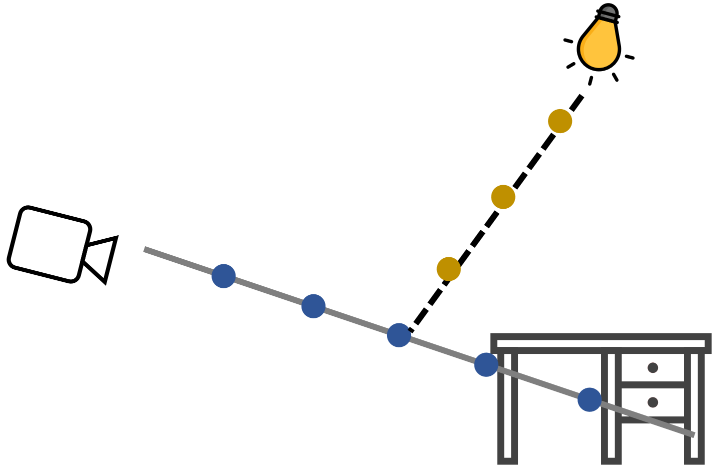
$$N_{\text{pixel}} N_{\text{sample}} \quad ?$$



Method

- Efficiency
  - How many queries do we need to render an N_{pixel} = HxW image?

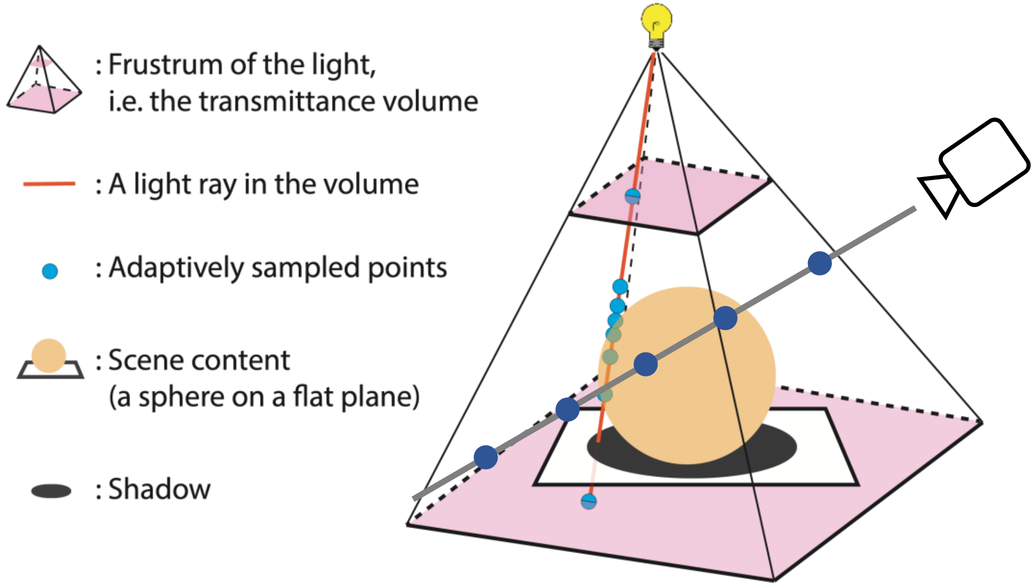
$$N_{\text{pixel}} N_{\text{sample}} N_{\text{light}} N_{\text{lsample}}$$



Method

- Speed-Up Inference
  - Precompute a light transmittance volume
  - Query for light transmittance will be interpolated from the pre-computation

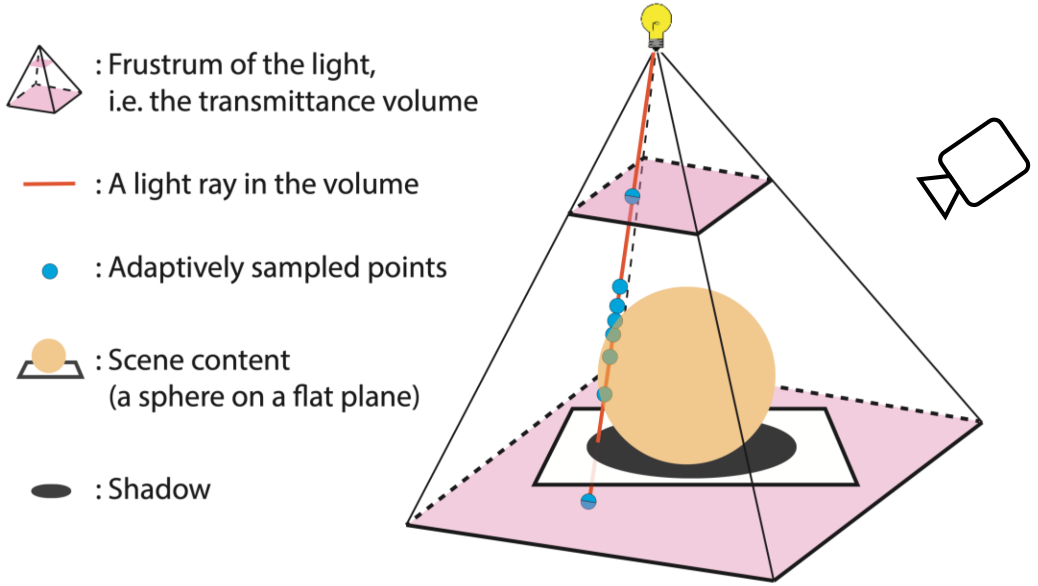
$$N_{\text{light}} N_{\text{lpixel}} N_{\text{lsample}}$$



Method

- Speed-Up Inference
  - Precompute a light transmittance volume
  - Query for light transmittance will be interpolated from the pre-computation

$$N_{\mathrm{pixel}} N_{\mathrm{sample}} + N_{\mathrm{light}} N_{\mathrm{lpixel}} N_{\mathrm{lsample}}$$



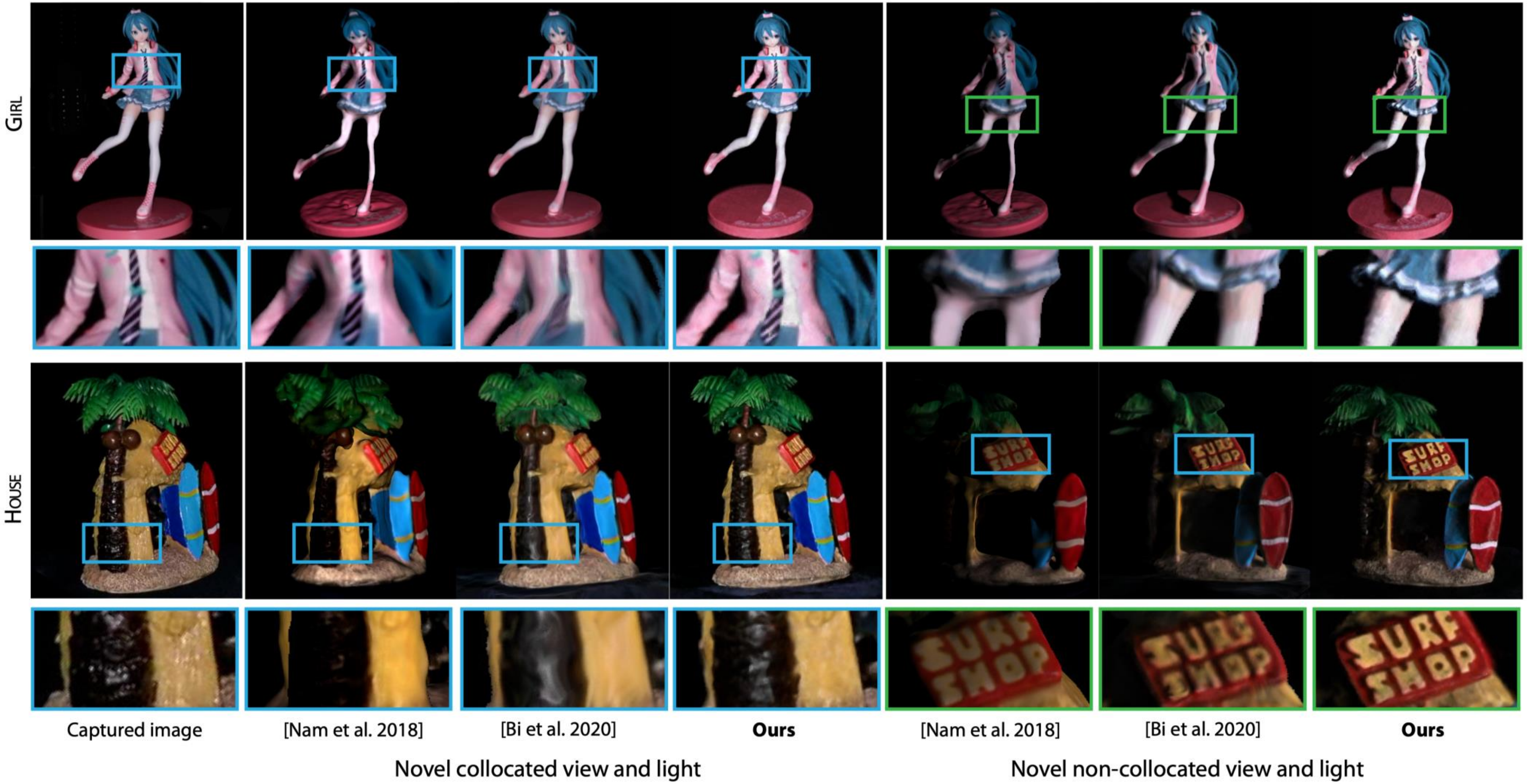
Results

- Efficiency
  - Training time: ~2 days on 4 RTX 2080Ti
  - Inference time: 30 seconds for a 512x512 image.

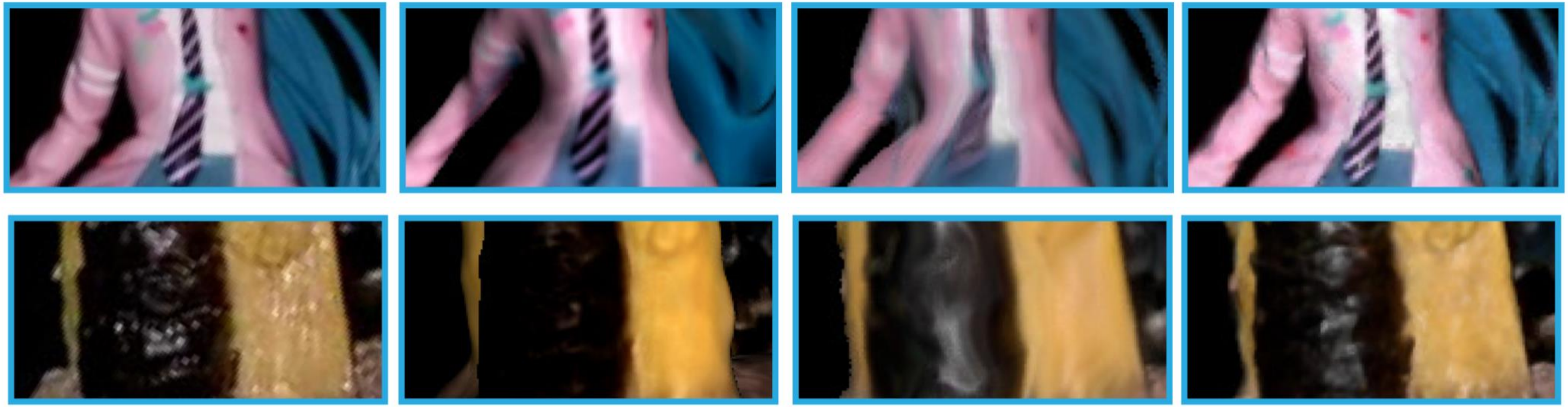
Results

- Evaluation
  - Comparison with prior works on relighting
  - Results of re-rendering and relighting
  - Generality
    - Results on a human face.
    - Results on a furry object.
  - Object Insertion Demo

Results

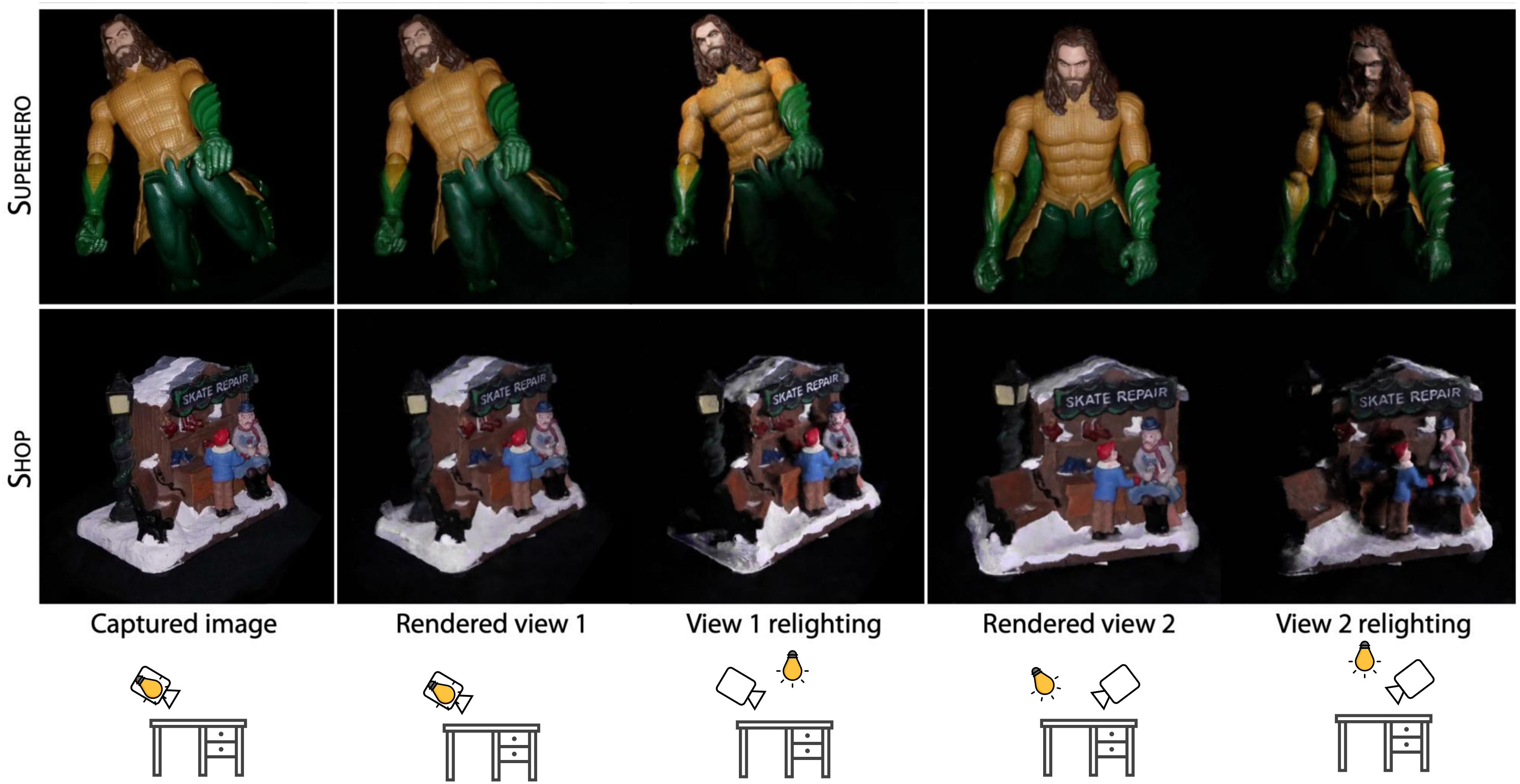

GIRL
HOUSE
SURF
SHOP
Captured image
[Nam et al. 2018]
[Bi et al. 2020]
Ours
[Nam et al. 2018]
[Bi et al. 2020]
Ours
Novel collocated view and light
Novel non-collocated view and light

Results



Captured image (GT) | [Nam et al. 2018] (Mesh) | [Bi et al. 2020] (Voxel) | **Ours** (Implicit Function)

Results



SUPERHERO
SHOP
SKATE REPAIR
Captured image
Rendered view 1
View 1 relighting
Rendered view 2
View 2 relighting

Results

- Human face

Captured images Our renderings

Results

- Furry Object
  - With a different BRDF

Captured images Our renderings

Results

- Object Insertion
  - Voxelize the implicit function (512^3) and render with Blender.

Discussion

- Conclusion from results
  - Neural Reflectance Fields enables high-quality relighting and view synthesis.
  - The method enables capturing fine details and improves material capturing.

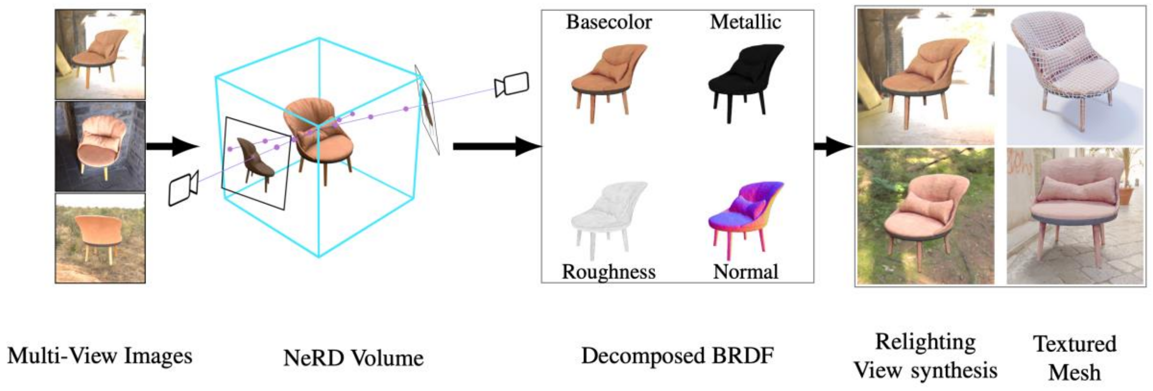
- Better-to-have results
  - Visualization of re-rendered normal and material properties.

Limitation

- Restricted lighting condition
  - It assumes that cellphone flash is the only light source. This is not convenient to satisfy in real-world.
  - Naive lighting model (point light).
- Rendering speed
  - The run-time efficiency during inference is not applicable in real-world.

Follow-up

- NeRD: Neural Reflectance Decomposition from Image Collections
  - Removes the lighting assumption during capturing
    - Input is multi-view images (Same as NeRF)
    - Output
      - a volumetric MLP encoding material and volume density per-location, and
      - per-image environment illumination.



Follow-up

- NeRD: Neural Reflectance Decomposition from Image Collections
  - Removes the lighting assumption during capturing
    - Input is multi-view images (Same as NeRF)
    - Output
      - a volumetric MLP encoding material and volume density per-location, and
      - per-image environment illumination.
  - How to constrain the additional ambiguity?
    - Introduce a bottleneck network structure for material to constrain its freedom.

Contributions (Recap)

- Task
  - The ill-posed problem of jointly capturing material and geometry from multi-view images.
  - Portable content capturing is important for Game / VR / AR.
  - Prior works either do not handle material (NeRF) or cannot capture details (Mesh, Voxel).
- Key Insight
  - Following prior works, using controlled lighting condition to constrain ambiguity.
  - Extend NeRF to predict material properties and optimize with photometric loss.
  - Adapt NeRF's ray marching to render radiance with geometry, lighting and material.
- Result
  - Given cellphone captured videos (under controlled lighting condition),
  - We get relightable high-quality (fine details) implicit function representation of objects.