ReLight My NeRF: A Dataset for Novel View Synthesis and Relighting of Real World Objects

November 22, 2023

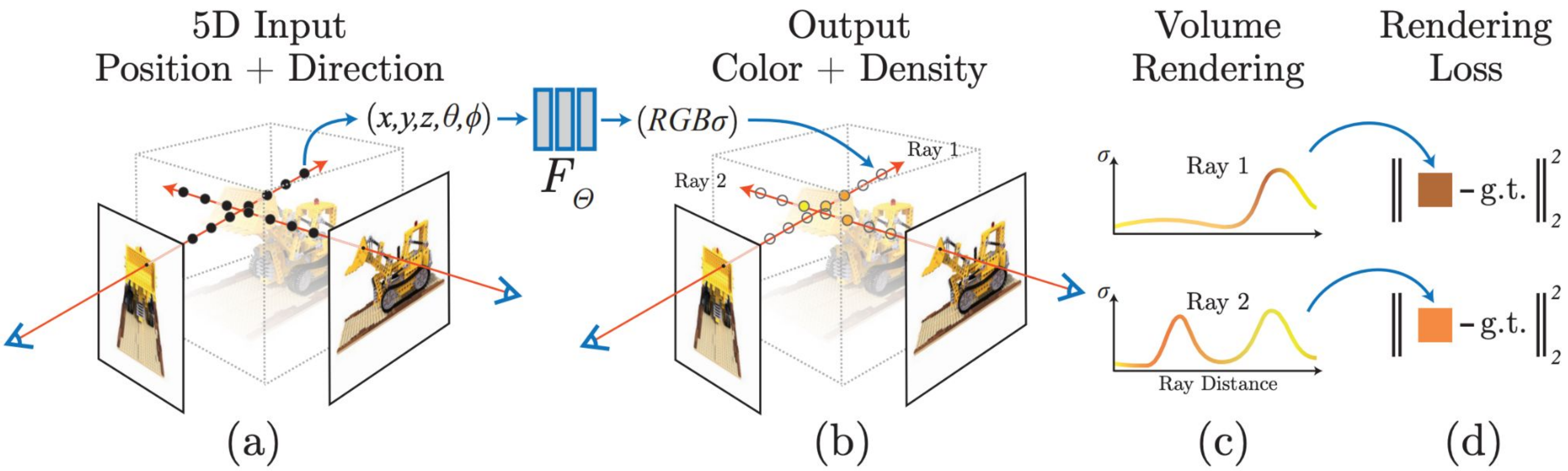
M. Toschi, and et al. In CVPR, 2023.
(Highlight)

Team 5
20233091 Donghwan Kim
20233105 Minje Kim

Contents

1. Background

NeRF (Neural Radiance Fields)



Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, Ren Ng. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In *ECCV*, 2020.

2. Introduction

Limitations of Original NeRF

1. Computation cost and train/inference time
2. Hard to represent deformable objects
3. Hard to generalize to novel scenes
4. Hard to adopt dynamic scenes
5. **Relighting ability**

2. Introduction

Relighting

Varying direction

https://github.com/majedelhelou/VIDIT/tree/master#vidit-virtual-image-dataset-for-illumination-transfer

2. Introduction

Previous Relighting Datasets

Dataset	Multiple categories	Real-World	Background Shadows	Public	Light Supervision
Gross *et al.* [14]	✗	✓	✗	✓	✓
Sun *et al.* [57]	✗	✓	✗	✓	✓
Wang *et al.* [68]	✗	✓	✗	✓	✓
Zhang *et al.* [74]	✗	✓	✗	✓	✓
Srinivasan *et al.* [55]	✓	✗	✗	✓	✓
Zhang *et al.* [75]	✓	✓	✓	✓	✗
Zhang *et al.* [75]	✓	✗	✗	✓	✓
Bi *et al.* [2]	✓	✓	✗	✗	✓
ReNe	✓	✓	✓	✓	✓

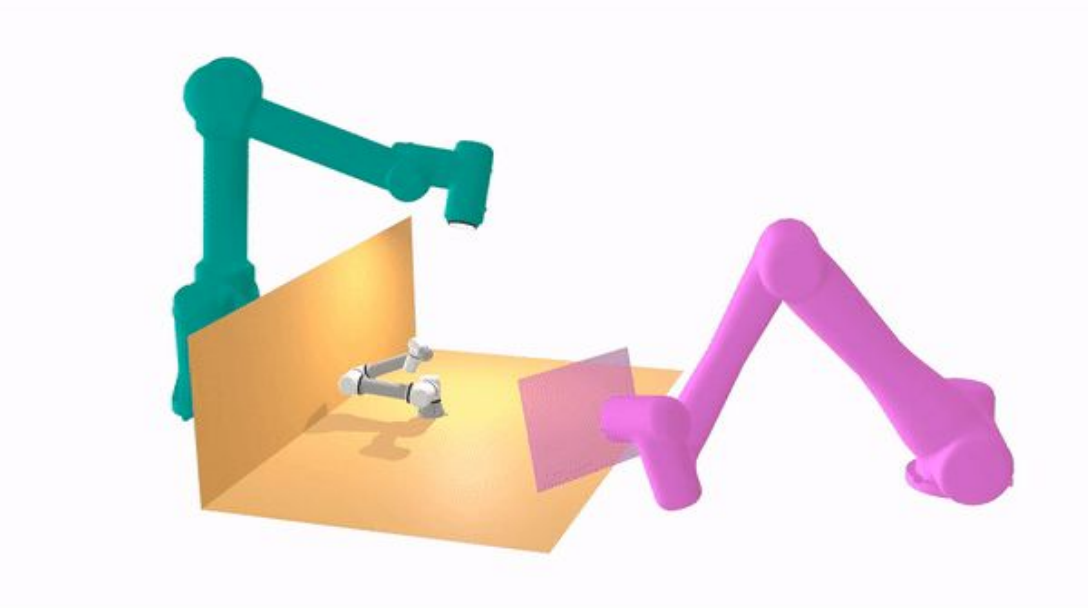
2. Introduction

Previous Relighting Datasets

1. Publicly available
2. Realistic ↔ Synthetic
3. Light supervision
4. Various object categories (↔ portraits)
5. Background shadow (↔ masked-out)
6. Background texture

3. Dataset Acquisition

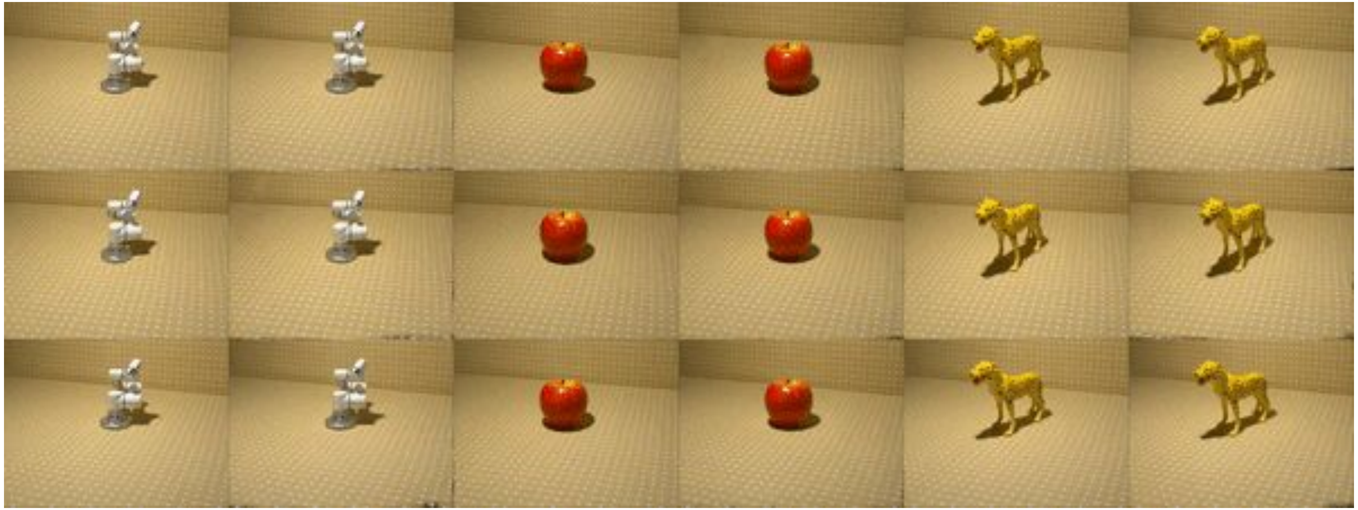
Varying Lights and Cameras Positions



3. Dataset Acquisition

Varying Lights and Cameras Positions

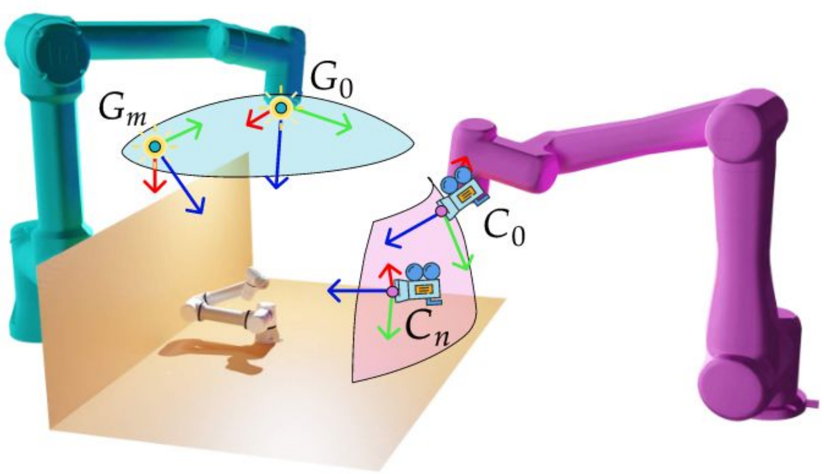
Dataset with images providing both **lights and cameras poses**



3. Dataset Acquisition

Trajectories

Using industrial robots for accurate positions (±0.03mm) and calibration.



3. Dataset Acquisition

Dataset Structure

Total 40,000 images with (S=20) scenes x (N=50) camera positions x (M=40) lighting positions

$$S = \{\mathcal{I}_s\}_{s=1}^{S}$$

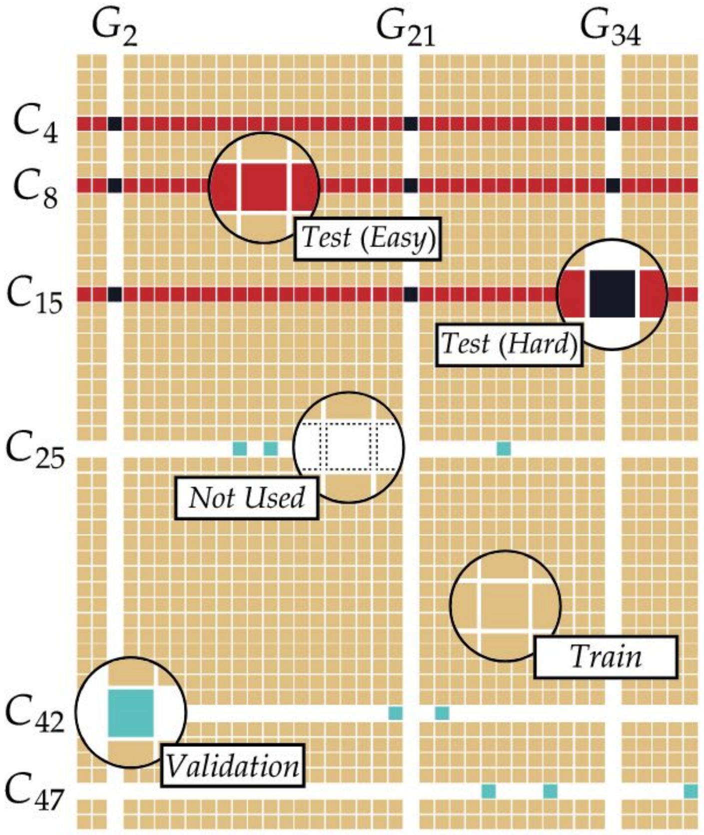
$$\mathcal{I}_s = \{I_{s,n,m} | n = 1, \ldots, N, m = 1, \ldots, M\}$$

$$\mathcal{C}_s = \{C_{s,n}\}_{n=1}^{N} \qquad \mathcal{G}_s = \{G_{s,m}\}_{m=1}^{M}$$

3. Dataset Acquisition

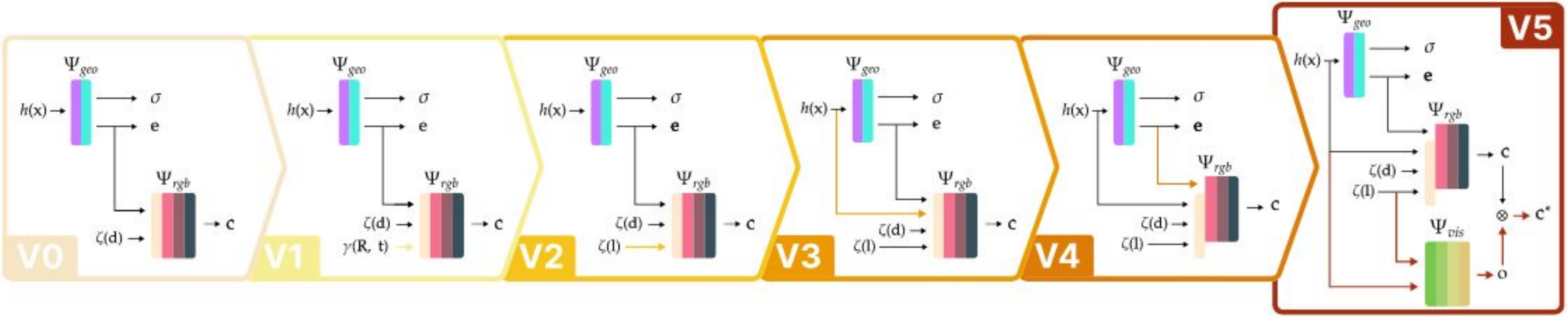
Dataset Split

1. Validation: unseen viewpoint, seen light
2. Test (easy): unseen viewpoint, seen light
3. Test (hard): unseen viewpoint, unseen light



4. Baseline

Overview



4. Baseline

V0 (Instant NGP)

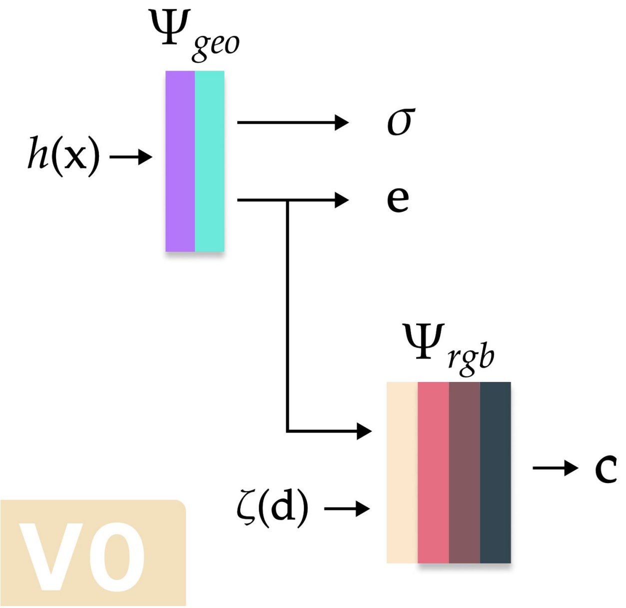
MLP for geometry: Ψ_{geo}

MLP for view-dependent colors: Ψ_{rgb}

Viewing Direction: $\mathbf{d} = (\theta, \phi)$

Multiresolution hash encoding of point: $h(\mathbf{x})$

Spherical harmonics basis: $\zeta(\mathbf{d})$



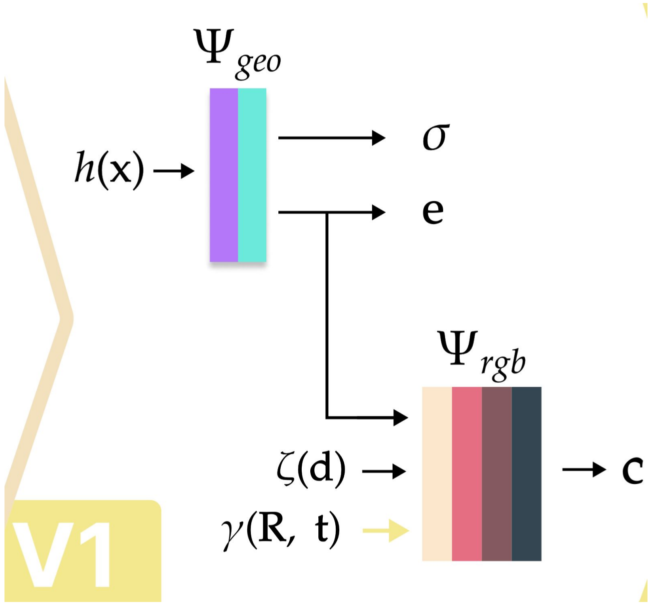
Thomas Müller, Alex Evans, Christoph Schied, Alexander Keller. Instant Neural Graphics Primitives with a Multiresolution Hash Encoding. In SIGGRAPH 2022.

4. Baseline

V1

Straightforward way to enable relighting

Encode rotation and translation of light source using Fourier Features: $\gamma(\mathbf{R}, \mathbf{t})$



4. Baseline

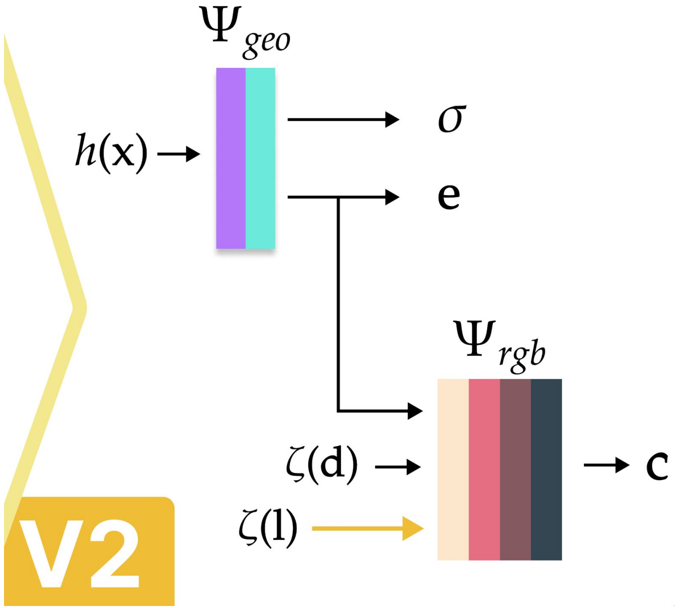
V2

Encode rotation and translation of light source using Fourier Features: $\gamma(\mathbf{R}, \mathbf{t})$

$\rightarrow$

Construct the 3D unit vector: $\mathbf{l} = \frac{\mathbf{t}-\mathbf{x}}{||\mathbf{t}-\mathbf{x}||}$

Encode using the spherical harmonics: $\zeta(\mathbf{l})$



4. Baseline

V3

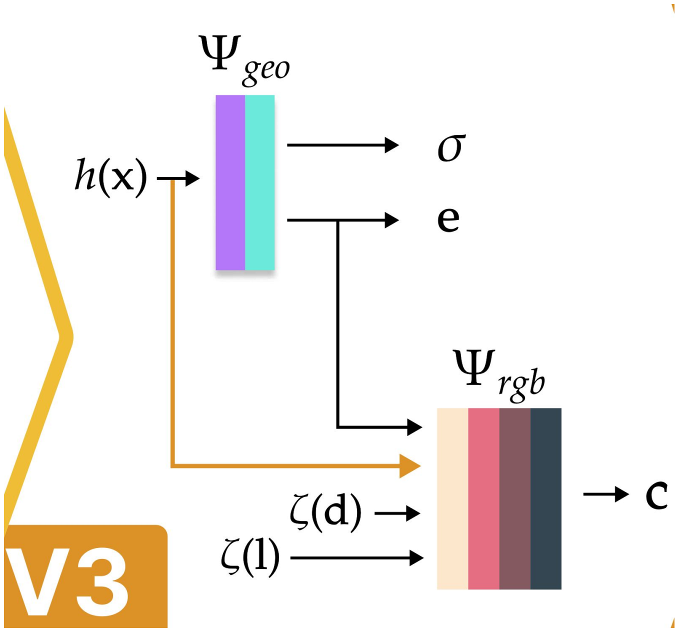
Approximate BRDF with Ψ_{rgb}

incoming light: $\mathbf{l} = \frac{\mathbf{t} - \mathbf{x}}{||\mathbf{t} - \mathbf{x}||}$

outgoing light: $\mathbf{d} = (\theta, \phi)$

feature of the normal vector at **x**: **e**

→ they are relative to **x**, add skip connection

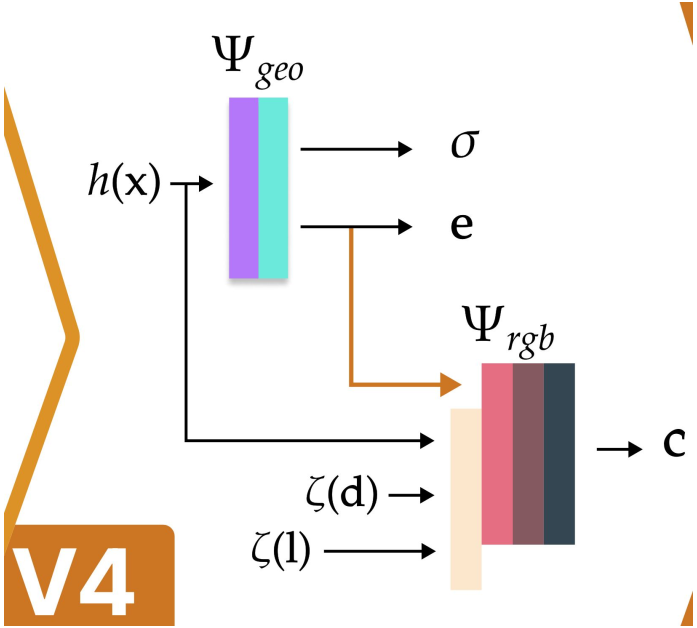


4. Baseline

V4

Consider different scale of input

concatenate feature **e** with output of the first layer



4. Baseline

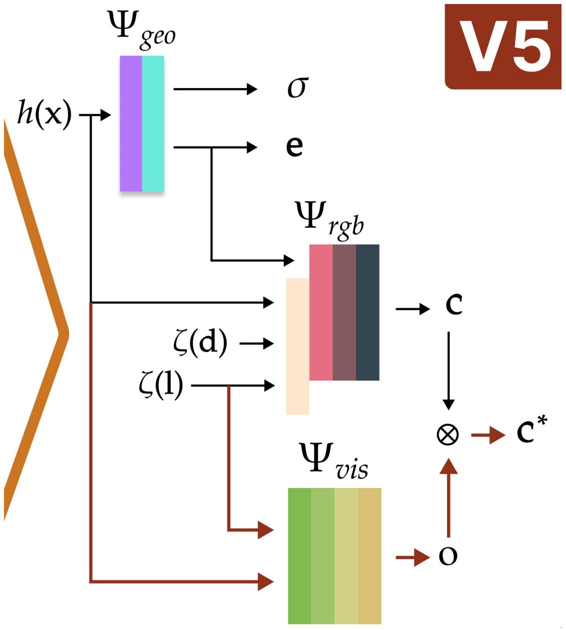
V5

MLP for point-light visibility: Ψ_{vis}

BRDF: $f_p \approx \mathbf{c} = \Psi_{rgb}(h(\mathbf{x}), \mathbf{e}, \zeta(\mathbf{d}), \zeta(\mathbf{l}))$

incoming radiance: $L_i \approx o = \Psi_{vis}(h(\mathbf{x}), \zeta(\mathbf{l}))$

$$L_r(\mathbf{x}, \omega_o) = \int_S f_p(\mathbf{x}, \omega_o, \omega_i) L_i(\mathbf{x}, \omega_i) d\omega_i$$
$$\approx o \cdot \mathbf{c}$$



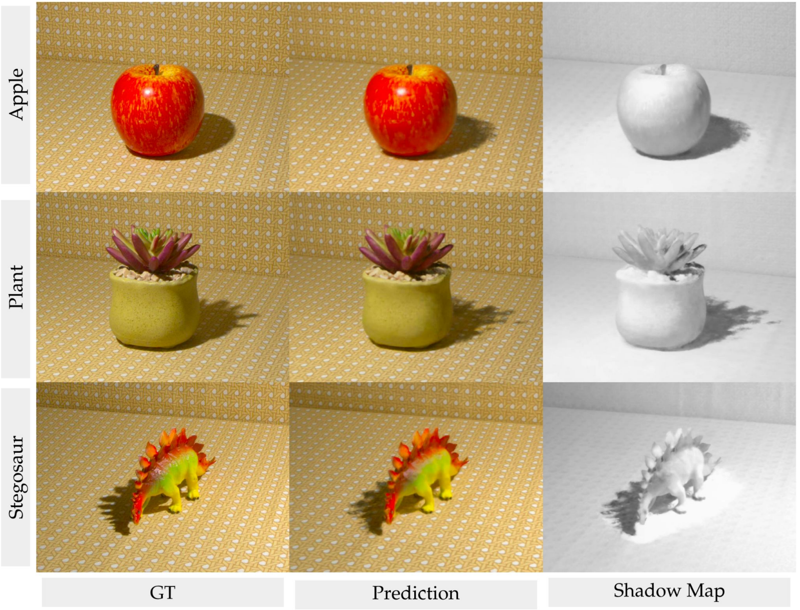
5. Results

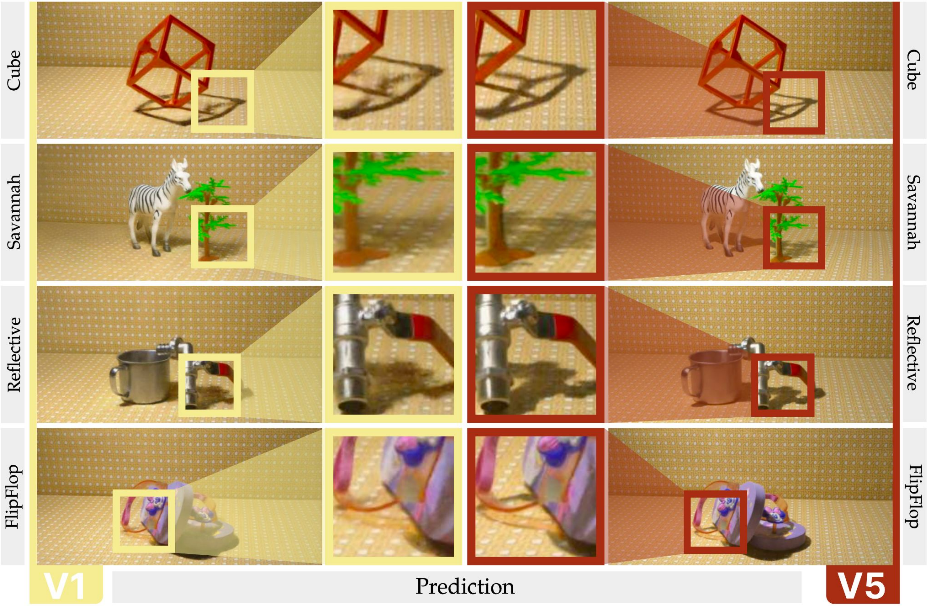
Training time

Methods		Training time	Computation resources
NeRV		1 day	128 TPU cores
NeRD		1.5 days	4 NVIDIA 2080 Ti
NeRFactor	NeRF	6-8 hours	4 NVIDIA TITAN RTX
	normals and visibility (per view)	30 min	1 NVIDIA TITAN RTX
	geometry	20 min	1 NVIDIA TITAN RTX
	joint optimization (per view)	30 min	1 NVIDIA TITAN RTX
Baseline of Relight My NeRF		5 hours	1 NVIDIA 2080 Ti

5. Results

Qualitative Evaluation





5. Results

Quantitative Evaluation (baselines)

Method	$\gamma(\mathbf{R}, \mathbf{t})$	$\mathbf{l}$	*skip*	*inputs*	Ψ_{vis}	Cube		Savannah		Reflective		FlipFlop	
						PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
V1	✓	✗	✗	✗	✗	24.37	0.52	22.53	0.44	23.57	0.51	24.12	0.51
V2	✗	✓	✗	✗	✗	24.73	0.54	23.70	0.52	23.68	0.52	24.42	0.56
V3	✗	✓	✓	✗	✗	25.38	0.56	24.39	0.55	24.65	0.58	25.06	0.57
V4	✗	✓	✓	✓	✗	25.41	0.57	24.79	0.58	24.24	0.56	25.27	0.58
V5	✗	✓	✓	✓	✓	**26.11**	**0.61**	**25.23**	**0.61**	**25.00**	**0.59**	**25.46**	**0.60**

5. Results

Quantitative Evaluation

vs. Instant-NGP

	Ours				*Instant-NGP*			
Name	*Easy*		*Hard*		*Easy*		*Hard*	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Apple	26.44	0.62	26.25	0.62	20.89	0.45	20.95	0.45
Cheetah	25.66	0.61	24.64	0.60	19.37	0.44	19.60	0.44
Cube	24.90	0.54	23.98	0.53	20.14	0.42	20.31	0.42
Dinosaurs	25.75	0.65	24.98	0.64	19.58	0.42	19.66	0.41
FlipFlop	25.85	0.61	25.42	0.61	20.38	0.45	20.36	0.45
Fruits	25.93	0.62	25.72	0.62	20.16	0.45	20.21	0.44
Garden	25.74	0.66	25.08	0.66	19.76	0.45	19.70	0.45
Helicopters	25.12	0.61	24.73	0.61	19.34	0.37	19.37	0.37
Kittens	25.90	0.64	24.96	0.63	18.52	0.37	18.65	0.37
Lego	26.07	0.61	25.77	0.61	20.75	0.46	20.76	0.46
Lunch	25.84	0.60	24.71	0.59	19.32	0.46	19.38	0.45
Plant	26.55	0.67	25.93	0.67	20.62	0.44	20.66	0.44
Reflective	25.79	0.61	25.28	0.61	20.09	0.43	20.11	0.42
Robotoy	26.24	0.65	25.55	0.65	20.77	0.50	20.78	0.50
Savannah	25.15	0.62	24.31	0.61	19.08	0.40	19.18	0.40
Shark	25.59	0.57	25.32	0.56	20.54	0.42	20.53	0.41
Stegosaurus	25.87	0.63	25.65	0.63	20.84	0.43	20.91	0.42
Tapes	25.84	0.58	25.41	0.57	19.34	0.41	19.55	0.41
Trucks	25.80	0.67	25.16	0.66	19.81	0.44	19.87	0.44
Wooden toys	25.69	0.61	25.24	0.60	20.19	0.48	20.22	0.48
Average	**25.79**	**0.62**	**25.20**	**0.61**	19.97	0.43	20.04	0.43

6. Conclusion

Discussion

- Strength
  - A novel dataset for relightable NeRF
  - Baseline that enables relighting of NeRF
- Weakness
  - Limited to single point light
  - Does not support varying light temperature/color

References

Author’s project page and video

https://github.com/majedelhelou/VIDIT/tree/master#vidit-virtual-image-dataset-for-illumination-transfer

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, Ren Ng. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In *ECCV*, 2020.

Thomas Müller, Alex Evans, Christoph Schied, Alexander Keller. Instant Neural Graphics Primitives with a Multiresolution Hash Encoding. In SIGGRAPH 2022.