VMINer: Versatile Multi-view Inverse Rendering with Near- and Far-field Light Sources

Fan Fei [1,3] Jiajun Tang [1,3] Ping Tan [2,3] Boxin Shi [1]

[1] Peking University [2] Hong Kong University of Science and Technology 3 Light Illusions

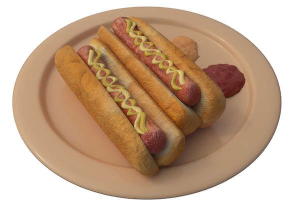
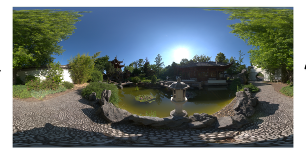
Project page: https://costrice.github.io/vminer

CVPR SEATTLE, WA JUNE 17-21, 2024

Camera Intelligence
A Computational Photography Lab @ PKU
http://camera.pku.edu.cn

Motivation

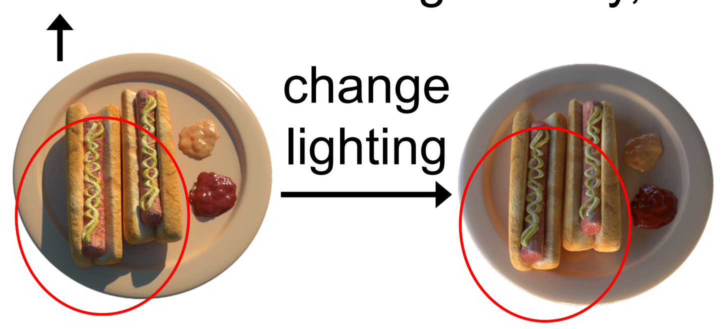
Make 3D reconstructions relightable

What will  look like under  ?

- 3D reconstruction methods usually represent only the radiance, a product of geometry, material, and lighting, making their results unrelightable.
- Inverse rendering further separates material and lighting, making the reconstruction relightable.

Leverage varied lighting to disambiguate

Severe Inherent Ambiguity: dark region in the red circle due to geometry, material, or lighting?



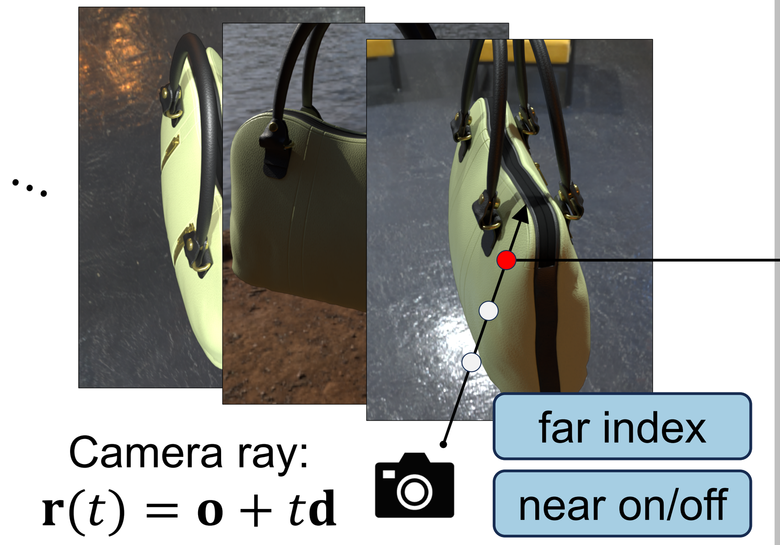
Finding: not due to material, or the dark region would sustain

Observation: it disappears under another lighting

- We propose to **make the most out of whatever lighting conditions (both far- and near-field) are at hand** to disentangle lighting from the material, resulting in a versatile framework.
- A higher degree of lighting variation gives better reconstruction results but requires a more burdensome capture process. It is up to you.
- Use flashlights as handy, effective light sources.

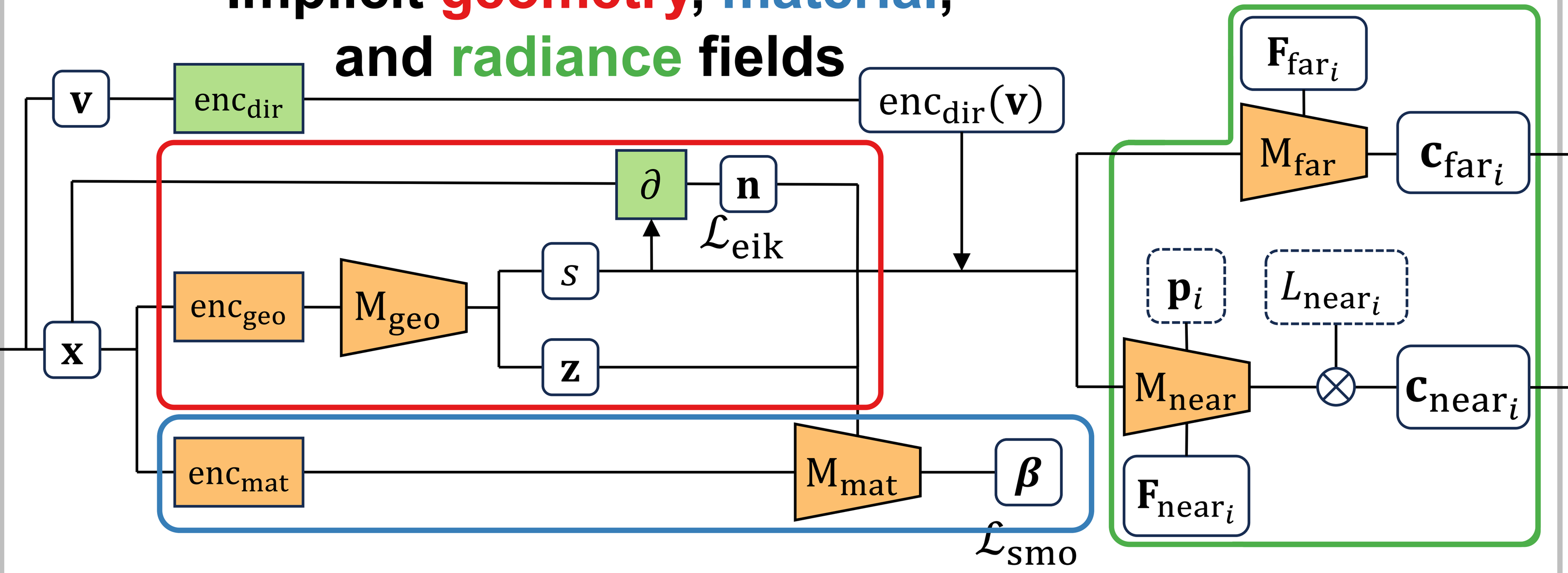
Method

Input multi-view images under varied lighting



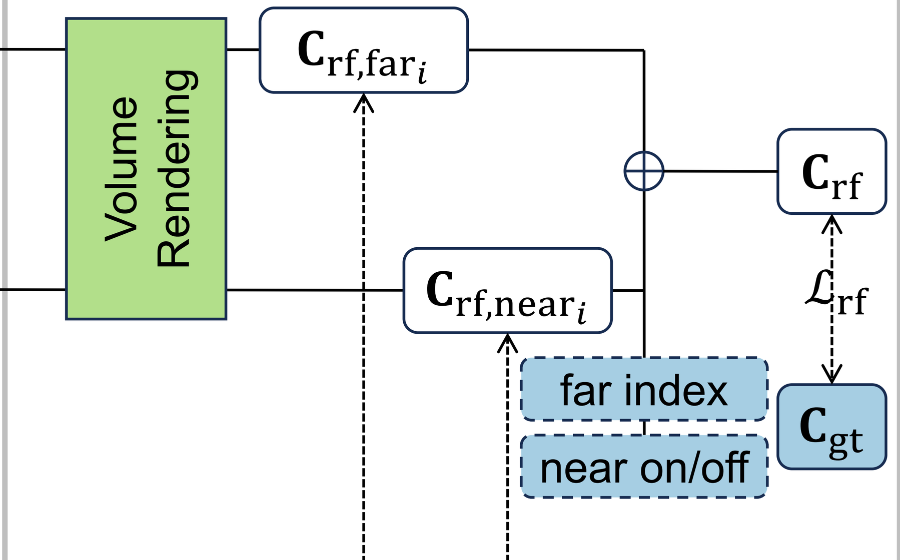
- Each of the input multi-view image is illuminated by:
  - one far-field lighting with index `far index`
  - some near-field lights with states `near on/off`

Implicit geometry, material, and radiance fields



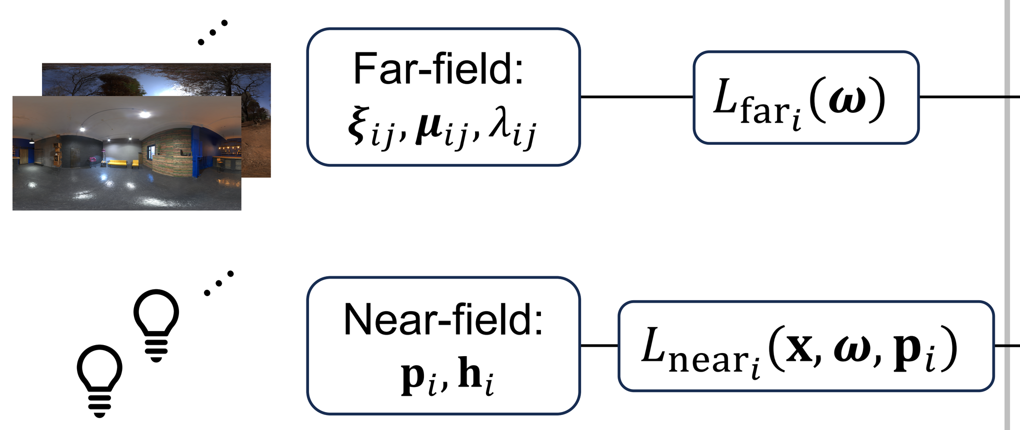
- The geometry MLP M_{geo} processes the hash grid-encoded position $\mathbf{x}$ to get the SDF s and the appearance descriptor $\mathbf{z}$.
- The material MLP M_{mat} processes the geometric information and predicts SVBRDF parameters $\boldsymbol{\beta}$.
- The radiance MLPs M_{far} M_{near} additionally take SH-encoded view direction $\mathbf{v}$ and lighting embeddings $\mathbf{F}_{far_i}$ $\mathbf{F}_{near_i}$ to get the neural radiance $\mathbf{c}_{far_i}$ $\mathbf{c}_{near_i}$ under each lighting condition.

Neural radiance field rendering



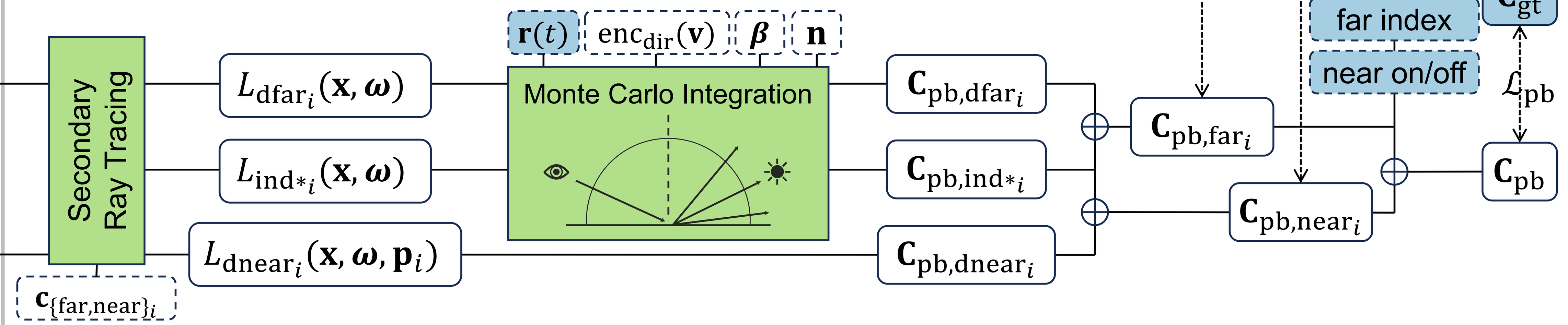
- Predict the radiance $\mathbf{C}_{rf,*}$ separately under each lighting,
- Add them up to $\mathbf{C}_{rf}$ according to the per-image lighting condition.

Versatile lighting model



- Model each far-field lighting as spherical Gaussians with per-lobe axis, amplitude, and sharpness $\xi_{ij}, \boldsymbol{\mu}_{ij}, \lambda_{ij}$
- Model near-field lighting as a point light with position and SH radiation $\mathbf{p}_i, \mathbf{h}_i$.

Physically-based surface rendering



- Compute the indirect illumination $L_{ind*_i}(\mathbf{x}, \boldsymbol{\omega})$ and the secondary visibility of the direct illumination $L_{dfar_i}(\mathbf{x}, \boldsymbol{\omega})$ $L_{dnear_i}(\mathbf{x}, \boldsymbol{\omega}, \mathbf{p}_i)$ using ray tracing and the neural radiance field $\mathbf{c}_{\{far,near\}_i}$.
- Evaluate the radiance under indirect illumination $\mathbf{C}_{pb,ind*_i}$ and far-field direct illumination $\mathbf{C}_{pb,dfar_i}$ by Monte Carlo Integration of the rendering equation.
- Compute radiance under near-field direct illumination $\mathbf{C}_{pb,dnear_i}$ efficiently *w/o* MC integration.
- Use neural radiance $\mathbf{C}_{rf,*}$ as additional signals to better separation and disambiguation.

Legend: input; variable; refer to another entry; untrainable operator; trainable operator / MLP; ⊗⊕: element-wise multiplication / add

Results

Comparison

Quantitative results

- We test with different variations of input lighting conditions, for each of which VMINer gives comparable or superior results.
- Adding lighting helps in geometry and material estimation.

Table 1. Quantitative comparison results with state-of-the-art methods averaged on 6 synthetic scenes. We show results of surface normal, diffuse albedo, view synthesis RGB, free-viewpoint (FV) relit RGB, the specular reflection part of FV relit RGB, and training time on a single RTX 3090 GPU. We mark the **best** and the second best results in each column. ↑ (↓) means bigger (smaller) is better.

Method	Input lighting conditions	Normal	Albedo		View synthesis		FV relit		FV relit (spec)		Time
		MAngE↓	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	
(1) TensoIR	Single far-field	17.66	26.48	0.921	29.48	0.912	28.18	0.901	28.30	0.861	300 mins
(2) NVDRMC	Single far-field	16.24	26.52	0.915	27.13	0.913	26.61	0.901	25.57	0.836	150 mins
(3) VMINer	Single far-field	12.39	24.50	0.882	28.20	0.934	27.46	0.921	27.56	0.871	45 mins
(4) WildLight	Single far-field + Flashlight	11.49	28.86	0.940	29.87	0.929	30.44	0.930	27.71	0.863	1440 mins
(5) Ours	Single far-field + Flashlight	10.89	31.62	0.953	32.09	**0.953**	32.00	0.953	30.75	0.906	60 mins
(6) TensoIR	Two far-field	16.24	27.18	0.929	29.68	0.912	28.66	0.902	28.46	0.863	300 mins
(7) VMINer	Two far-field	11.70	26.07	0.902	29.59	0.942	29.27	0.934	29.09	0.890	45 mins
(8) VMINer	Two far + Single near	**10.79**	**32.04**	**0.957**	**32.10**	0.950	**32.38**	**0.954**	**31.40**	**0.910**	60 mins

Real-world results

Applications