Uncalibrated Neural Inverse Rendering for Photometric Stereo of General Surfaces

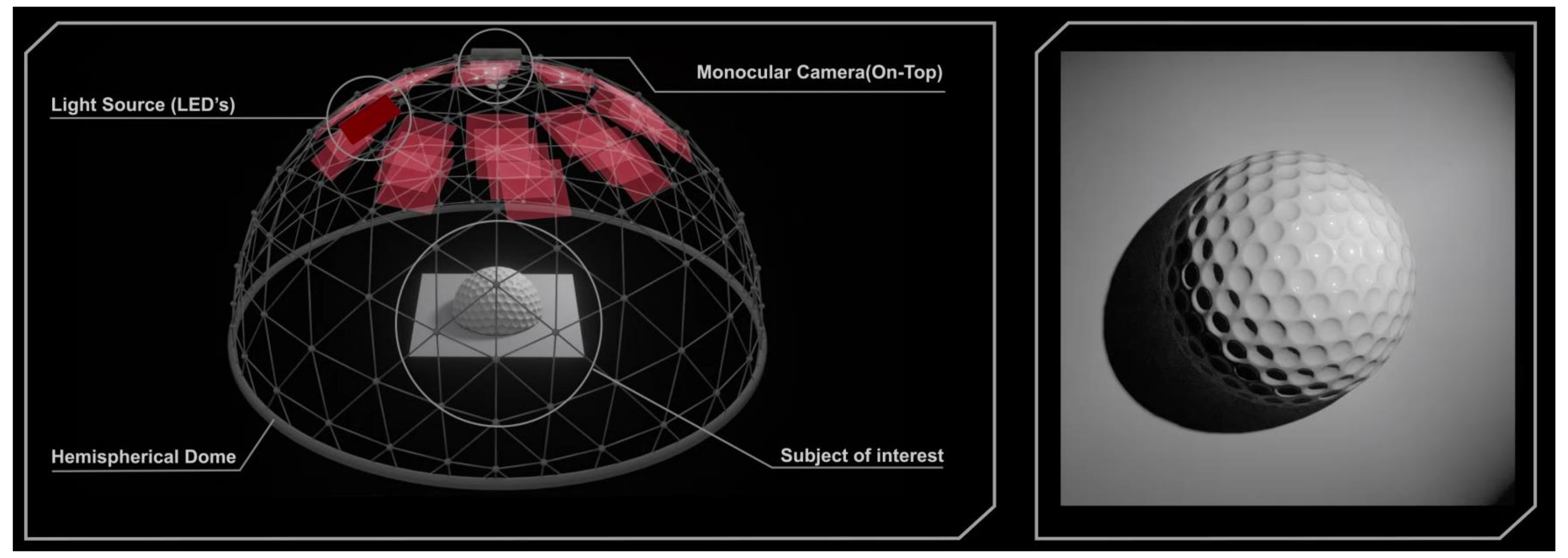
Berk Kaya[1], Suryansh Kumar[1], Carlos Oliveira[1], Vittorio Ferrari[2], Luc Van Gool[1,3]

Computer Vision Lab, ETH Zurich[1], Google Research[2], KU Leuven[3]

CVPR VIRTUAL JUNE 19-25

Overview

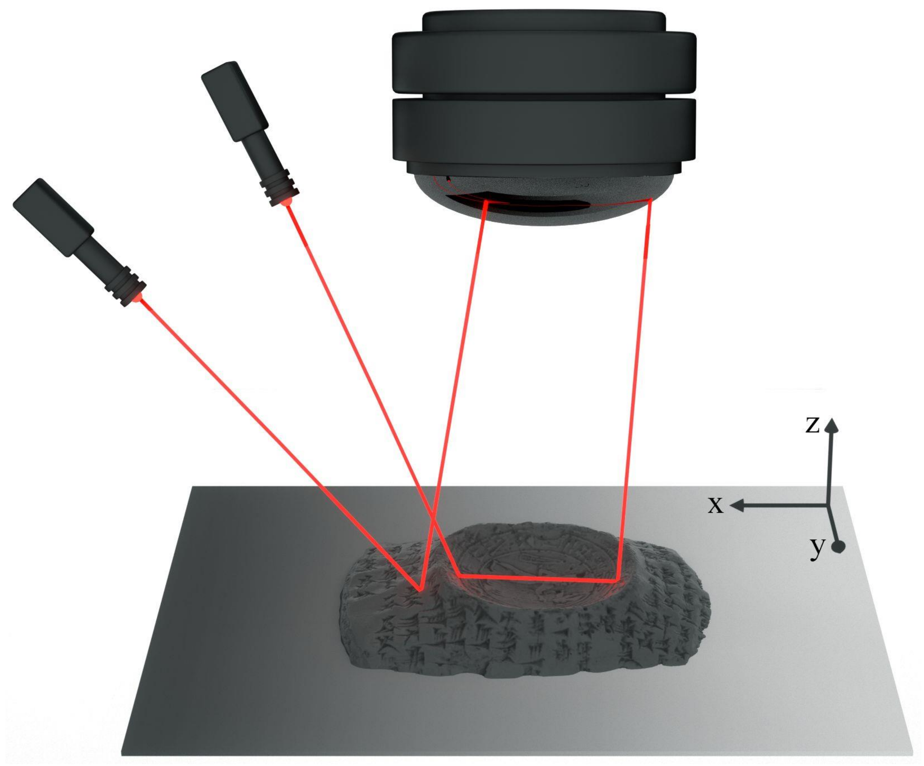
Goal: Estimate surface normals of an object from its light varying images.



Contributions:

- Uncalibrated deep photometric stereo method that does not require **ground-truth surface normals** for training.
- Considers the contribution of both the light source and **interreflections** to the image formation process → enables to handle objects with concave parts
- Leverages neural inverse rendering to infer the surface normal, depth and BRDF values from input images.

Image Formation Model



Classical Photometric Stereo Model:

$$\boldsymbol{X_s} = \rho \boldsymbol{N}^T \boldsymbol{L}$$

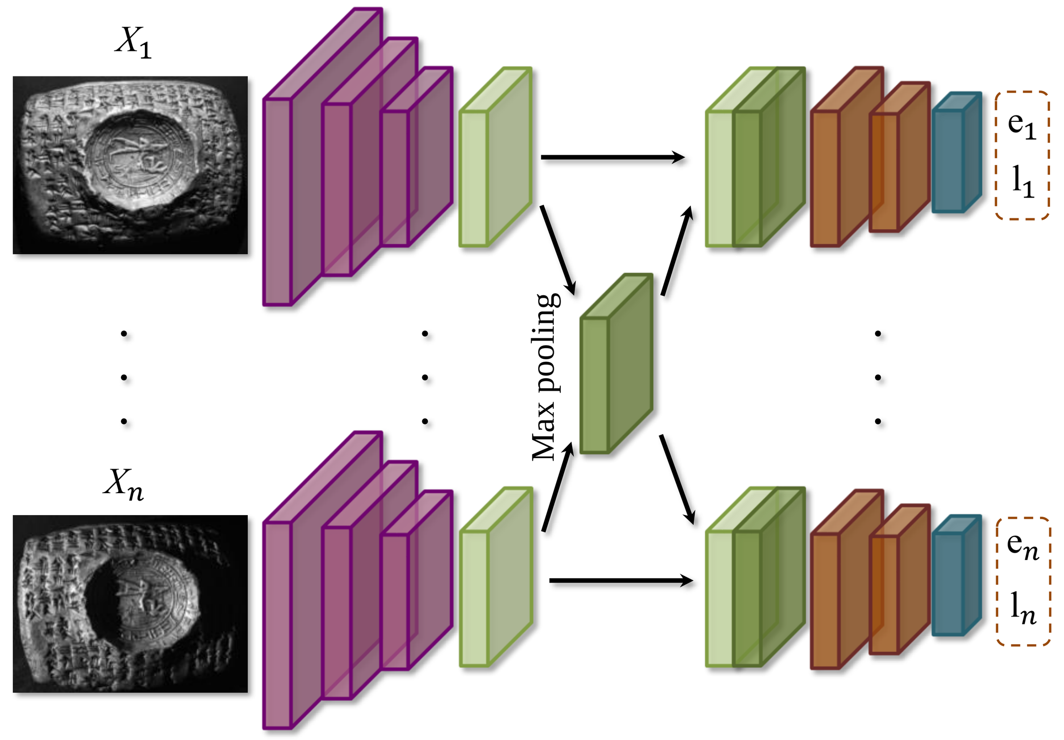
Interreflection Model:

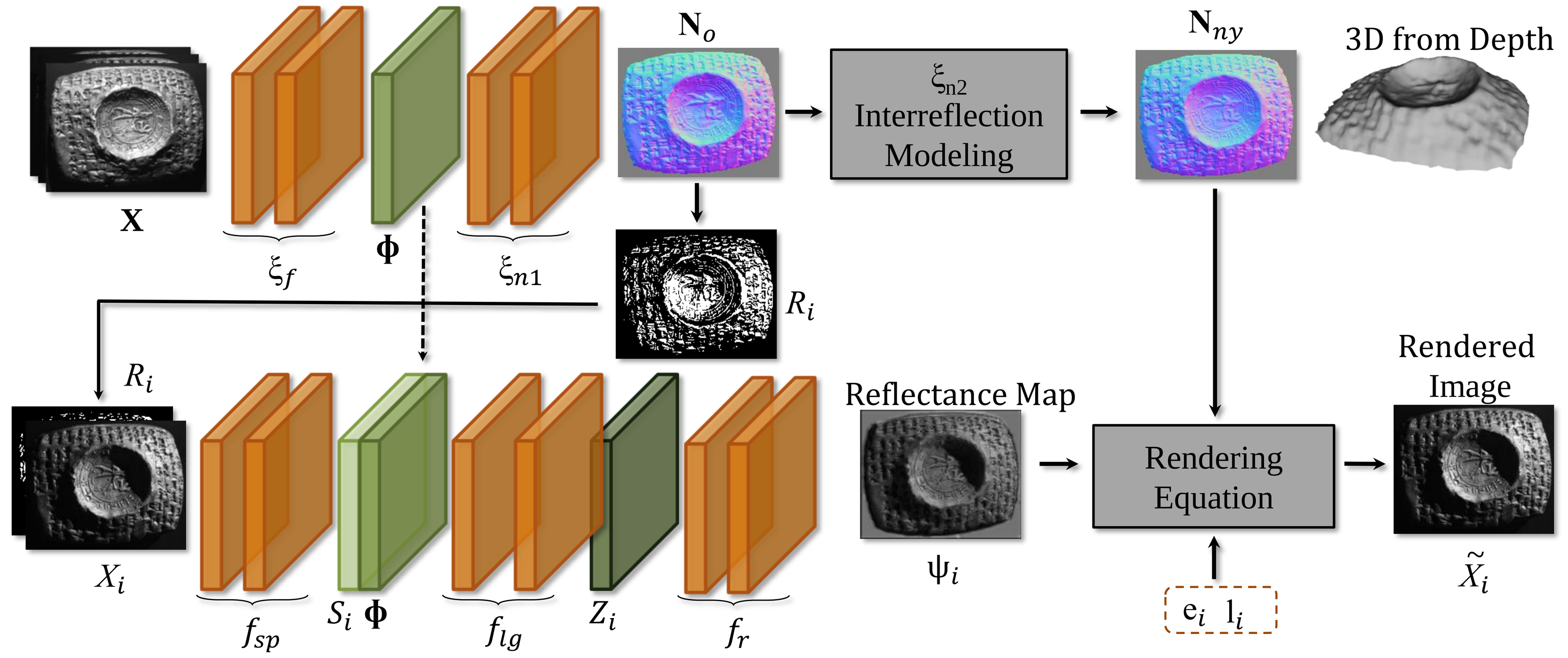
$$X(x) = X_s(x) + \frac{\rho(x)}{\pi} \int_{\Omega} K(x,x') X(x') dx'$$

$$K(x,x') = \left(\frac{\left(n(x)^T(-r)\right) \cdot \left(n(x')^T r\right) \cdot V(x,x')}{(r^T r)^2} \right)$$

$$\boldsymbol{X} = (I - PK)^{-1} \boldsymbol{X_s}$$

Network Architecture and Training





$$\mathcal{L} = \mathcal{L}_{rec}(\boldsymbol{X}, \tilde{\boldsymbol{X}}) + \lambda_w \mathcal{L}_{weak}(\boldsymbol{N}_{ny}, \boldsymbol{N}_{init})$$

Reconstruction Loss | Weak Supervision with Robust Initialization

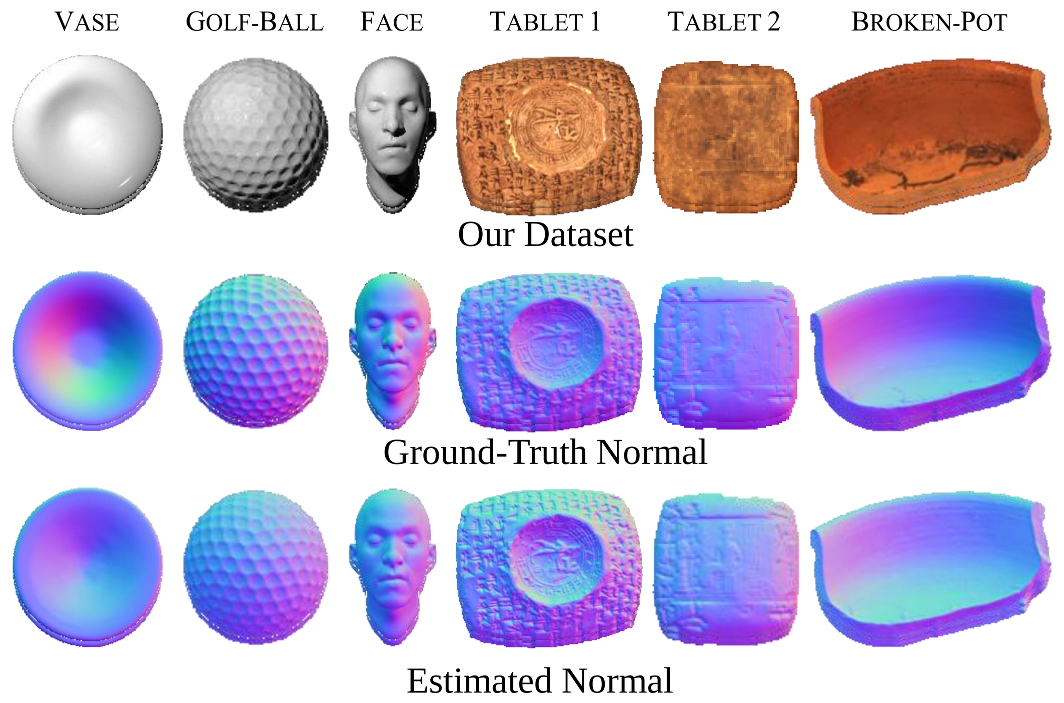
- **Light estimation network** initially predicts the light source directions and intensities to resolve Generalized Bas-Relief (GBR) ambiguity.
- Predicts azimuth, elevation and intensity value of each light source

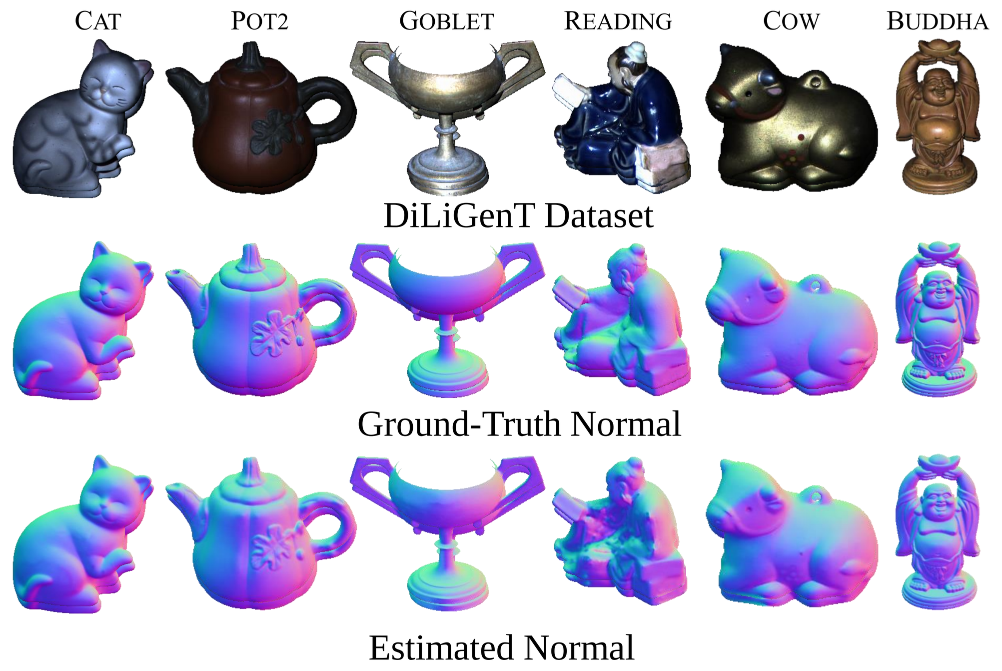
- **Inverse rendering network** estimates surface normals and reflectance map from input images.
- We use estimated normal and reflectance in rendering equation to reconstruct images.
- Our rendering equation uses interreflection modeling to handle concave surfaces.
- We use an initial normal estimate at early stages to warm-up the network.

Results

- We propose a new dataset to study complex surfaces, composed of concave and convex parts.
- Our method performs consistently well on complex surfaces without using ground-truth surface normals.

Type	G.T. Normal	Methods↓ | Dataset →	Vase	Golf-ball	Face	Tablet 1	Tablet 2	Broken Pot	Average
Classical	✗	Nayar et al.(1991)	28.82	11.30	13.97	19.14	16.34	19.43	18.17
NN-based	✓	Chen et al.(2018)	35.79	36.14	48.47	19.16	**10.69**	24.45	29.12
NN-based	✓	Chen et al.(2019)	49.36	31.61	13.81	16.00	15.11	**18.34**	24.04
NN-based	✗	**Ours**	**19.91**	**11.04**	**13.43**	**12.37**	13.12	18.55	**14.74**





Conclusion

- Uncalibrated neural inverse rendering framework with explicit interreflection modeling.
- Without using ground-truth normals, we perform comparable or better than supervised approaches.
- Our method is applicable to broader range of surfaces, including convex and concave parts.