CVPR VIRTUAL JUNE 19-25

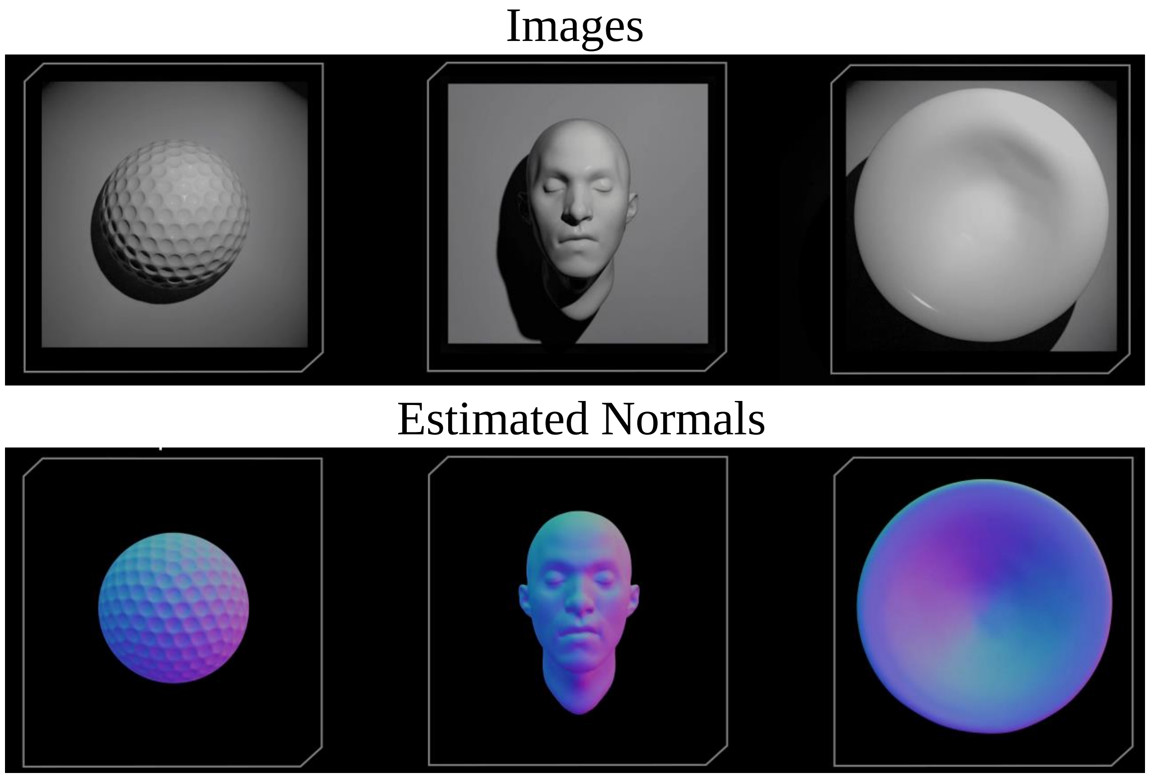
# Uncalibrated Neural Inverse Rendering for Photometric Stereo of General Surfaces

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

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

# Photometric Stereo

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




**Calibrated:** Known light sources

✗ Exact calibration is difficult

**Uncalibrated:** Unkown light sources

✓ Better generalizability

# Photometric Stereo

**Traditional Methods**

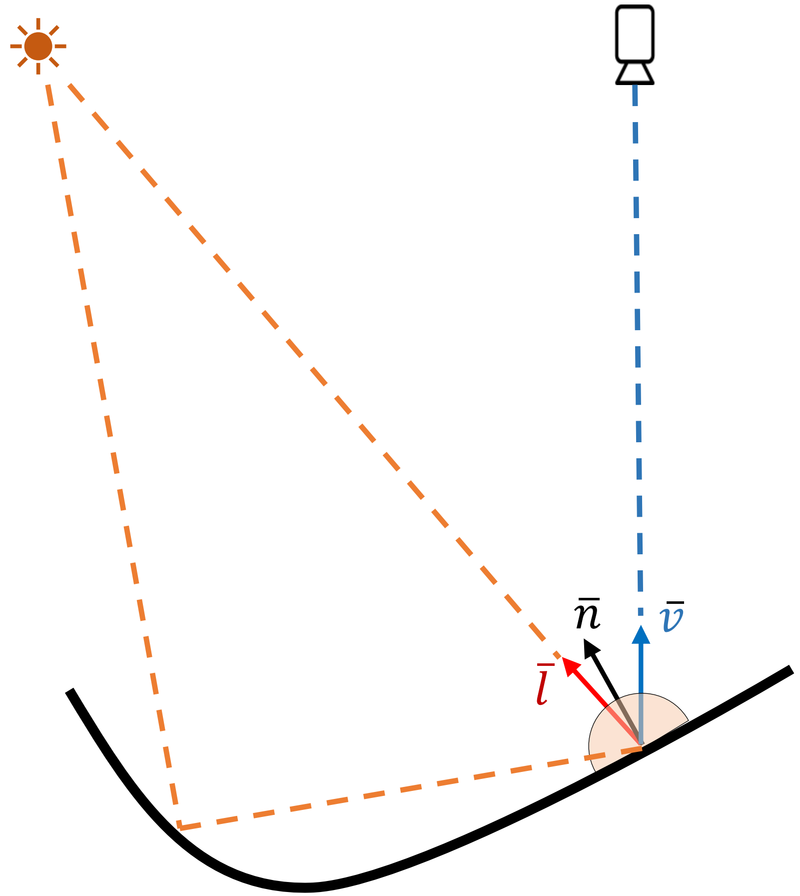
[Woodham 1980, Wu et al. 2010]

- ✗ Simple reflectance models

**Deep Learning Methods**

[Ikehata 2018, Chen et al. 2019]

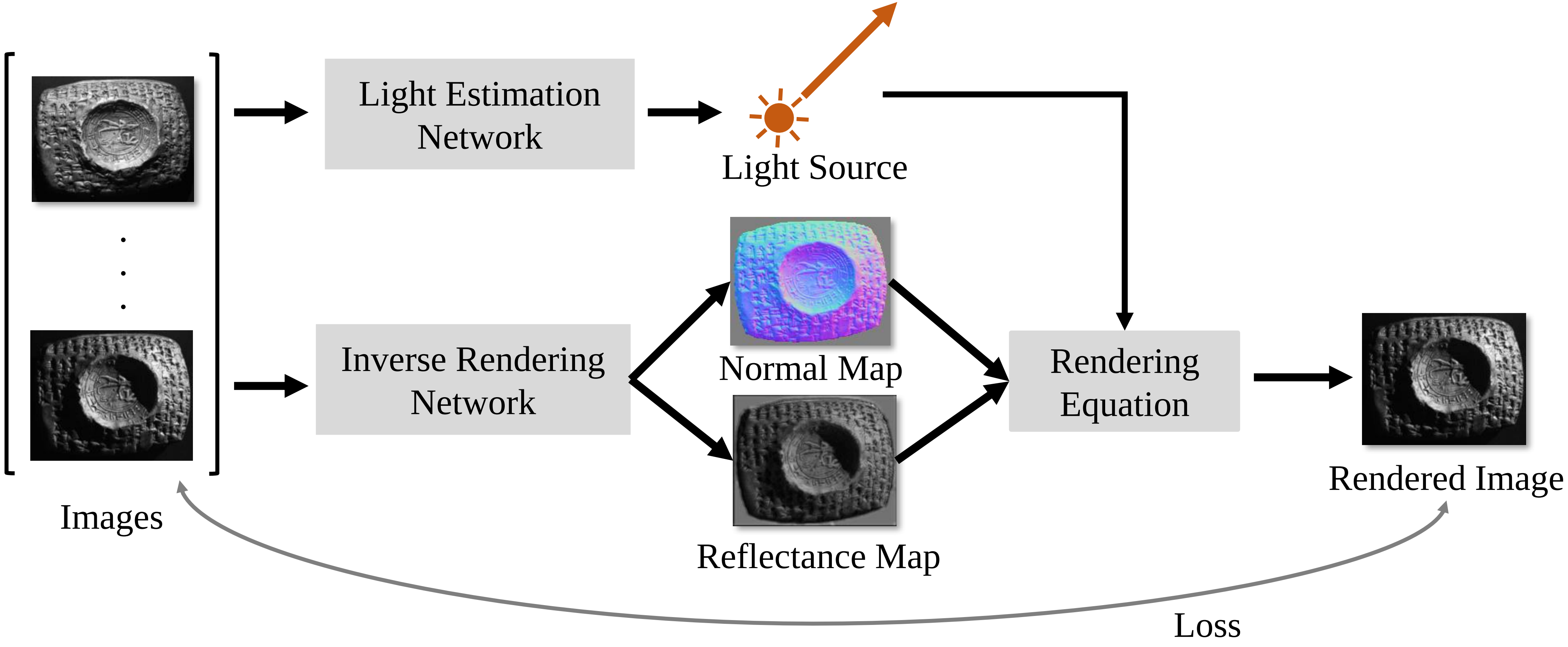
- ✓ Learn complex BRDF from data
- ✗ Require training data with ground-truth normals
- ✗ Cannot handle interreflections on concave surfaces




**Classical Photometric Stereo Model:**

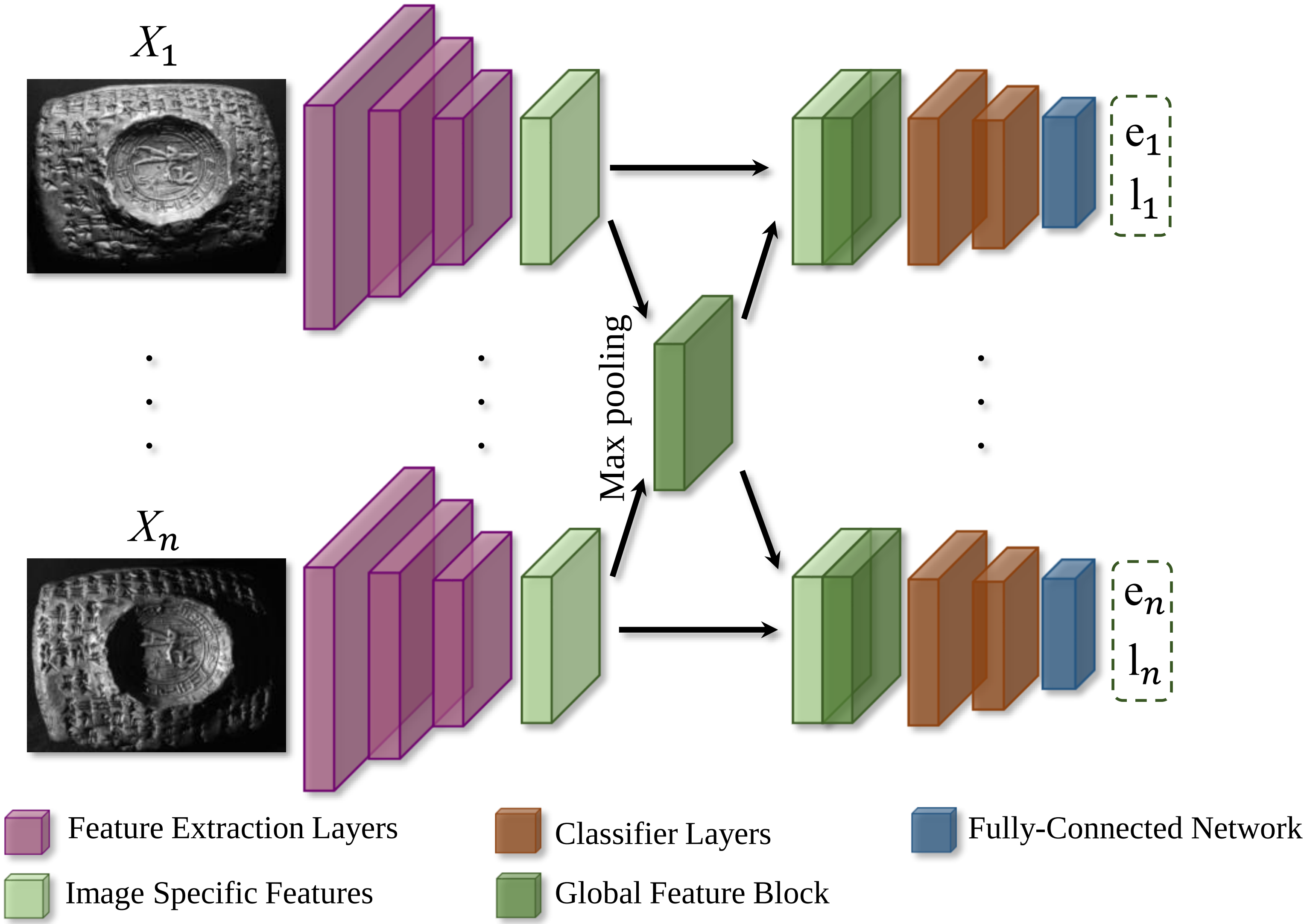
$$\boldsymbol{X_s} = \rho \boldsymbol{N}^T \boldsymbol{L}$$

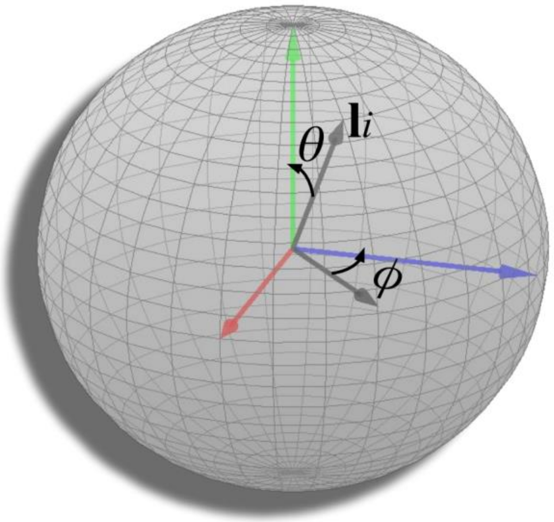
# Overview of Our Method




- Uncalibrated thanks to light estimation network
- Loss on the rendered image → does not require ground-truth surface normals
- Explicit interreflection modeling in rendering equation → enables to handle objects with concave parts

# Light Estimation Network

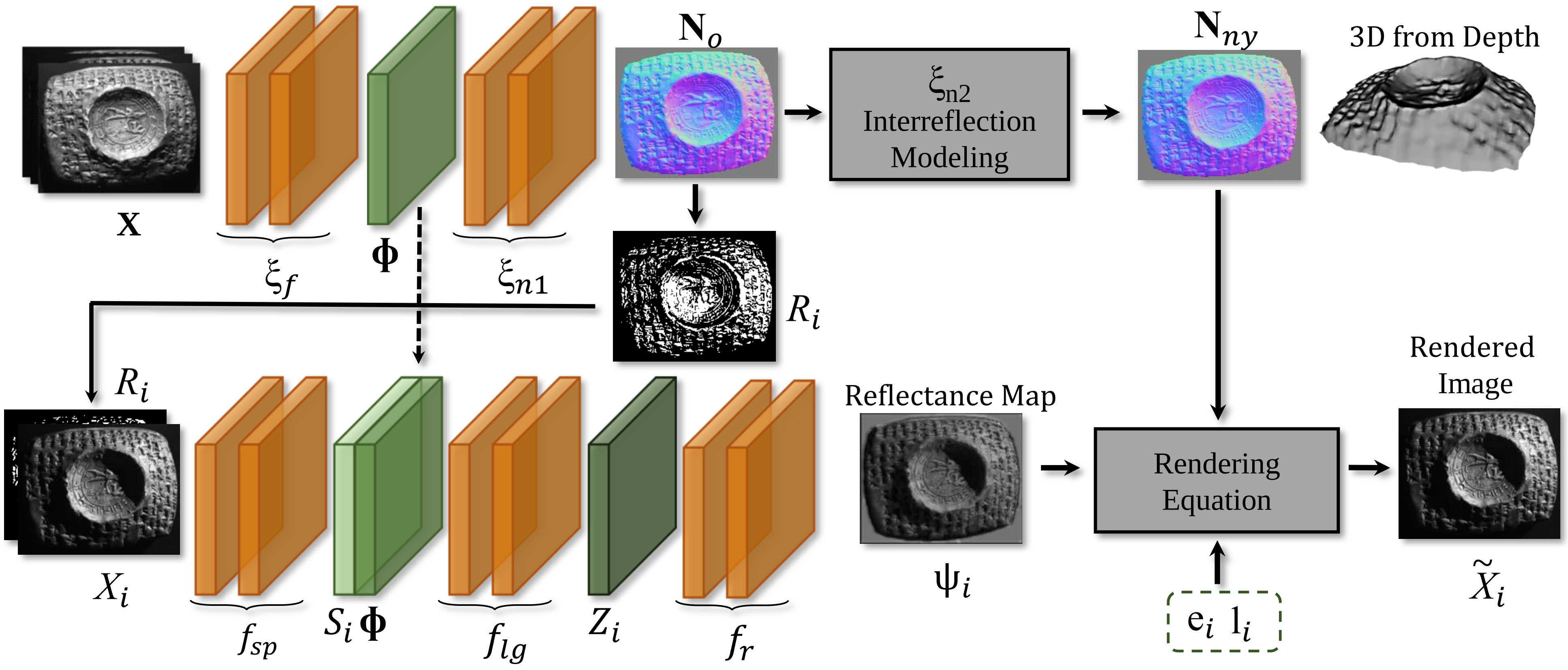



Light directions can be represented by azimuth and elevation angles

$$\mathcal{L}_{calib} = \mathcal{L}_{az} + \mathcal{L}_{ele} + \mathcal{L}_{in}$$

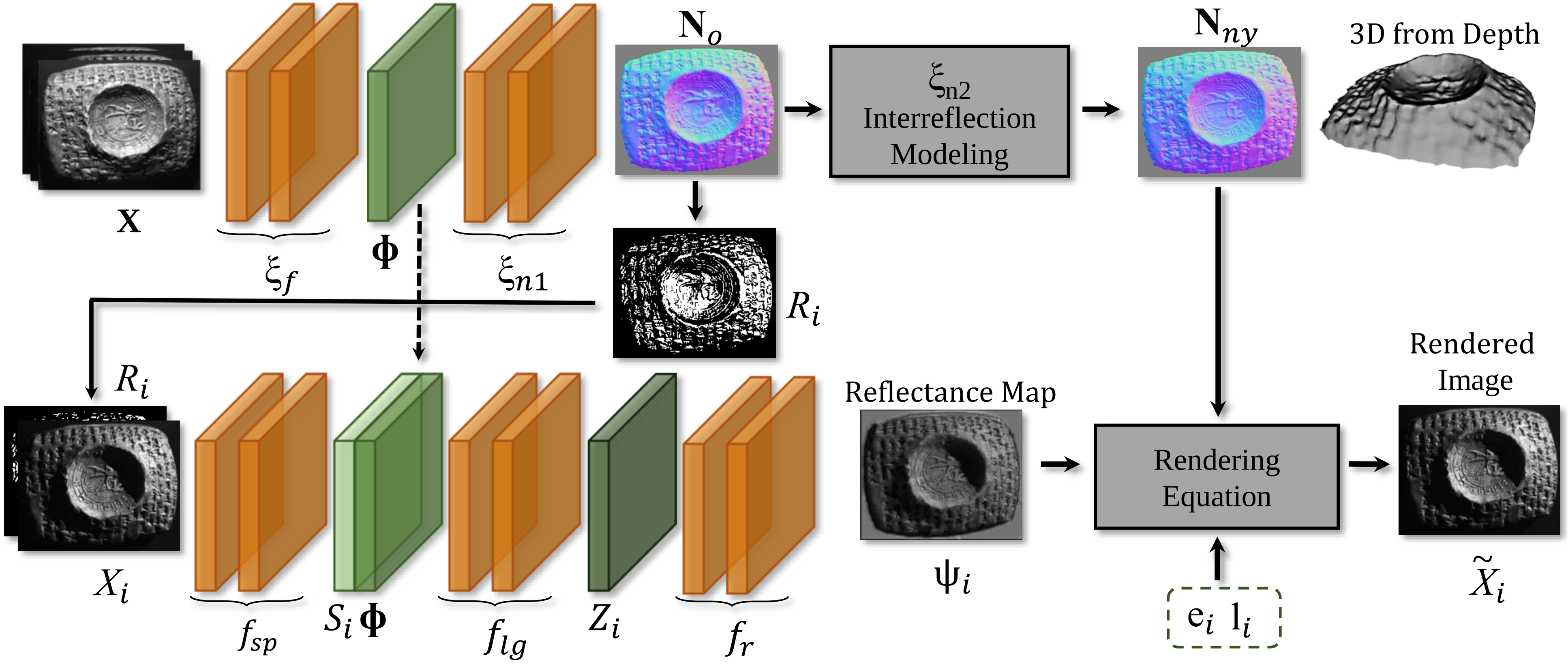
# Inverse Rendering Network




**Interreflection Model:**

$$X(x) = X_s(x) + \frac{\rho(x)}{\pi}\int_{\Omega} K(x,x')X(x')dx' \qquad K(x,x') = \left(\frac{\big(n(x)^T(-r)\big)\cdot\big(n(x')^T r\big)\cdot V(x,x')}{(r^T r)^2}\right) \longrightarrow \boldsymbol{X} = (I - PK)^{-1}\boldsymbol{X}_s$$

# Inverse Rendering Network

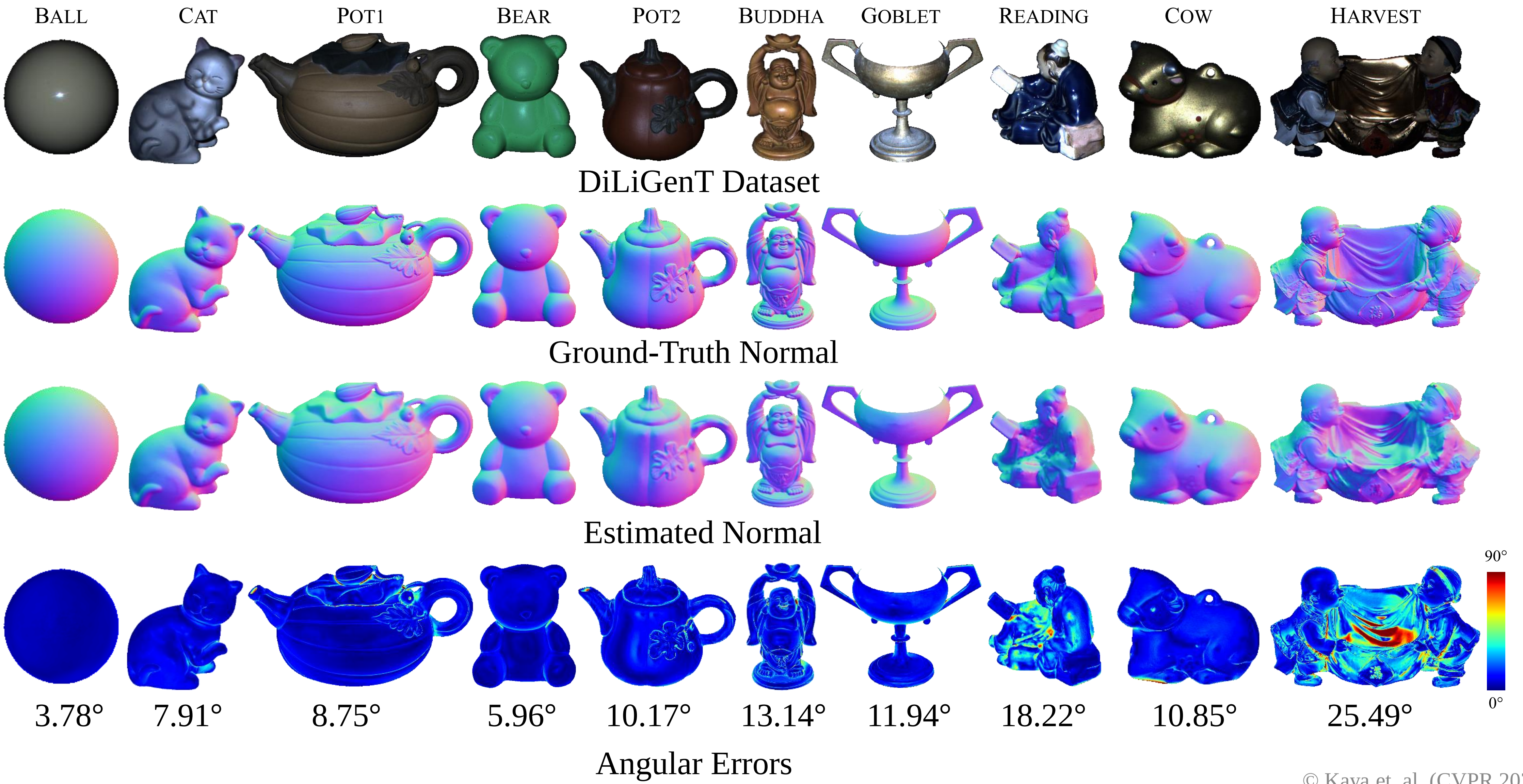



$$\mathcal{L} = \mathcal{L}_{rec}(\boldsymbol{X}, \widetilde{\boldsymbol{X}}) + \lambda_w \mathcal{L}_{weak}(\boldsymbol{N}_{ny}, \boldsymbol{N}_{init})$$

Reconstruction Loss

Weak supervision with robust initialization

# Results on DiLiGenT Dataset



# Results on DiLiGenT Dataset

| Type | G.T. Normal | Methods↓ \| Dataset → | Ball | Cat | Pot1 | Bear | Pot2 | Buddha | Goblet | Reading | Cow | Harvest | Average |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Classical | ✗ | Alldrin et al.(2007) | 7.27 | 31.45 | 18.37 | 16.81 | 49.16 | 32.81 | 46.54 | 53.65 | 54.72 | 61.70 | 37.25 |
| Classical | ✗ | Shi et al.(2010) | 8.90 | 19.84 | 16.68 | 11.98 | 50.68 | 15.54 | 48.79 | 26.93 | 22.73 | 73.86 | 29.59 |
| Classical | ✗ | Wu et al.(2013) | 4.39 | 36.55 | 9.39 | 6.42 | 14.52 | 13.19 | 20.57 | 58.96 | 19.75 | 55.51 | 23.93 |
| Classical | ✗ | Lu et al.(2013) | 22.43 | 25.01 | 32.82 | 15.44 | 20.57 | 25.76 | 29.16 | 48.16 | 22.53 | 34.45 | 27.63 |
| Classical | ✗ | Pap. et al.(2014) | 4.77 | 9.54 | 9.51 | 9.07 | 15.90 | 14.92 | 29.93 | 24.18 | 19.53 | 29.21 | 16.66 |
| Classical | ✗ | Lu et al.(2017) | 9.30 | 12.60 | 12.40 | 10.90 | 15.70 | 19.00 | 18.30 | 22.30 | 15.00 | 28.00 | 16.30 |
| NN-based | ✓ | Chen et al.(2018) | 6.62 | 14.68 | 13.98 | 11.23 | 14.19 | 15.87 | 20.72 | 23.26 | 11.91 | 27.79 | 16.02 |
| NN-based | ✓ | Chen et al.(2018)$^\dagger$ | 3.96 | 12.16 | 11.13 | 7.19 | 11.11 | 13.06 | 18.07 | 20.46 | 11.84 | 27.22 | 13.62 |
| NN-based | ✓ | Chen et al.(2019) | **2.77** | 8.06 | **8.14** | 6.89 | **7.50** | **8.97** | **11.91** | **14.90** | **8.48** | **17.43** | **9.51** |
| NN-based | ✗ | **Ours** | 3.78 | **7.91** | 8.75 | **5.96** | 10.17 | 13.14 | 11.94 | 18.22 | 10.85 | 25.49 | 11.62 |

Comparison against uncalibrated photometric stereo methods on DiLiGenT [Shi et al. 2016].
We report mean angular errors in degrees.

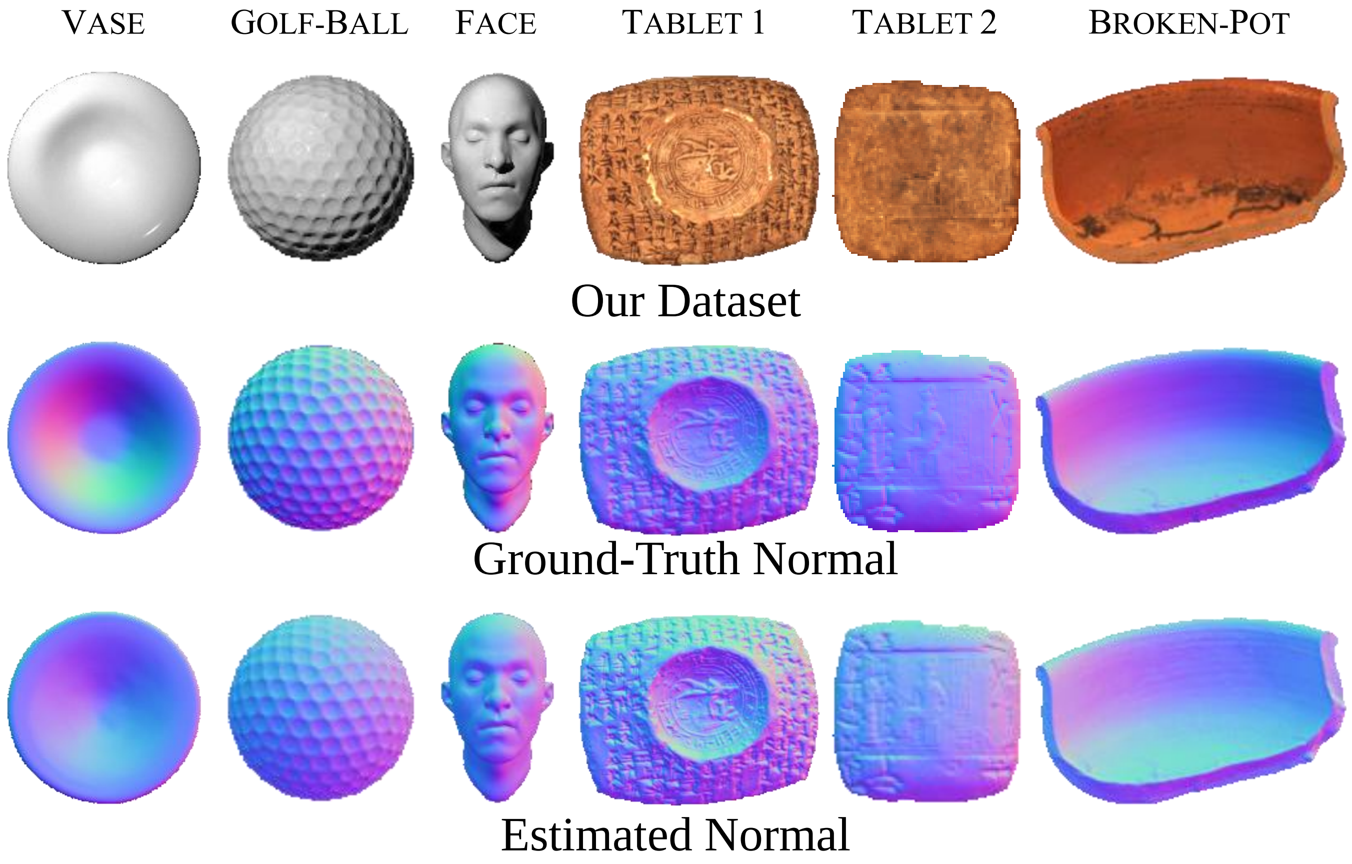
# Our Dataset

- We propose a new dataset for analyzing complex surfaces, including both convex and concave parts.

Acquisition Setup Images

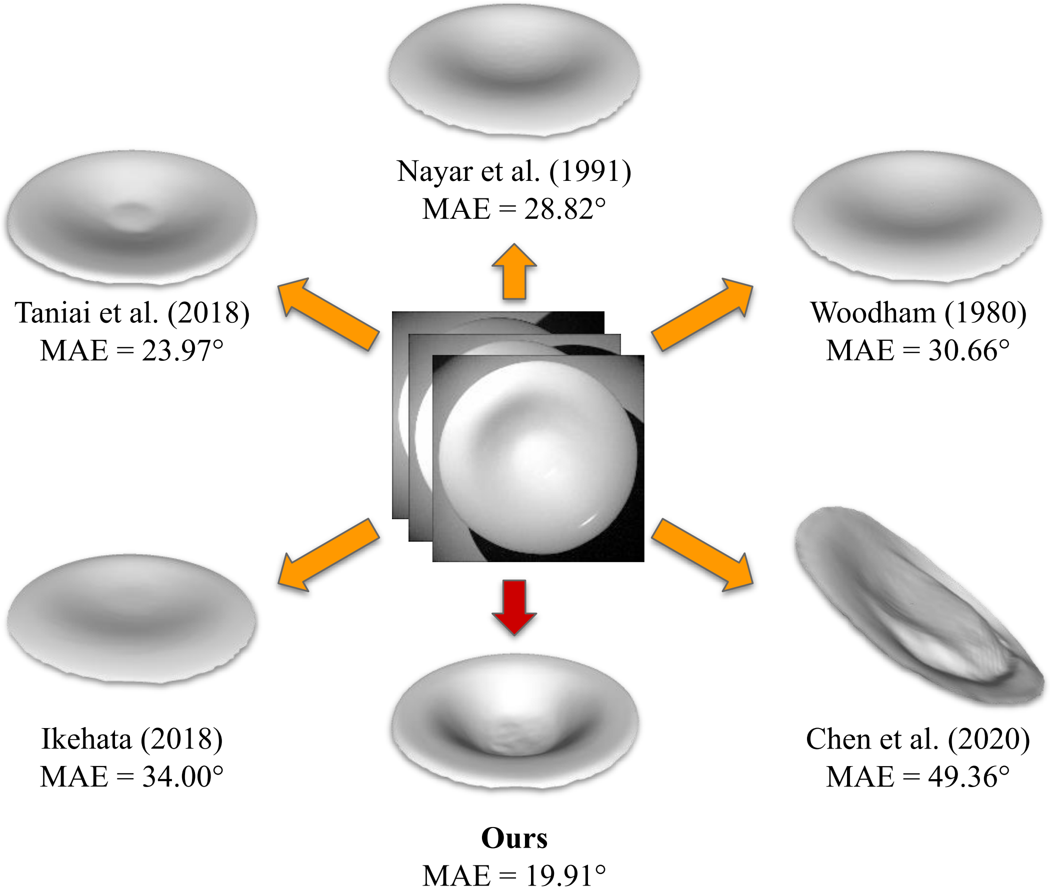
# Results on Our Dataset

| 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.
- Performs comparable or better than supervised approaches.
- Applicable to broader range of surfaces, composed of convex and concave parts.



Thank you for your attention!