

Glass Object Appearance Acquisition Using Learning-Based Inverse Rendering

Jeppe Revall Frisvad

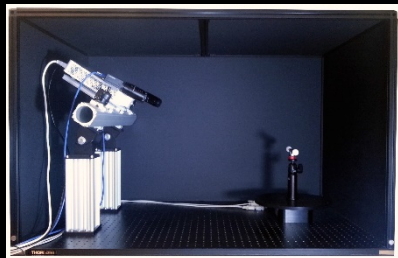
Image Section

Department of Applied Mathematics and Computer Science

Technical University of Denmark

Thanks to Jonathan Dyssel Stets *et al.* [2017; 2019].

[Eiriksson et al. 2016]



[Abildgaard et al. 2015]



[Frisvad et al. 2007]



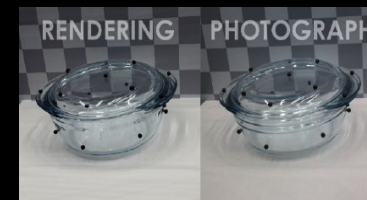
Sensors



Physical-digital
3D ecosystem



Synthesis, Prediction & Modeling



[Stets et al. 2017]

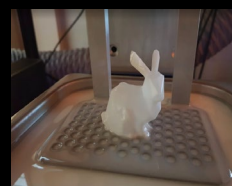
[Dal Corso et al. 2016]

Physical World

Digital Representation

Actuators

[Luongo et al. 2019]



[Eiriksson 2018]

Geometry & Appearance Digitization

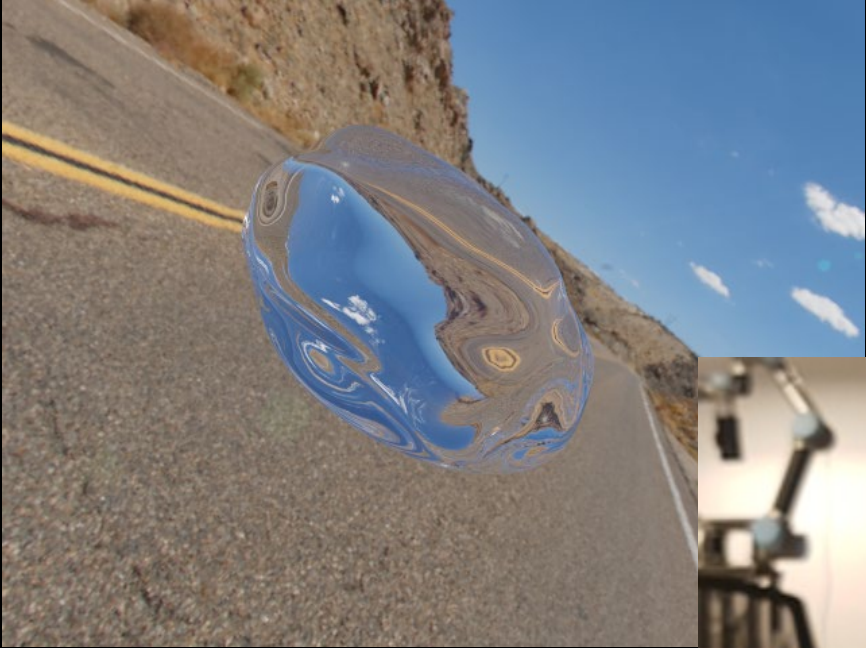


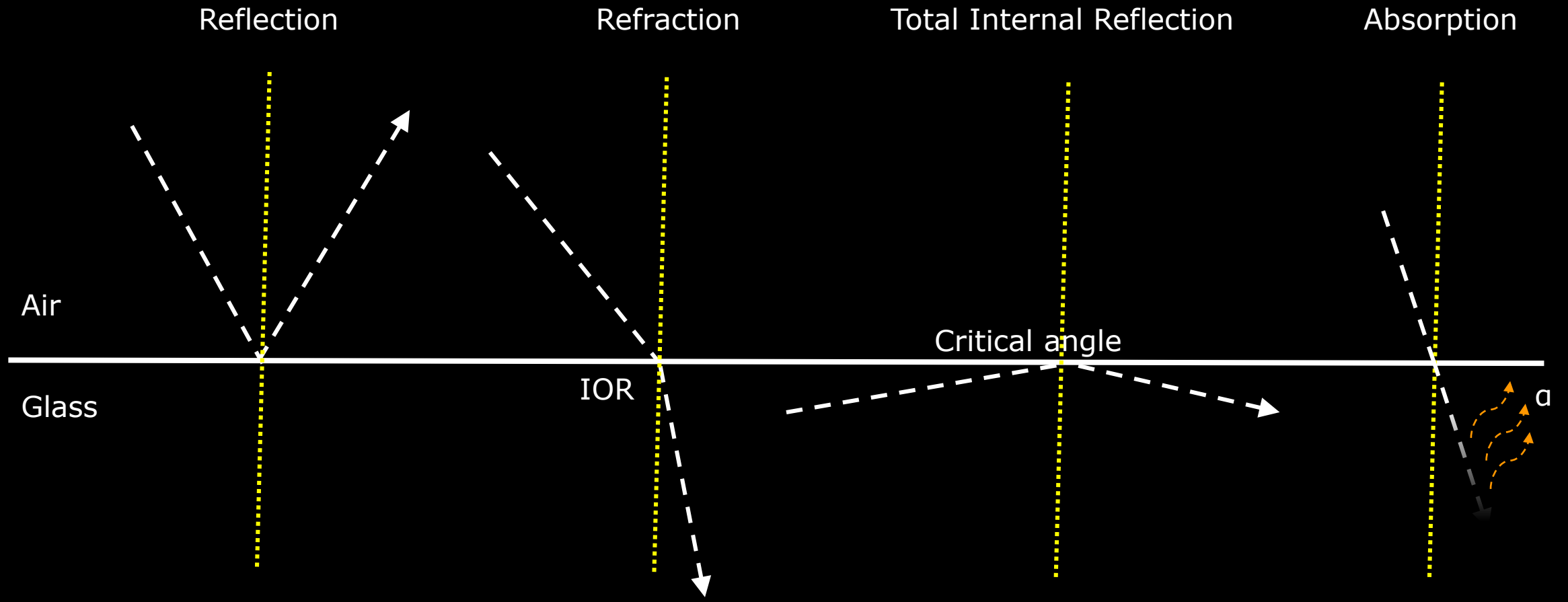
Geometry & Appearance Digitization



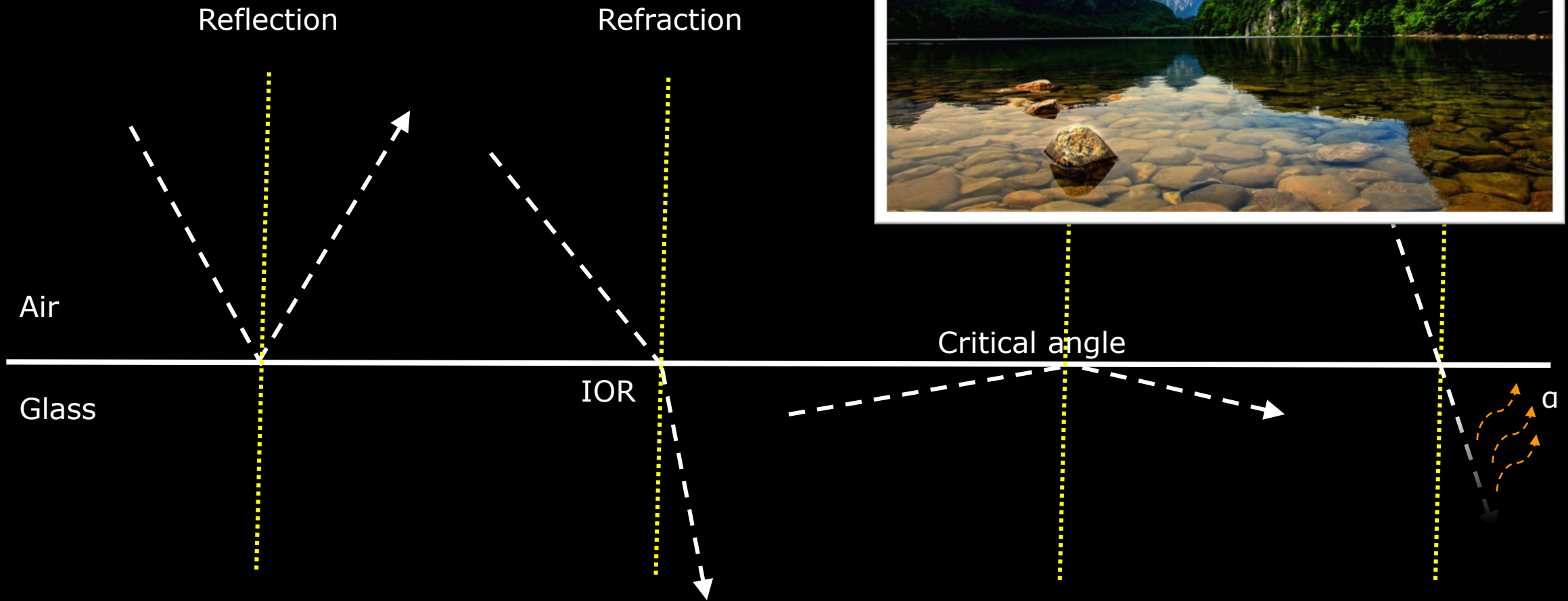
Geometry & Appearance Digitization

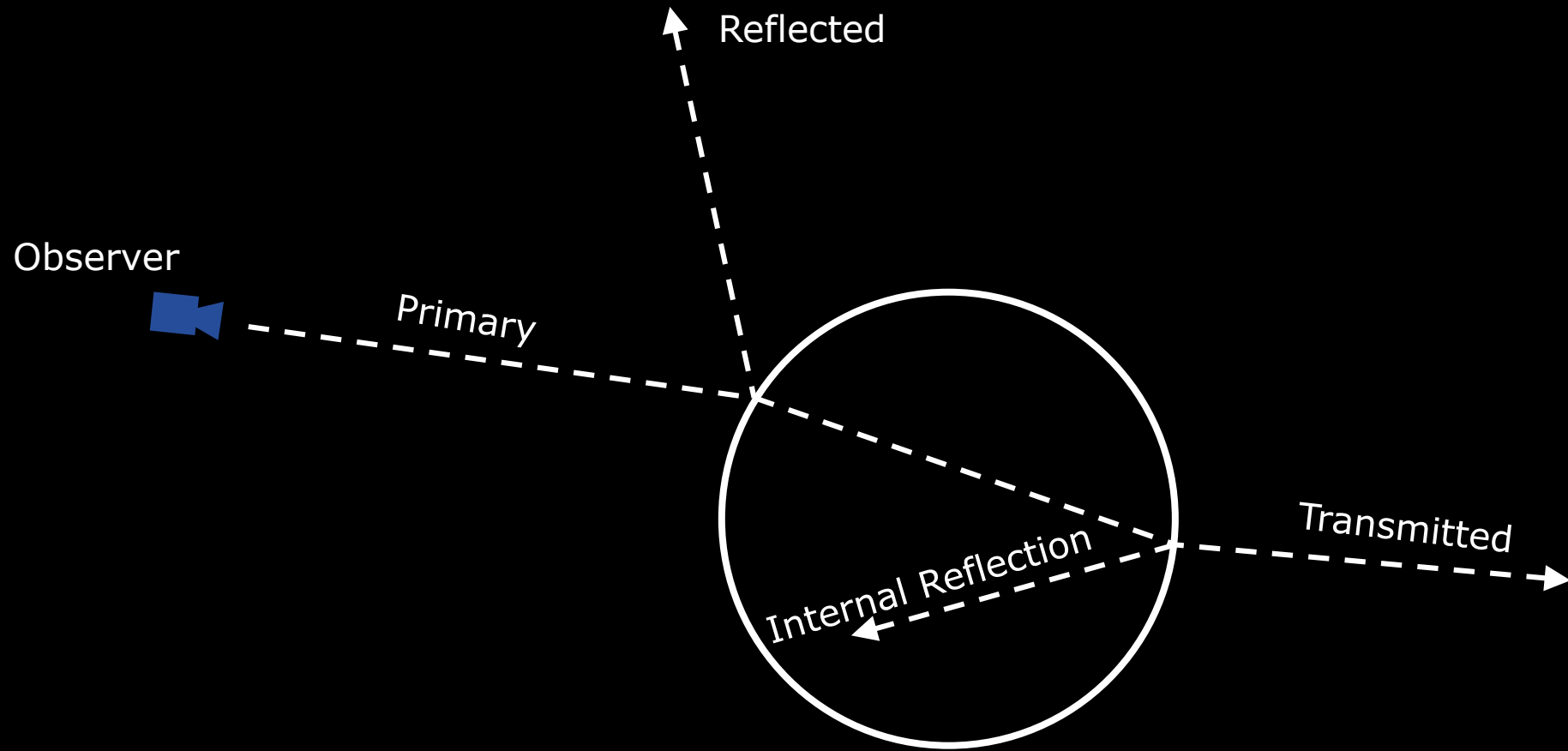


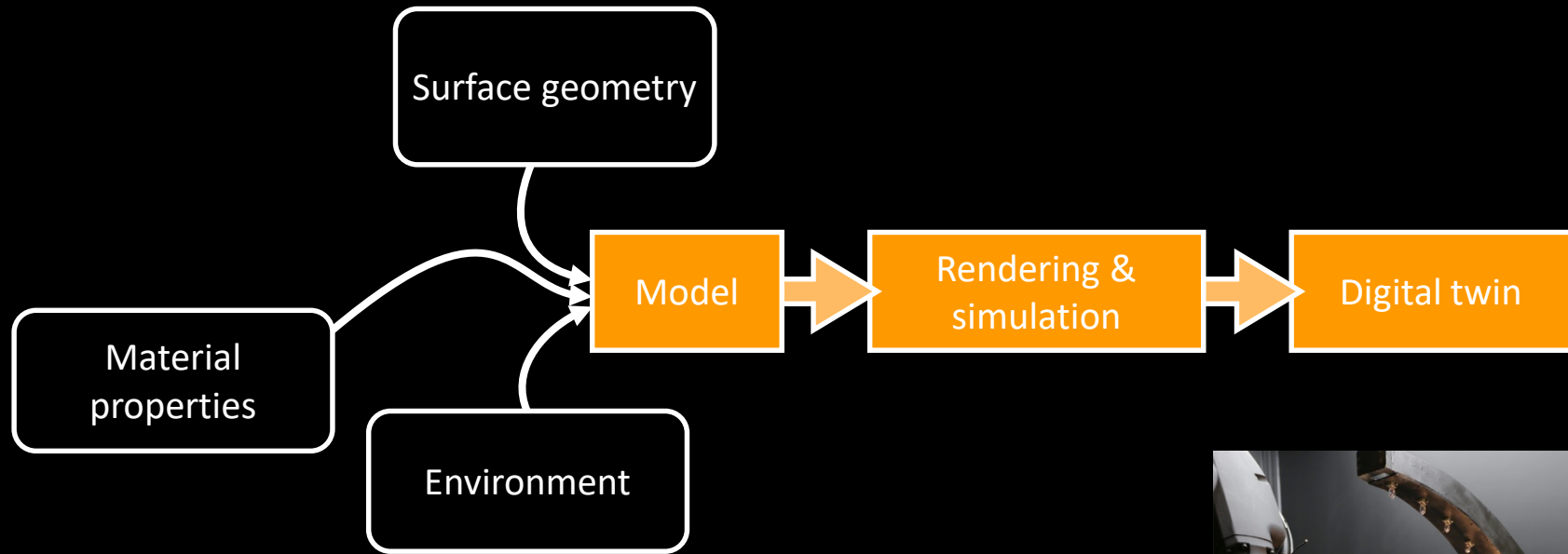




Geometry & Appearance Digitization





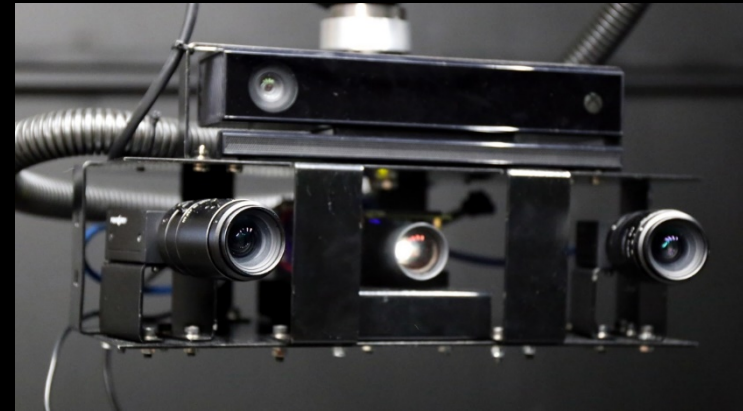


Appearance modelling (scattering theory)

[Frisvad *et al.* 2007; Frisvad 2011; Luongo *et al.* 2017; Frisvad 2018]

Acquiring model input

[Stets *et al.* 2017; Lyngby *et al.* 2019]



(robotic arm, illumination control, structured light scanner)

Photorealistic rendering (Monte Carlo ray tracing)

[Frisvad *et al.* 2012; Frisvad *et al.* 2014; Andersen *et al.* 2016; Dal Corso *et al.* 2016; Stets *et al.* 2017]

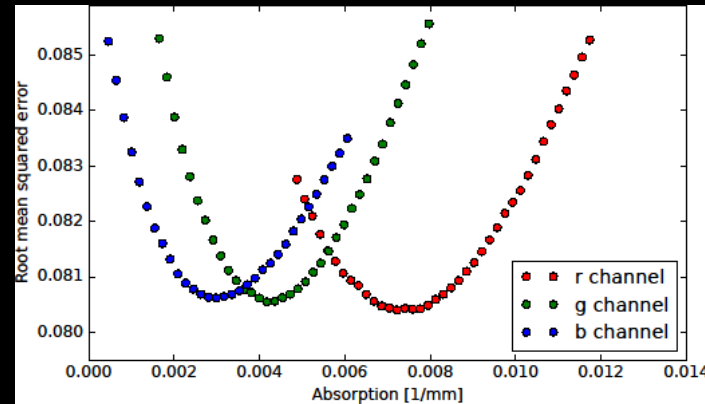
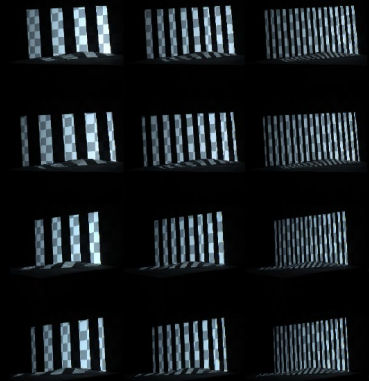


Image Section at Technical University of Denmark



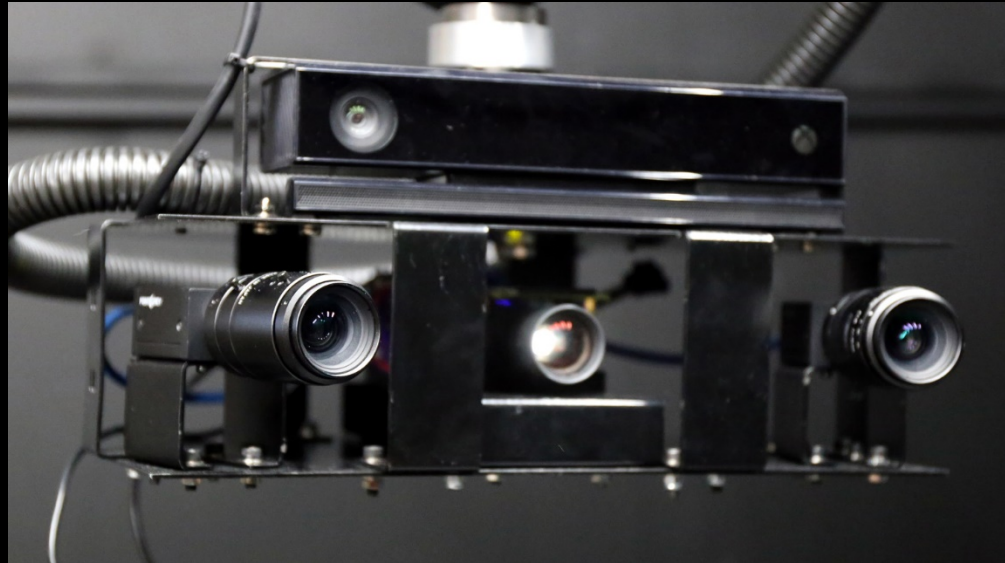
Acquiring input for appearance models

- Geometry from structured light and CT scans [Eiriksson *et al.* 2016; Stets *et al.* 2017]
- Reflectance properties from sparse sampling and PCA reconstruction [Nielsen *et al.* 2015; Xu *et al.* 2016; Stets *et al.* 2017; Nielsen *et al.* 2017; Lyngby *et al.* 2019]
- Scattering properties from directional subsurface scattering [Abildgaard *et al.* 2015; Abildgaard *et al.* 2016]
- Material properties from analysis-by-synthesis [Dal Corso *et al.* 2016; Stets *et al.* 2017]
- Environment mapping [Stets *et al.* 2017]



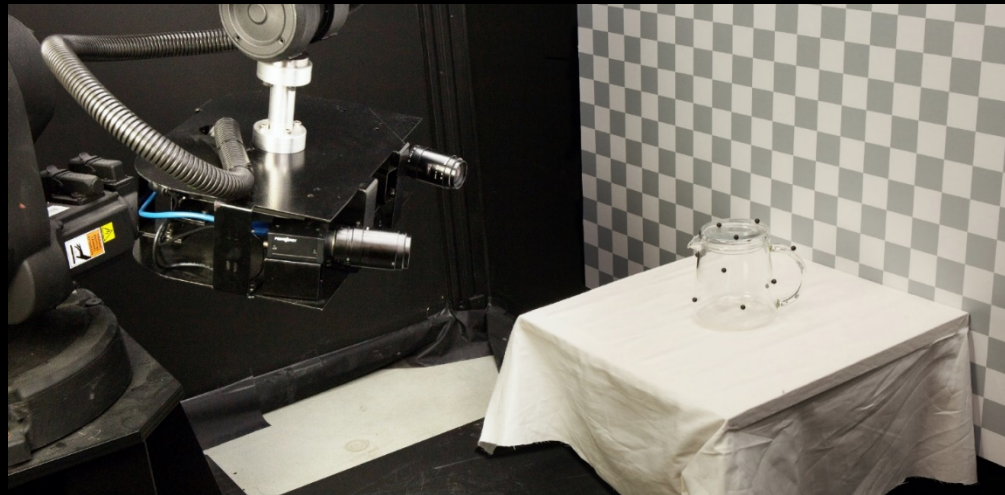
White checker Grey checker White cloth





Instruments

- Robot repeatability
- Illumination control
- Stereo camera setup
- Structured light scanner



Scene

- Checkered backdrop
- Table with cloth
- Glass object



Sphere



Bowl

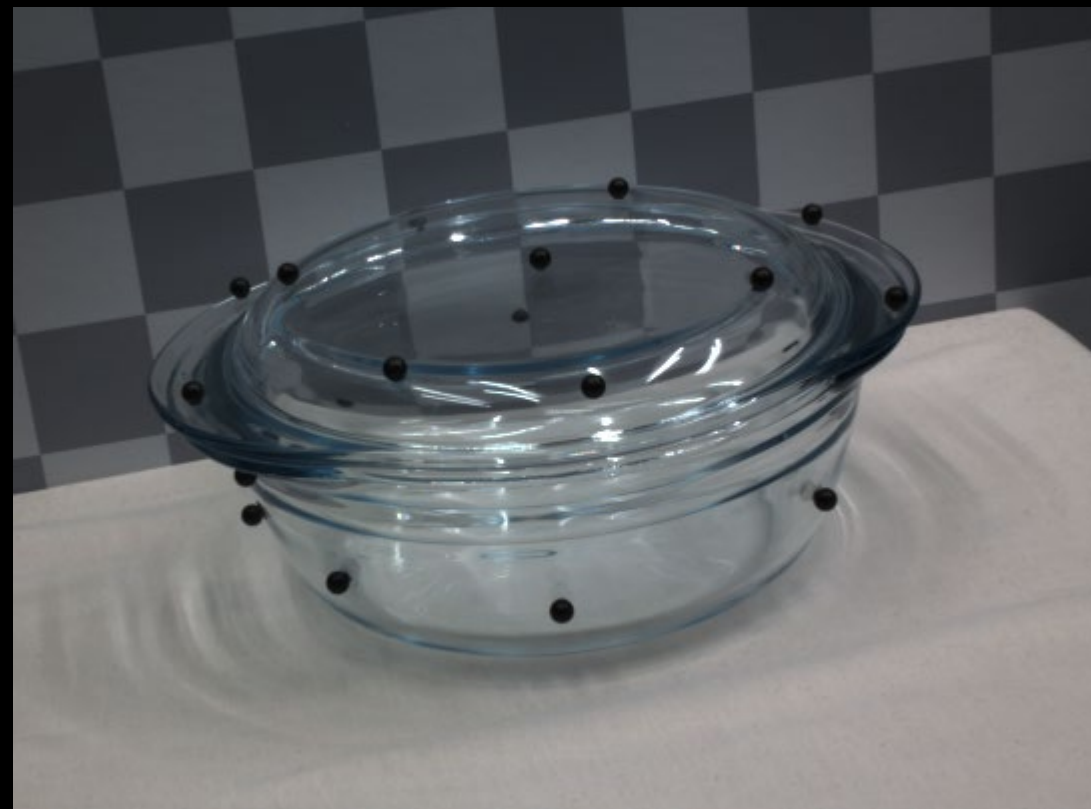


Teapot

Analysis by Synthesis

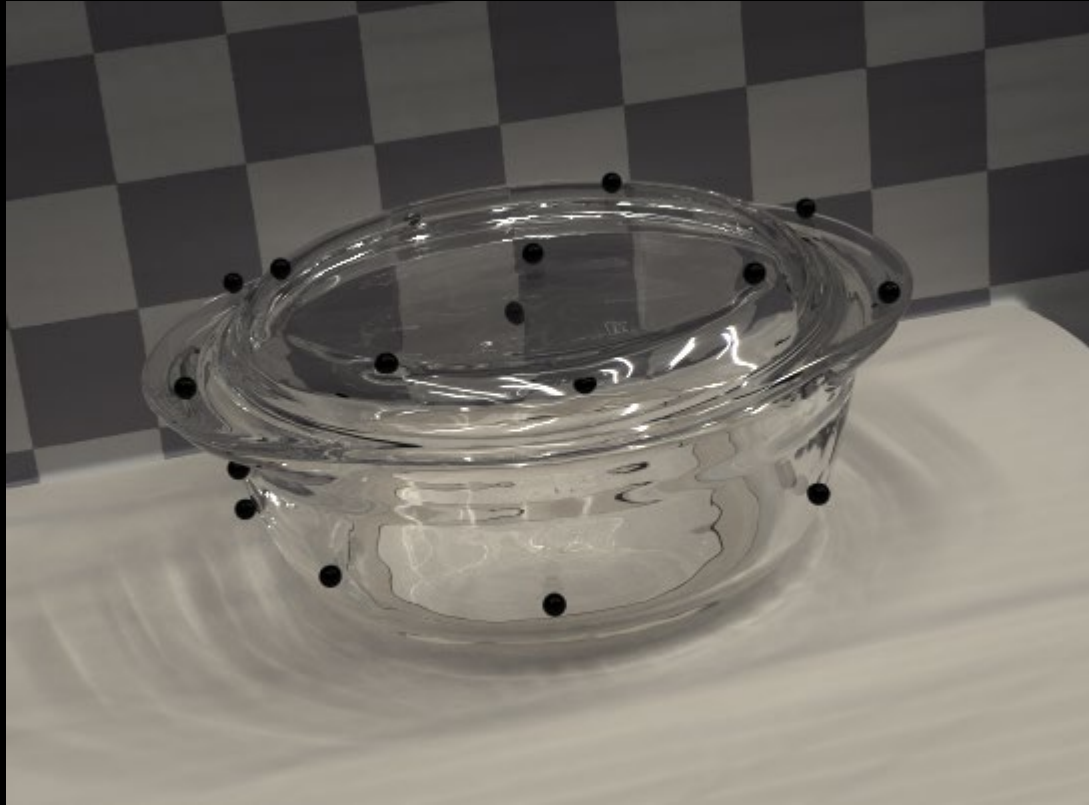


Rendered Image

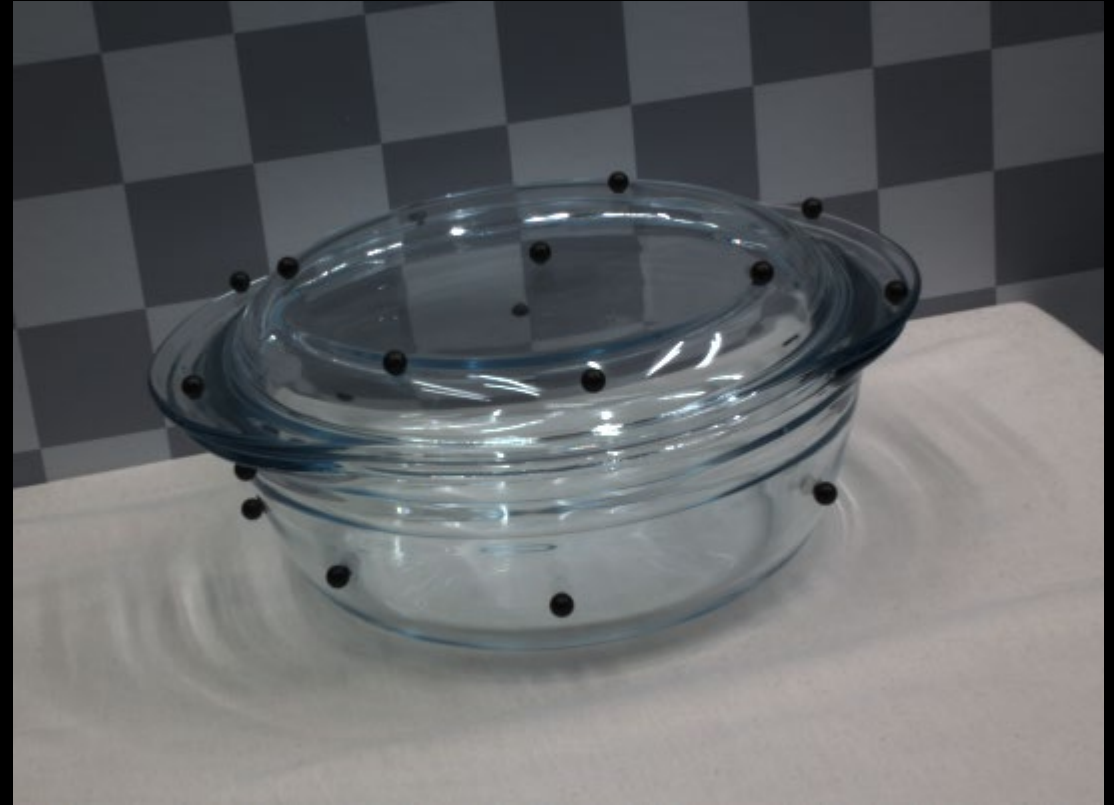


Reference Image

Analysis by Synthesis

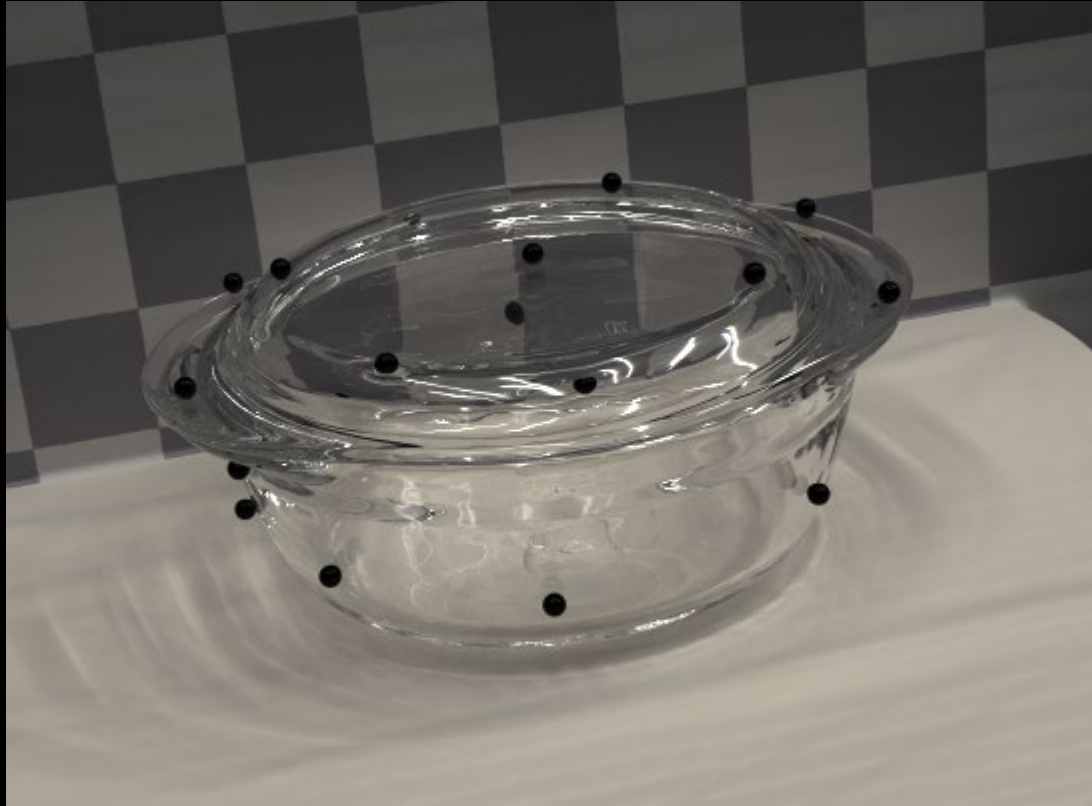


Rendered Image

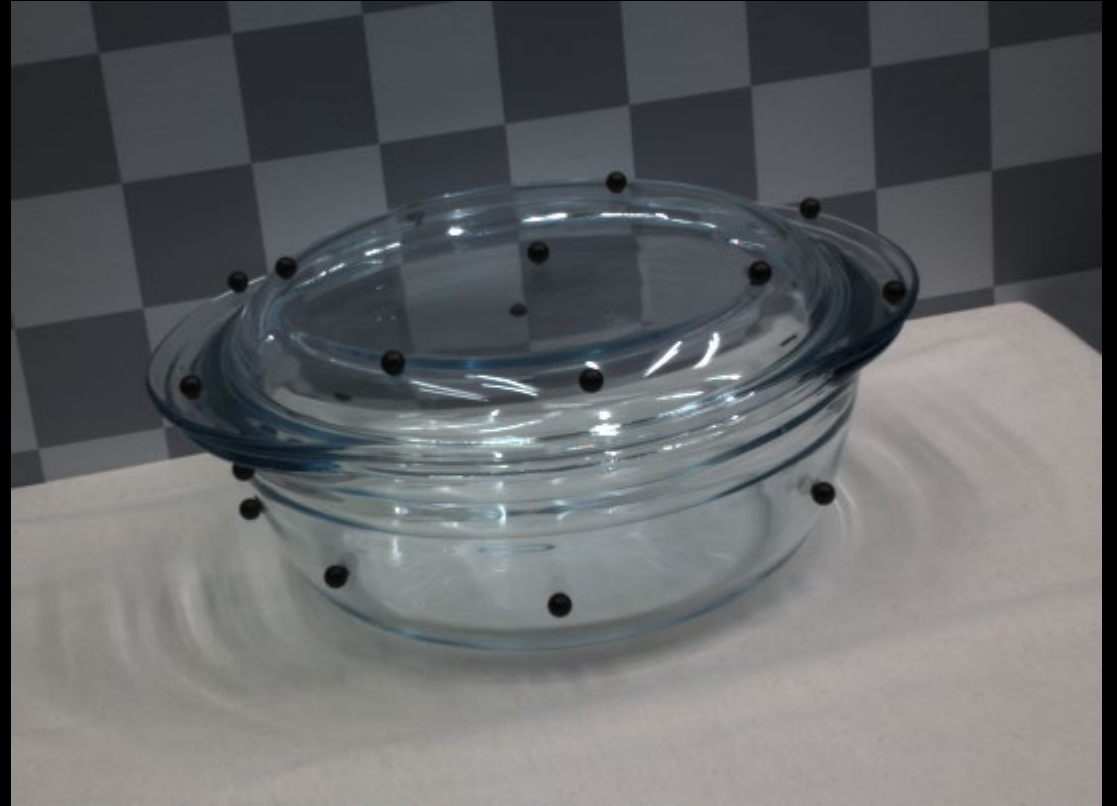


Reference Image

Analysis by Synthesis



Rendered Image

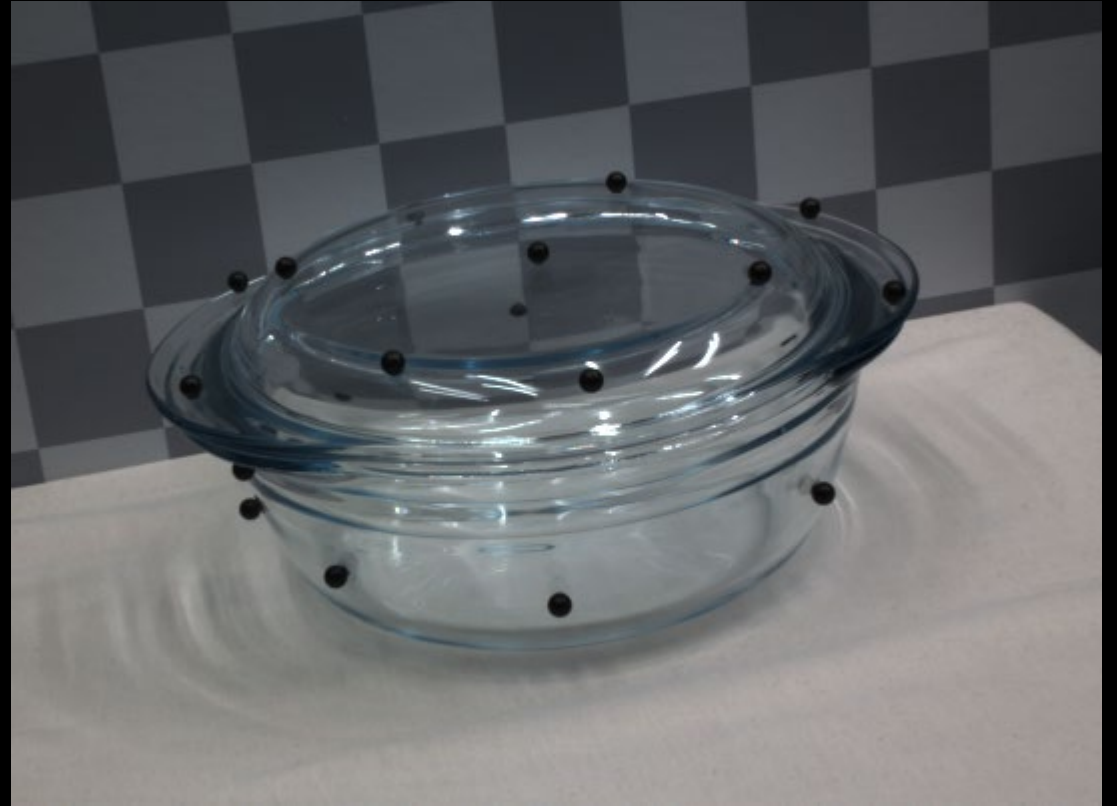


Reference Image

Analysis by Synthesis



Rendered Image

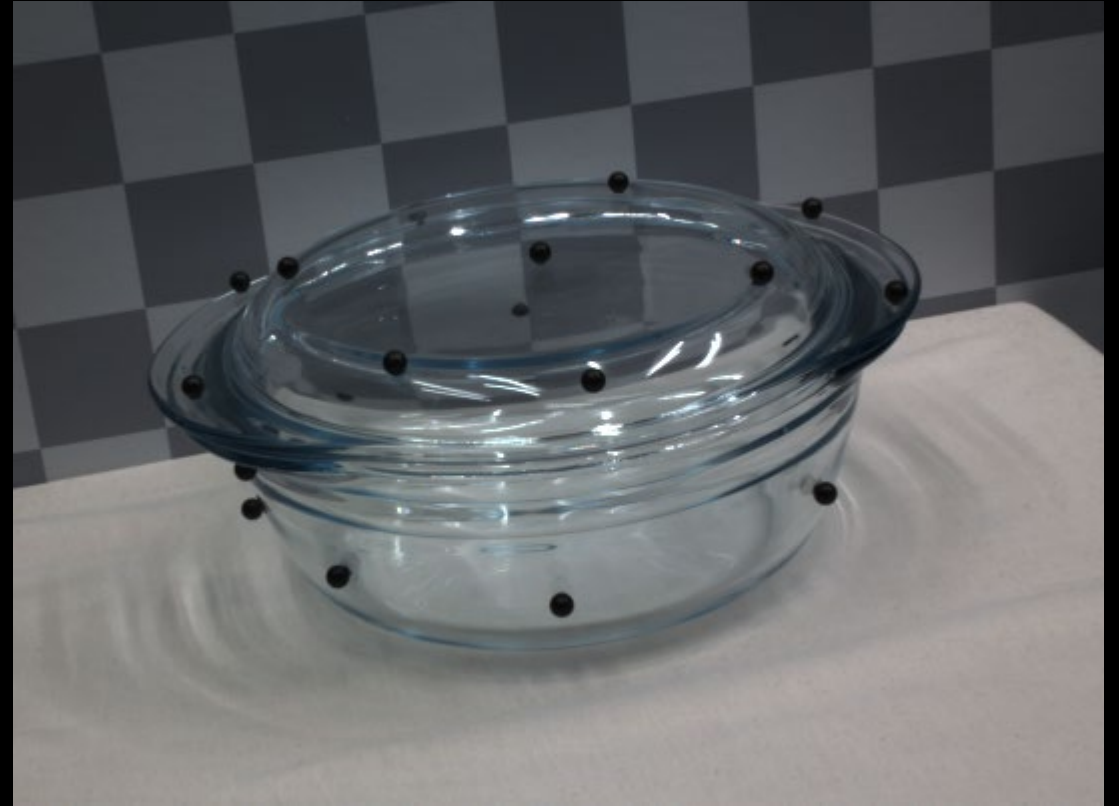


Reference Image

Analysis by Synthesis

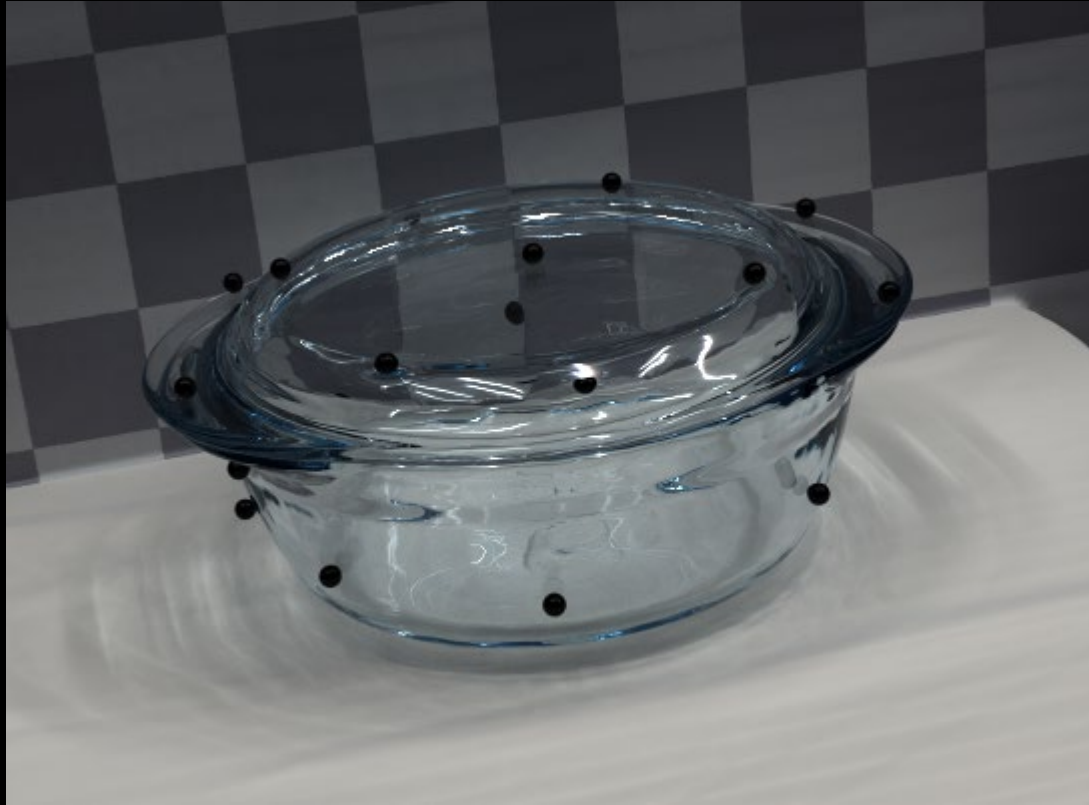


Rendered Image

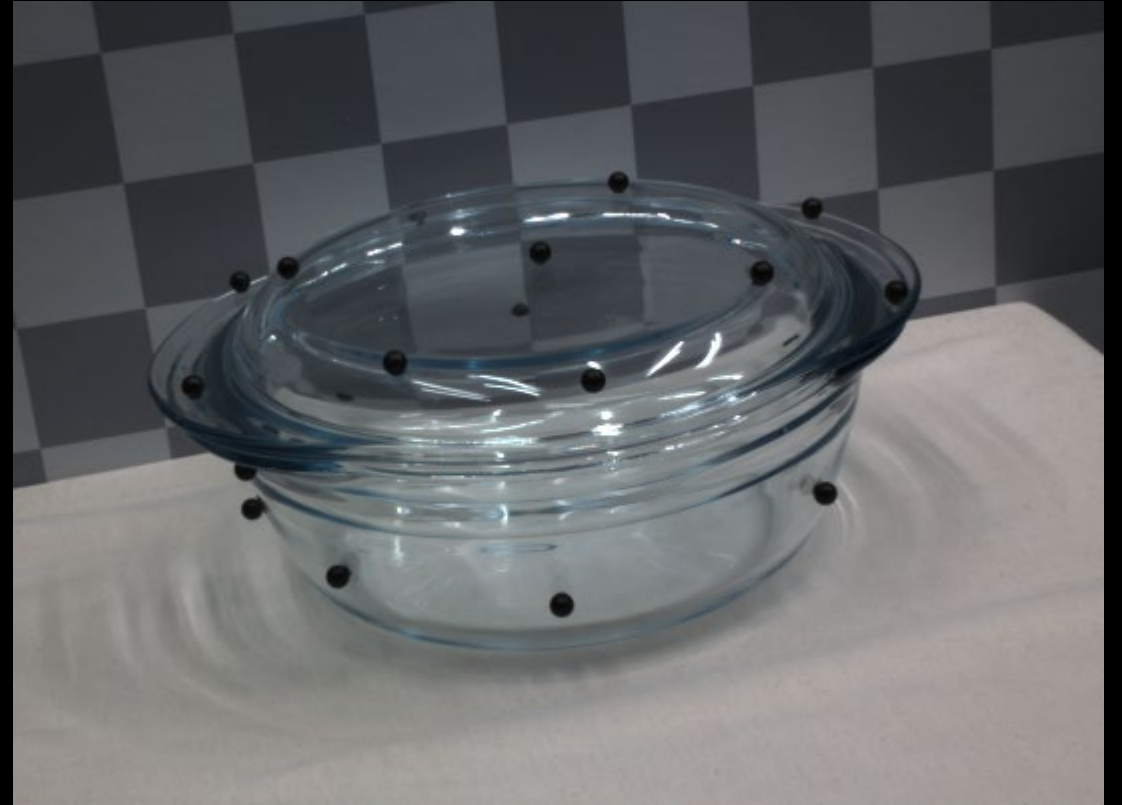


Reference Image

Analysis by Synthesis



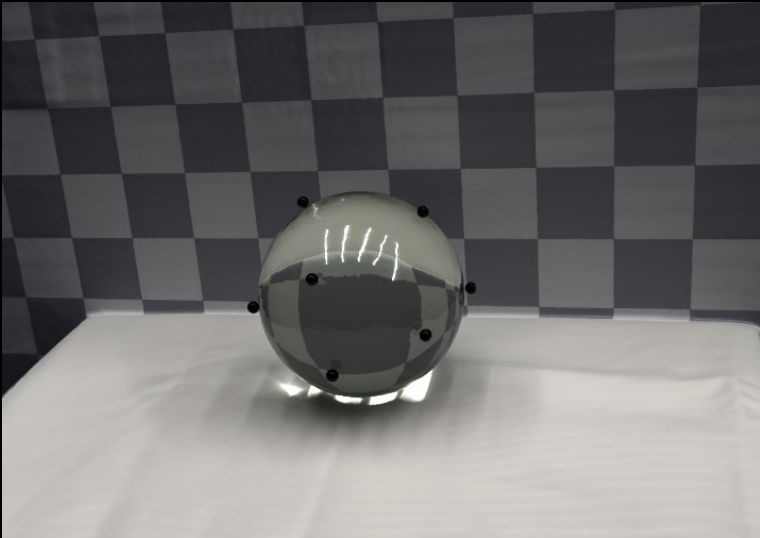
Rendered Image



Reference Image



Rendering



Rendering



Rendering

Geometry & Appearance Digitization



Geometry & Appearance Digitization

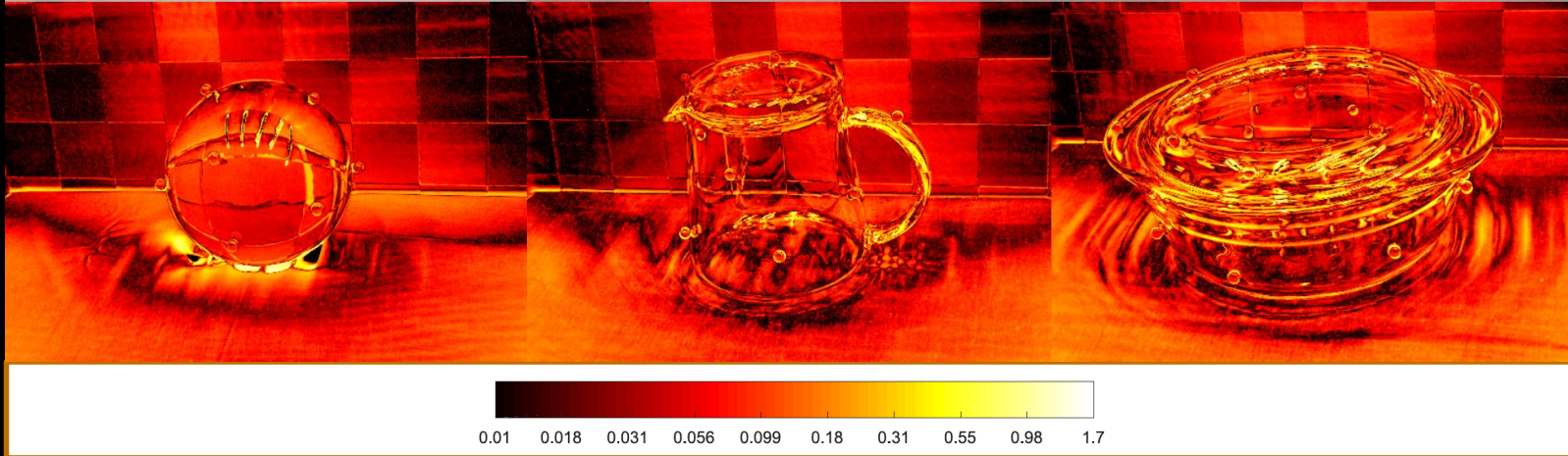
Renderings



Photos

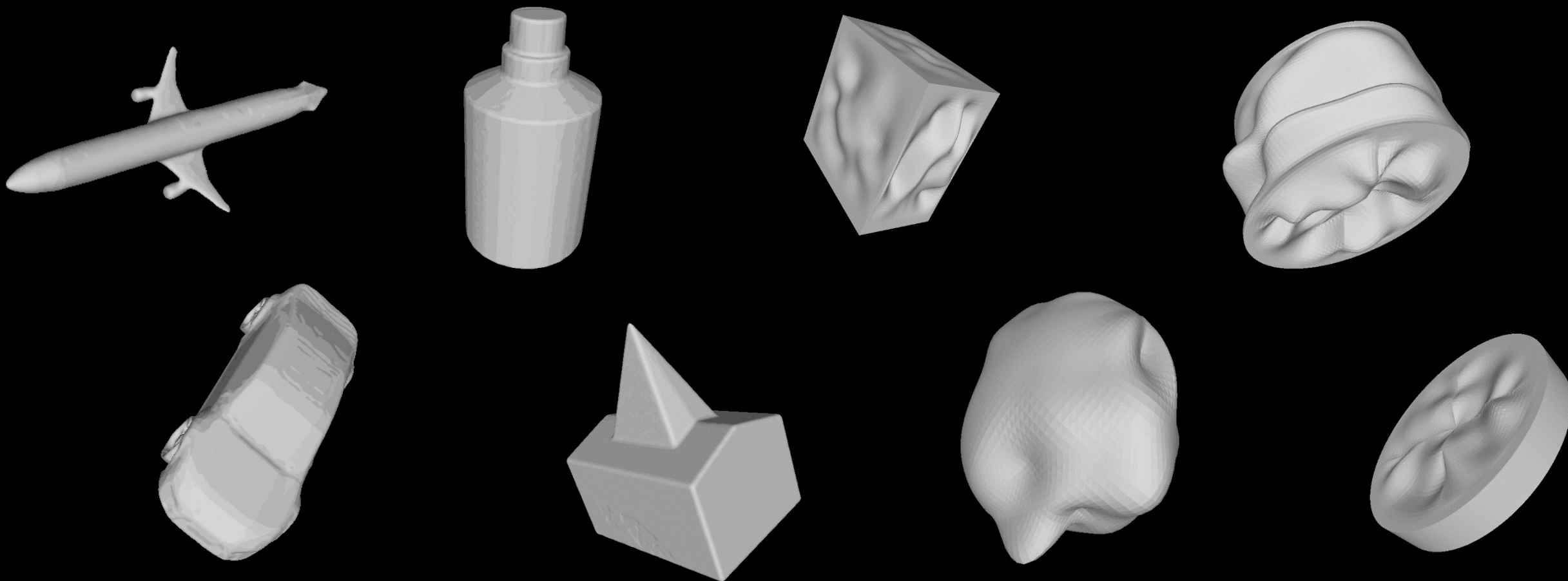


Pixelwise error
Log of L2 norm





Scene reassembly after multimodal digitization and pipeline evaluation using photorealistic rendering



300 Models
from ShapeNet

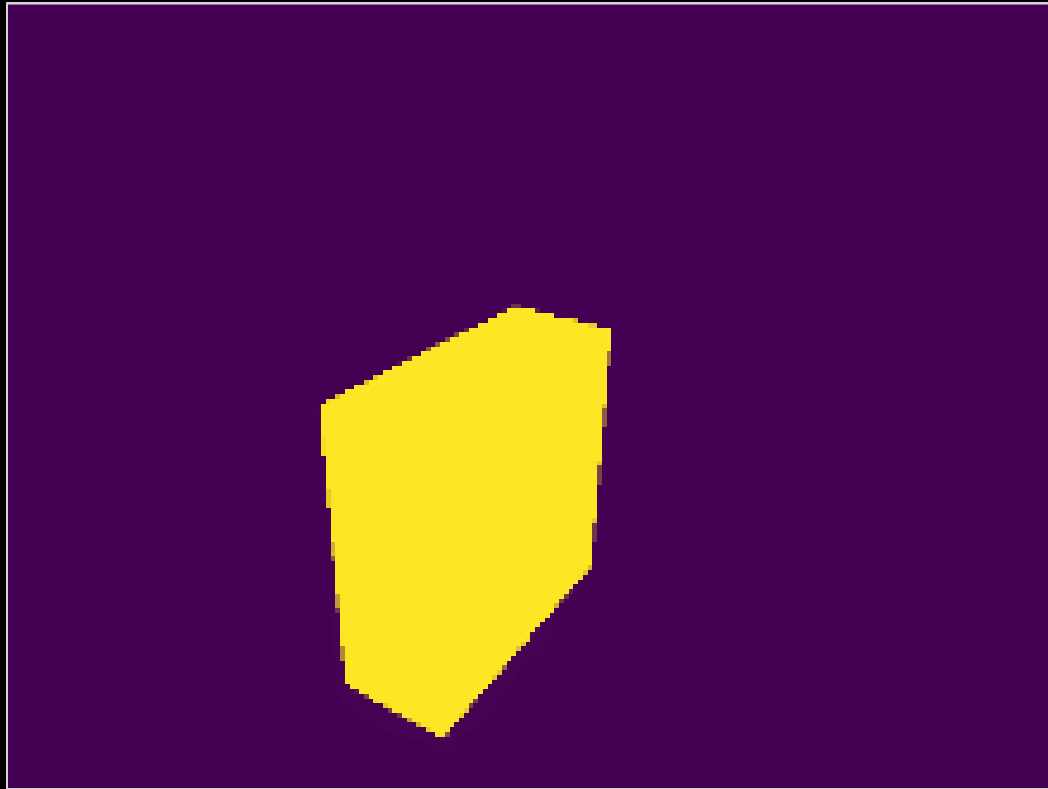
300 Models
of generated shapes



60 HDR Environment Maps

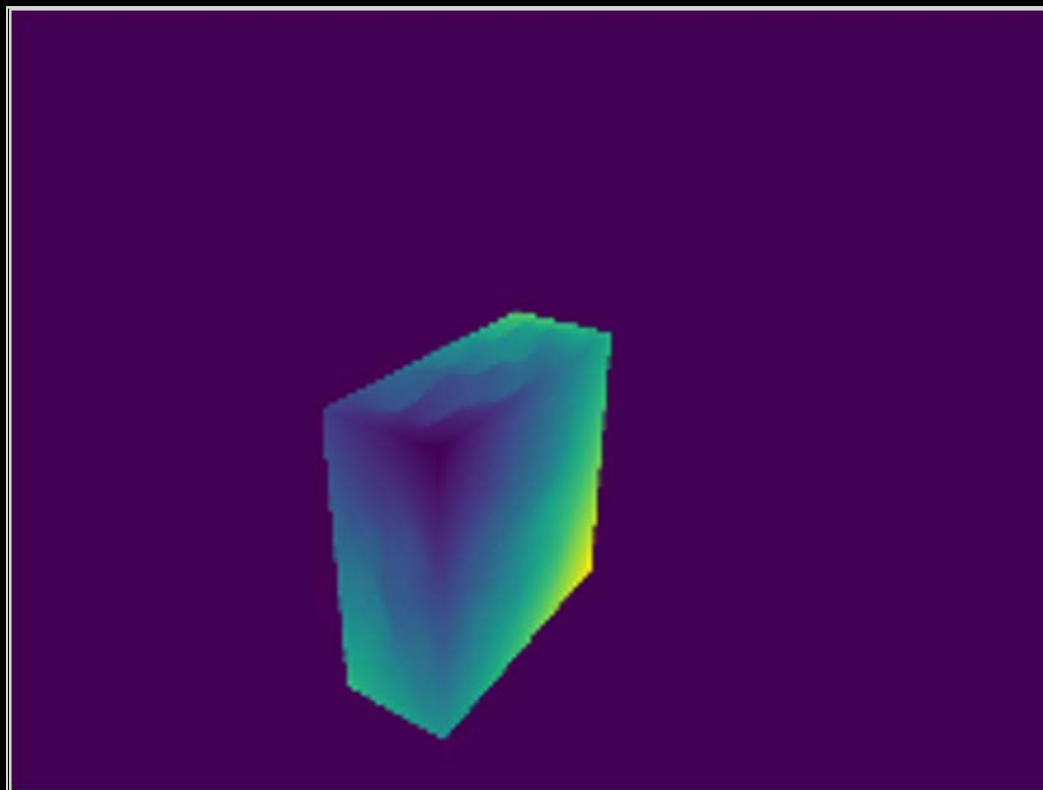


RGB-image



RGB-image

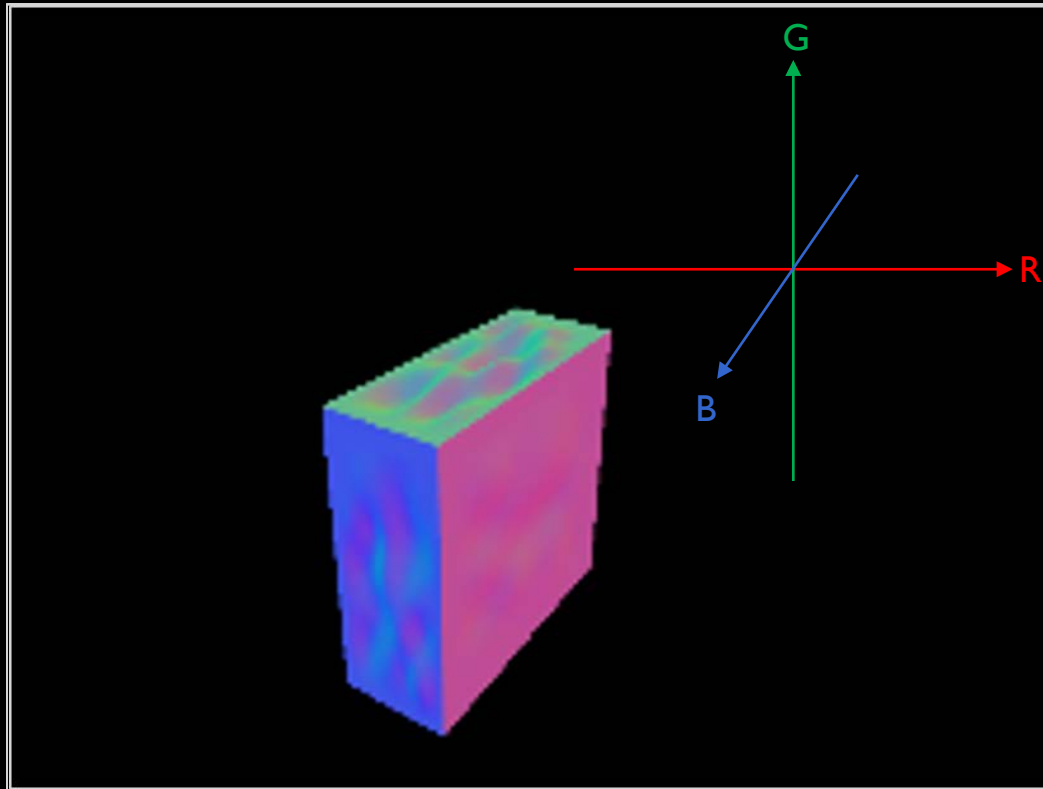
Mask



RGB-image

Mask

Depth

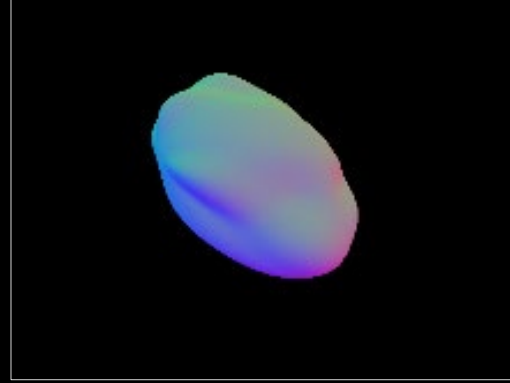
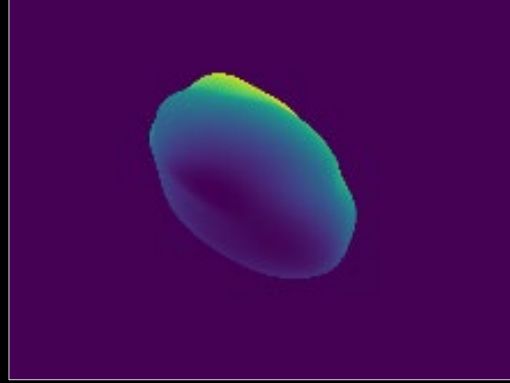
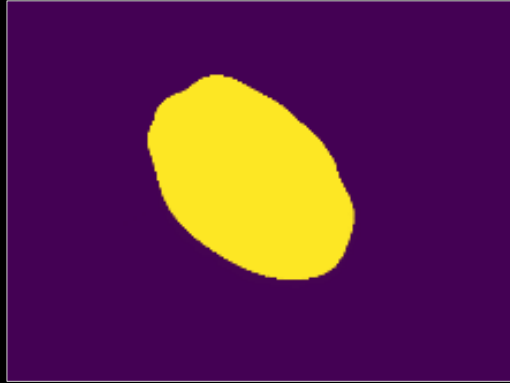
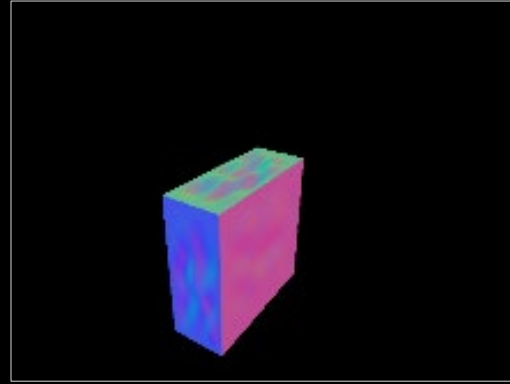
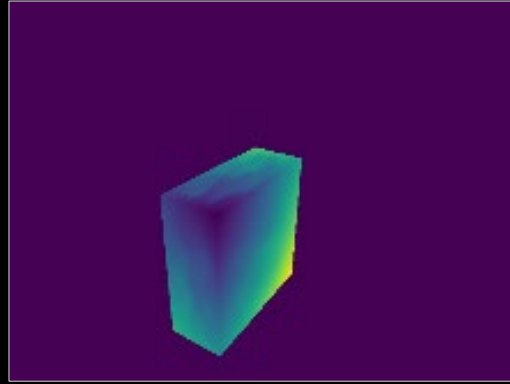
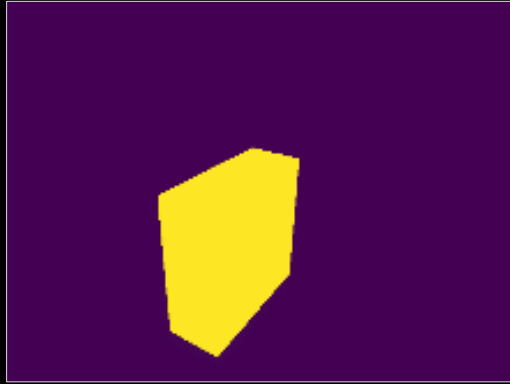
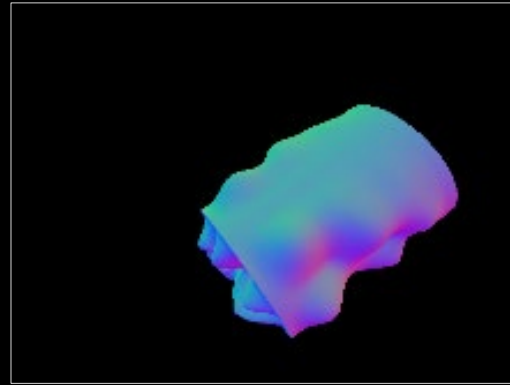
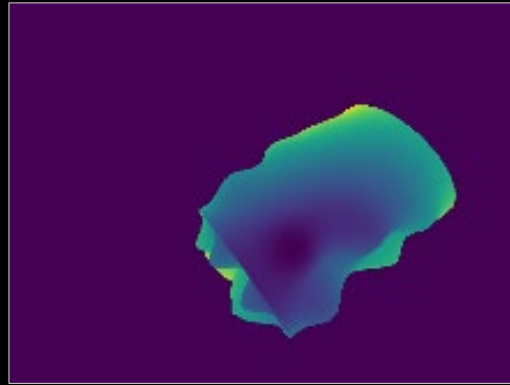
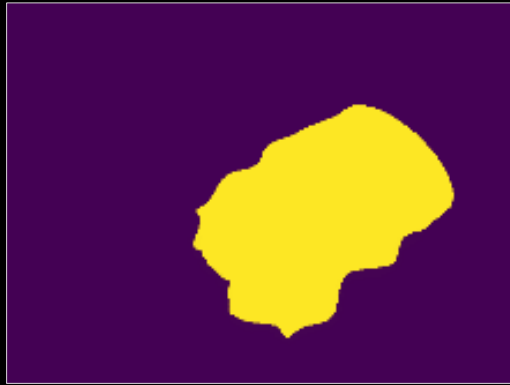


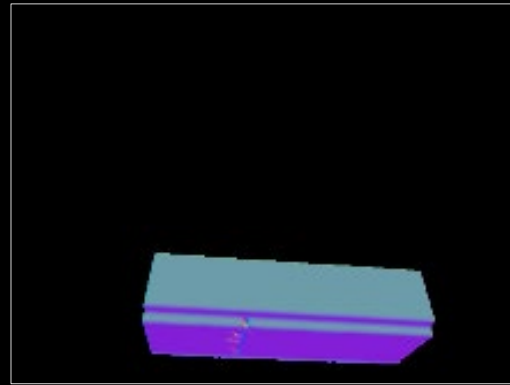
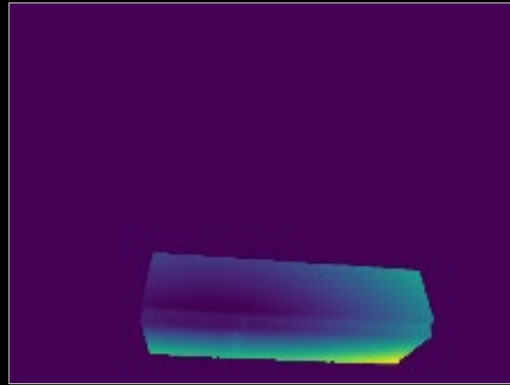
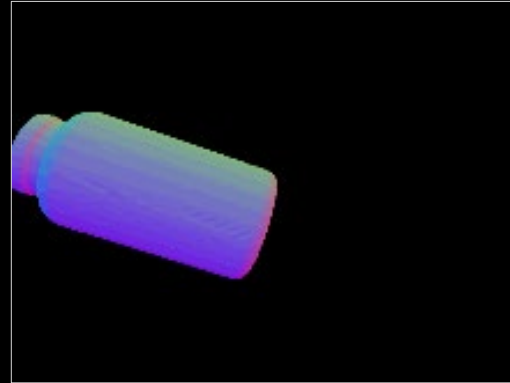
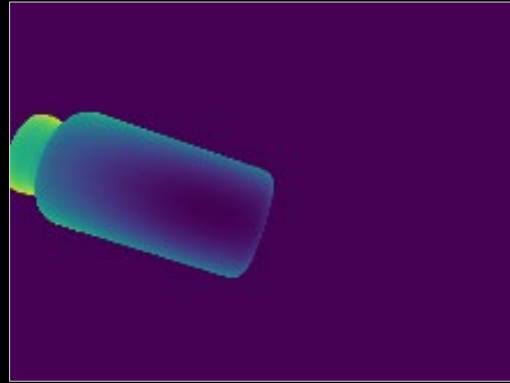
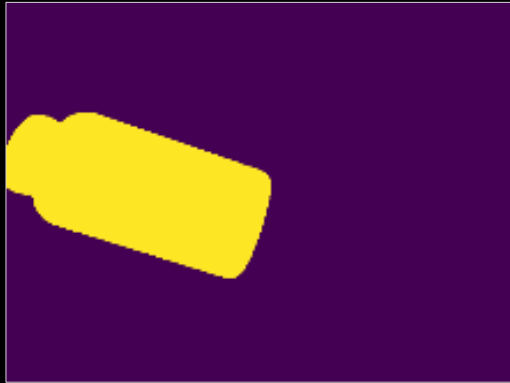
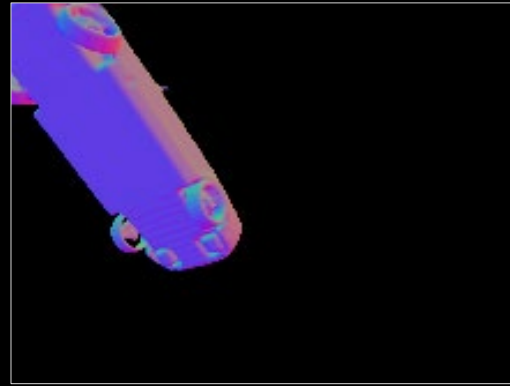
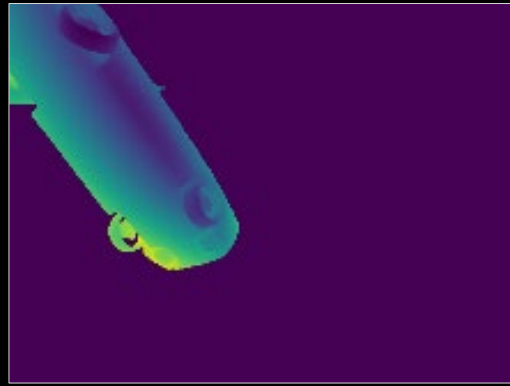
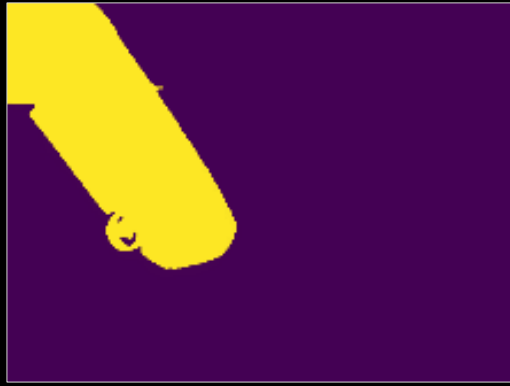
RGB-image

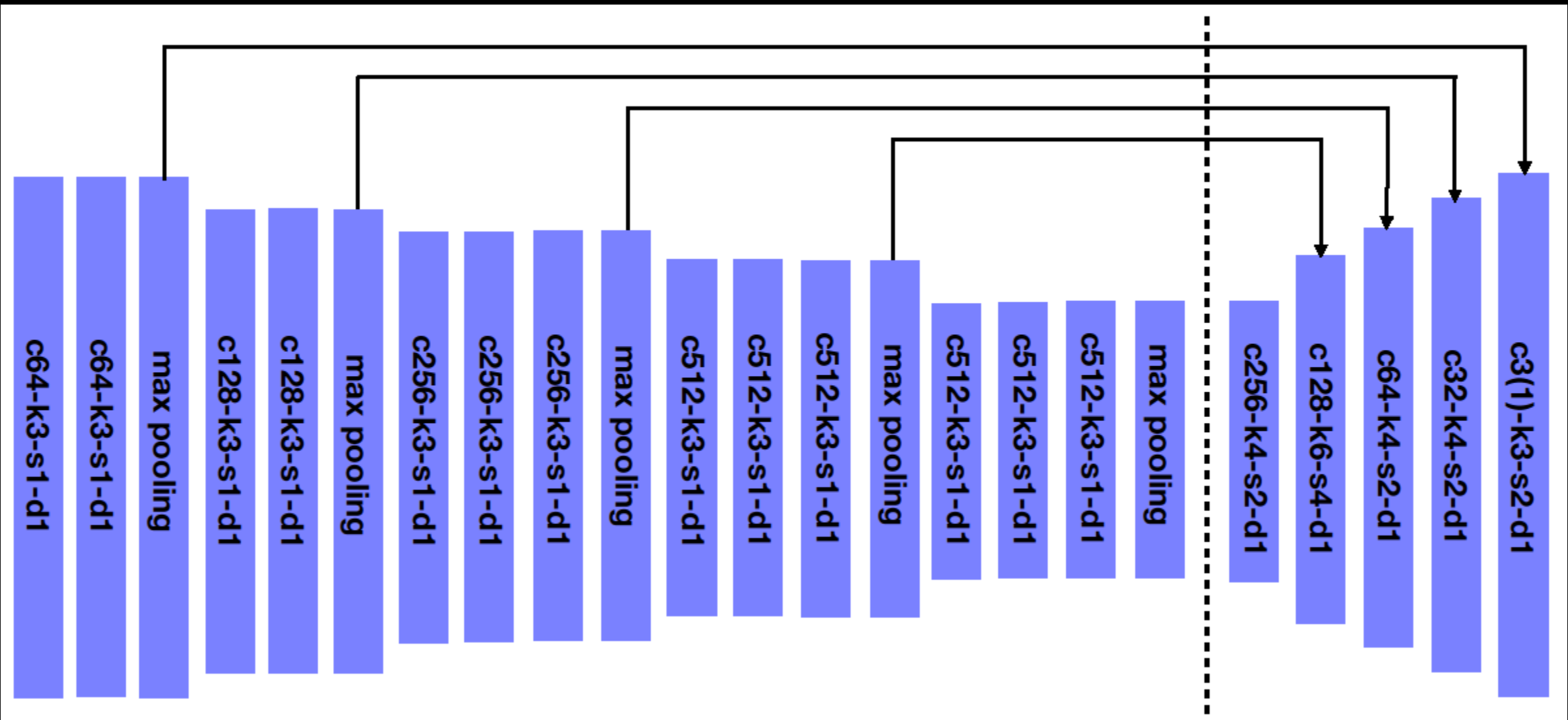
Mask

Depth

Normals







ENCODER

DECODER

Loss Functions

$$\frac{1}{|P|} \sum_{p \in P} (M_p - \hat{M}_p)^2$$

Mask

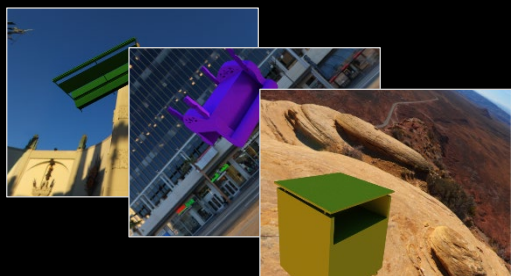
$$\frac{1}{|P_f|} \sum_{p \in P_f} (D_p - \hat{D}_p)^2$$

Depth

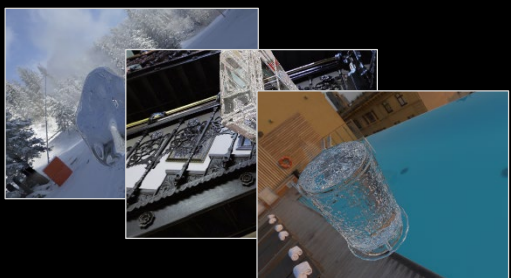
$$\frac{-1}{|P_f|} \sum_{p \in P_f} N_p \cdot \hat{N}_p$$

Normals

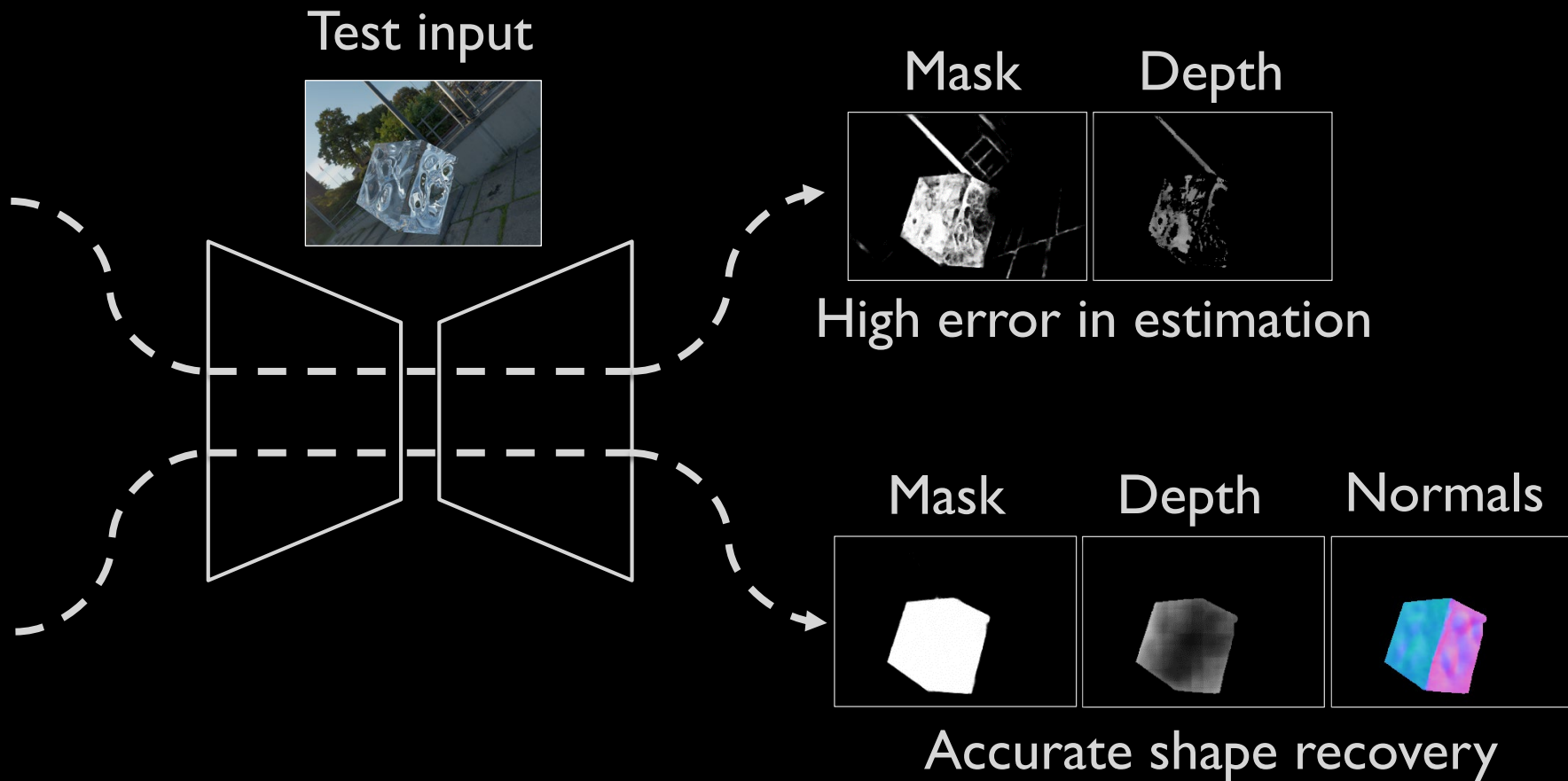
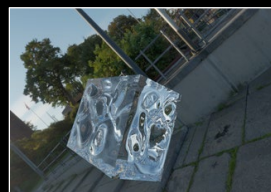
Diffuse dataset



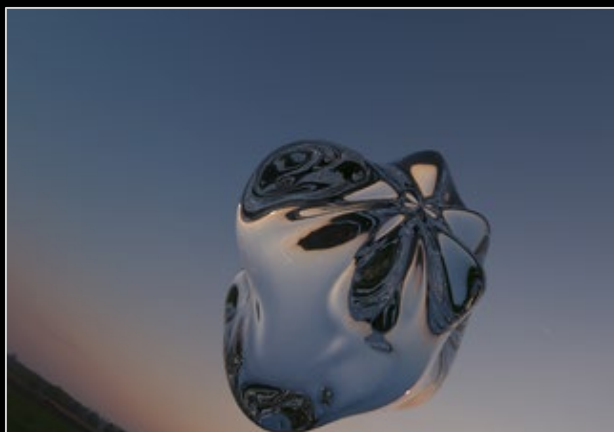
Transparent dataset



Test input



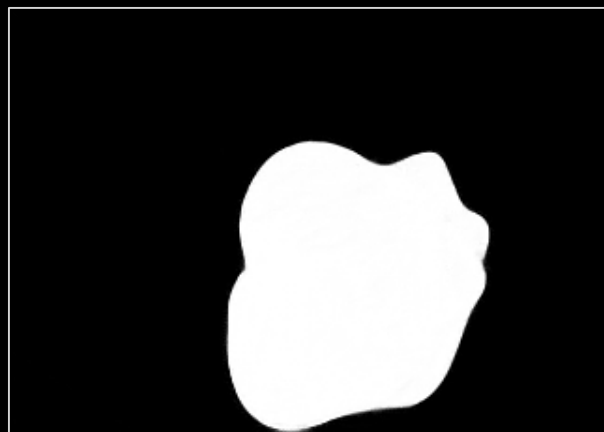
INPUT



GROUND TRUTH



PREDICTED



PREDICTED > 0.9

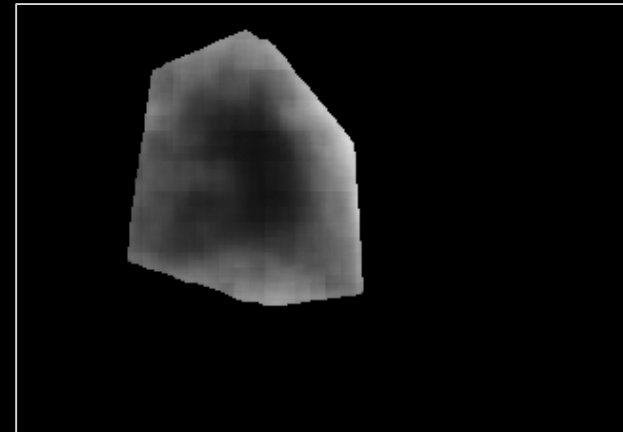
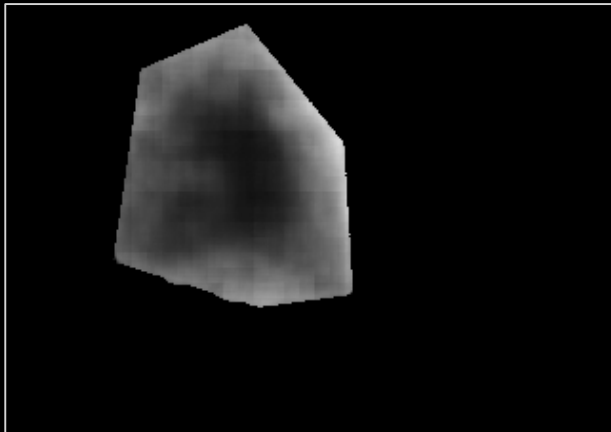
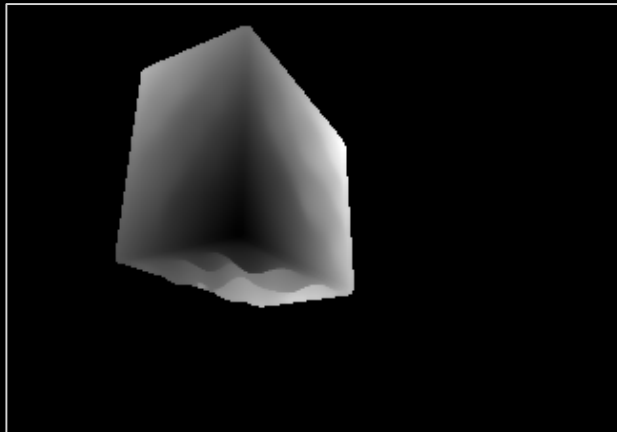


INPUT

GROUND TRUTH

PREDICTED

PREDICTED > 0.9

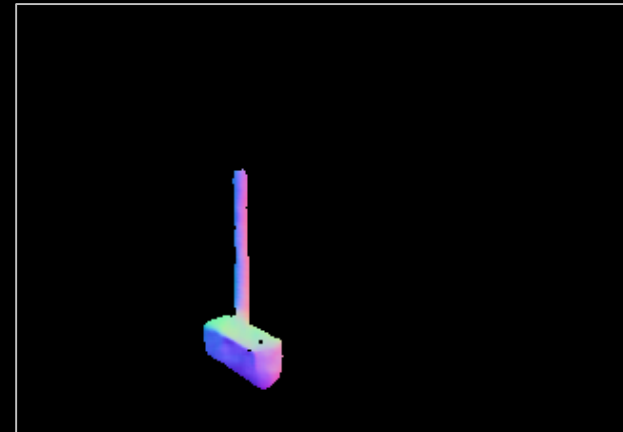
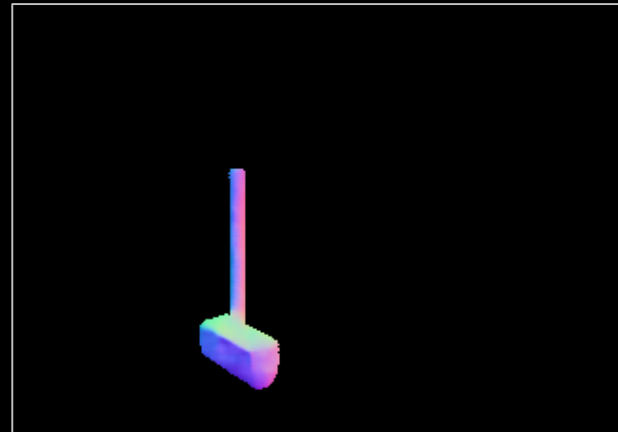
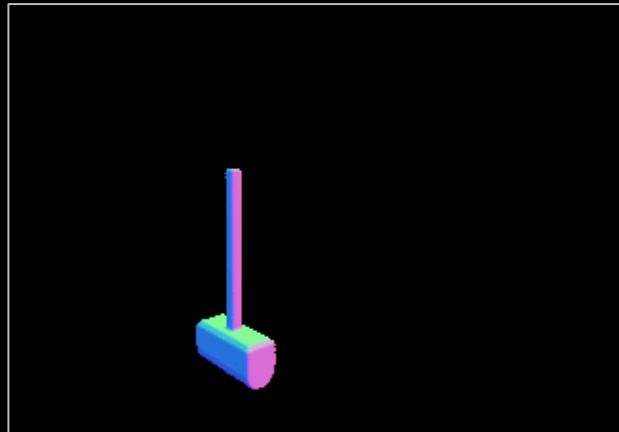


INPUT

GROUND TRUTH

PREDICTED

PREDICTED > 0.9





INPUT

MASK

DEPTH

NORMAL

PREDICTED

GT

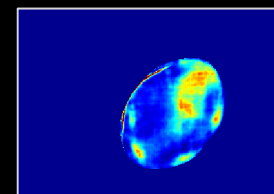
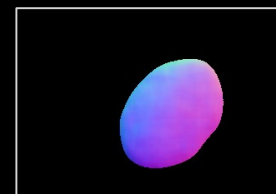
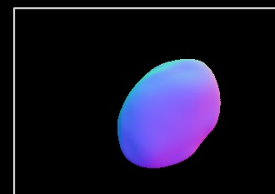
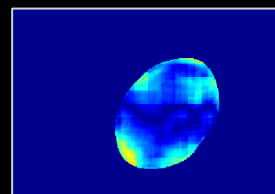
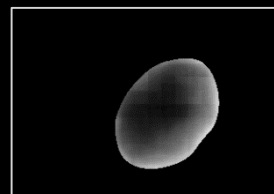
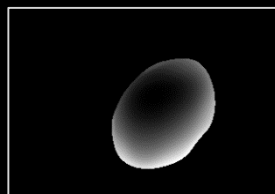
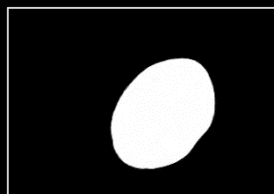
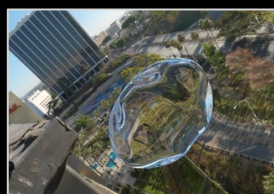
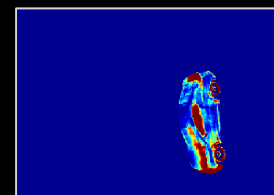
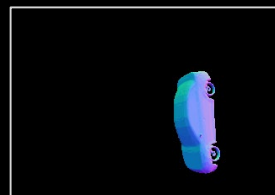
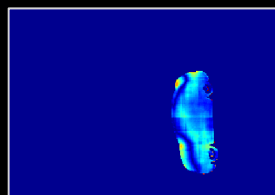
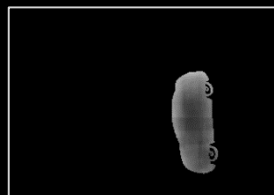
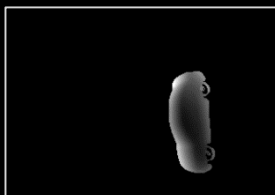
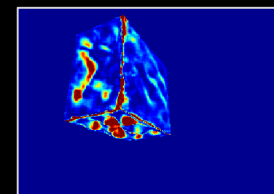
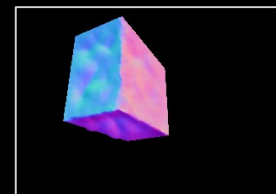
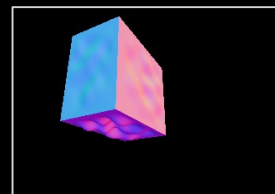
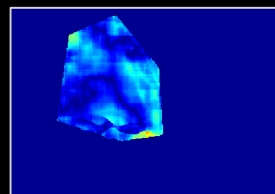
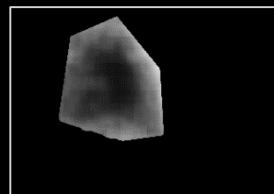
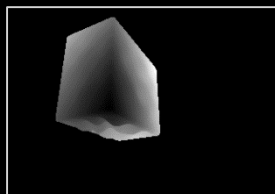
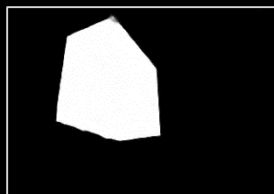
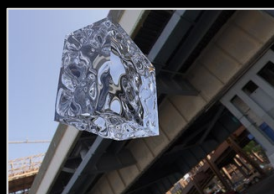
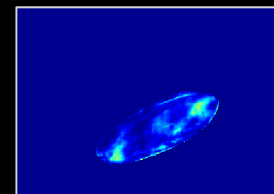
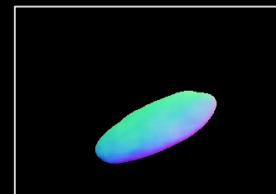
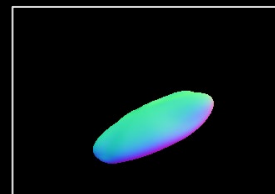
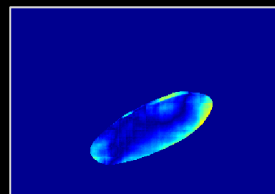
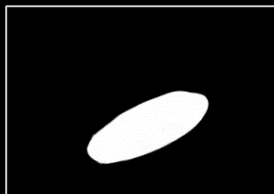
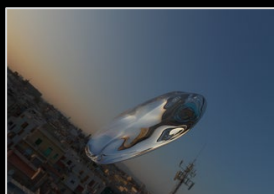
PREDICTED

ERROR

GT

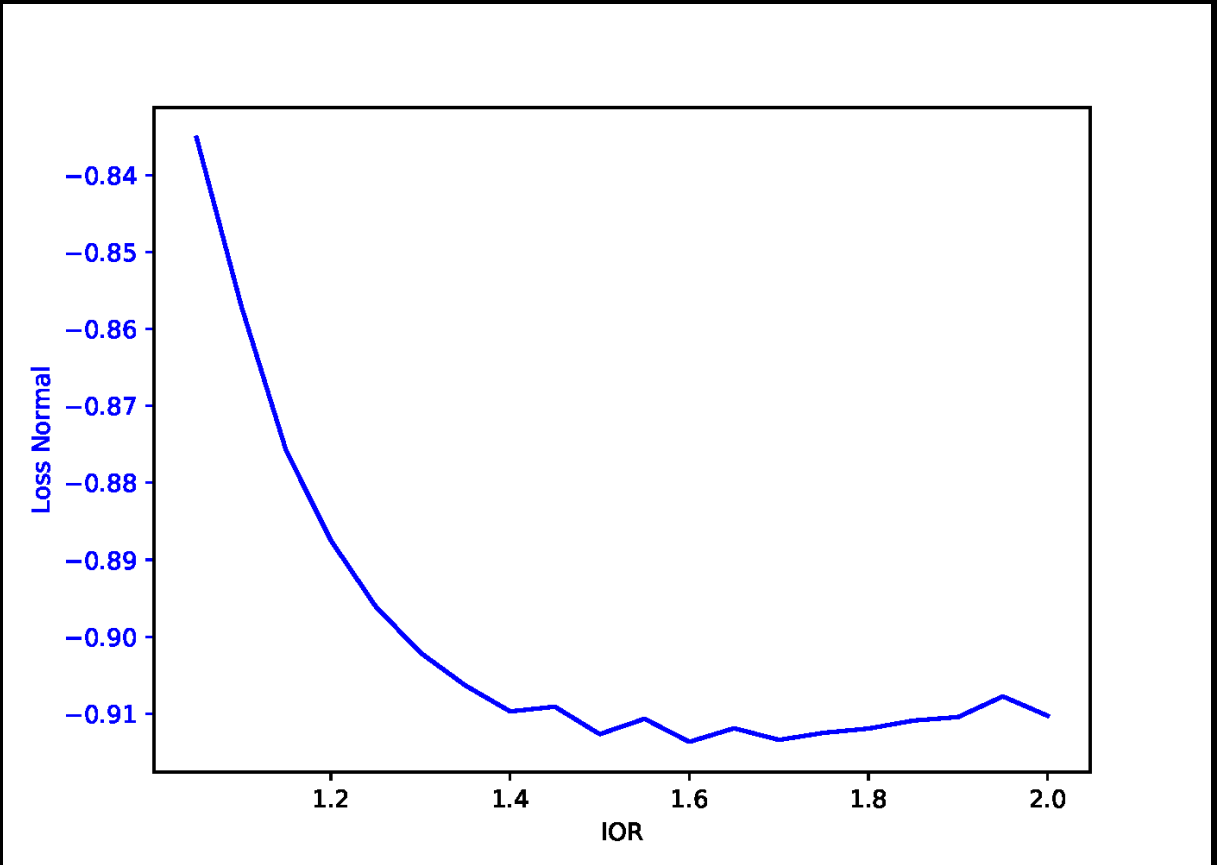
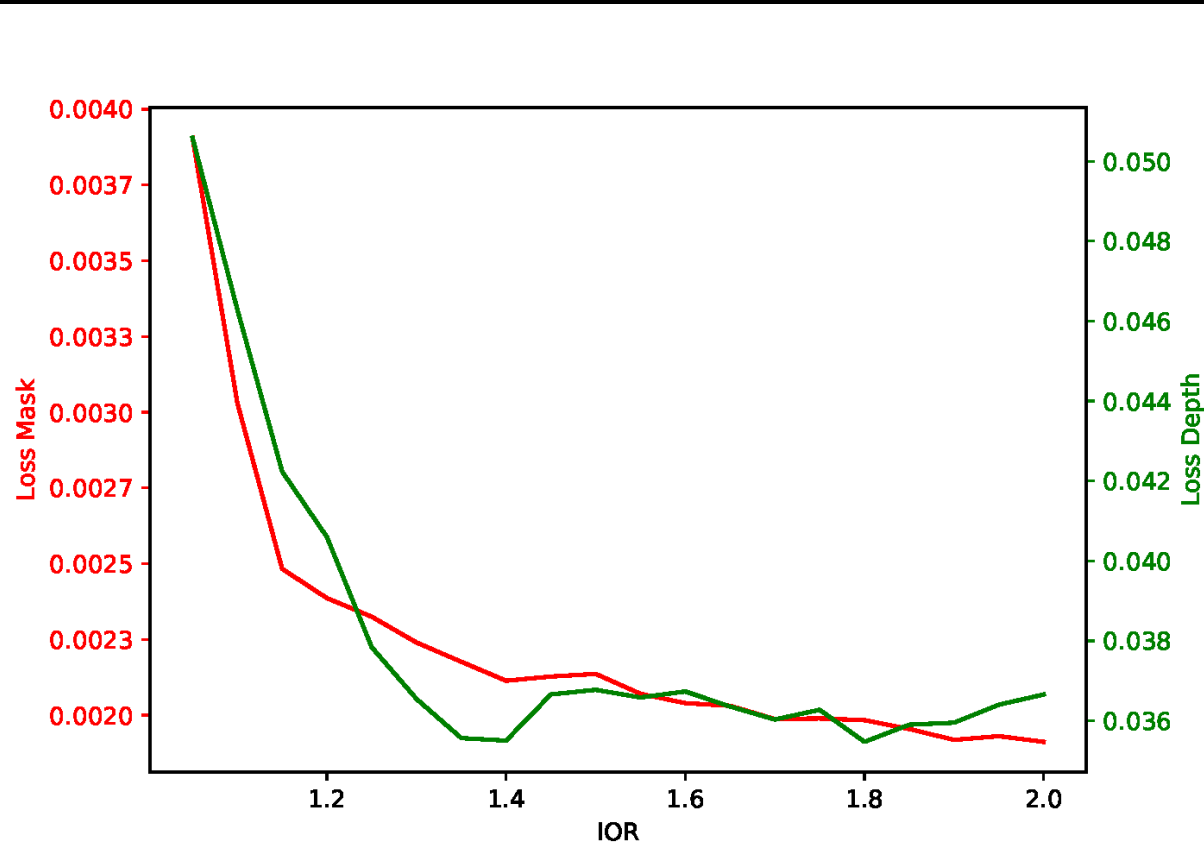
PREDICTED

ERROR





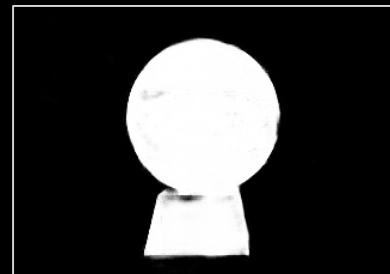
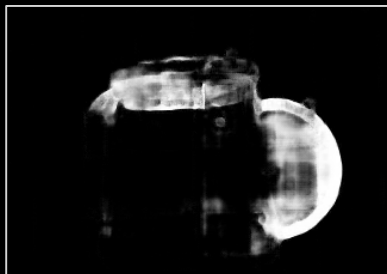
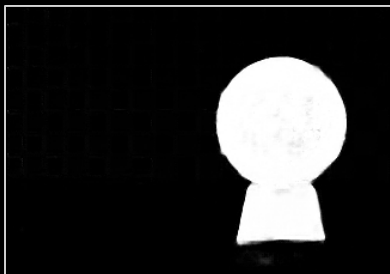
Varying the Index of Refraction



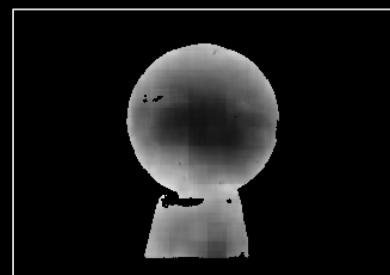
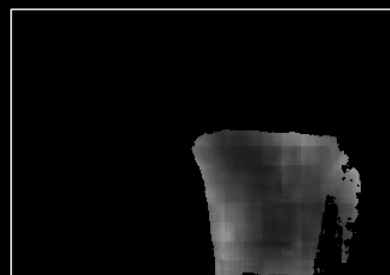
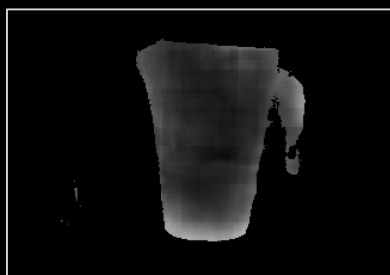
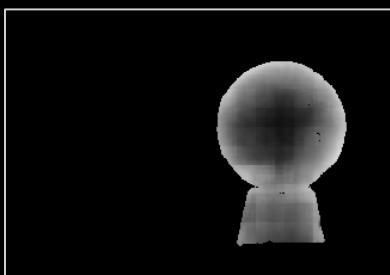
INPUT



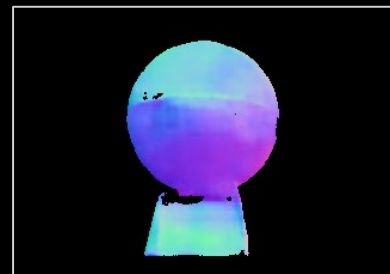
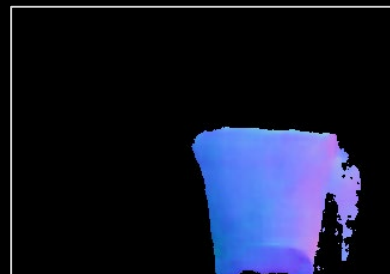
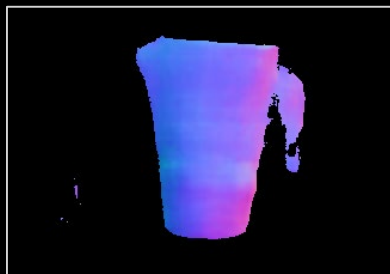
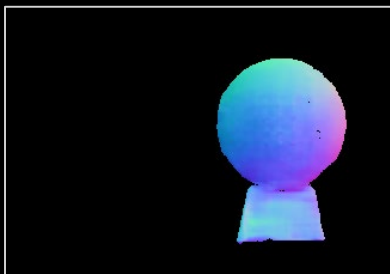
MASK



DEPTH



NORMAL

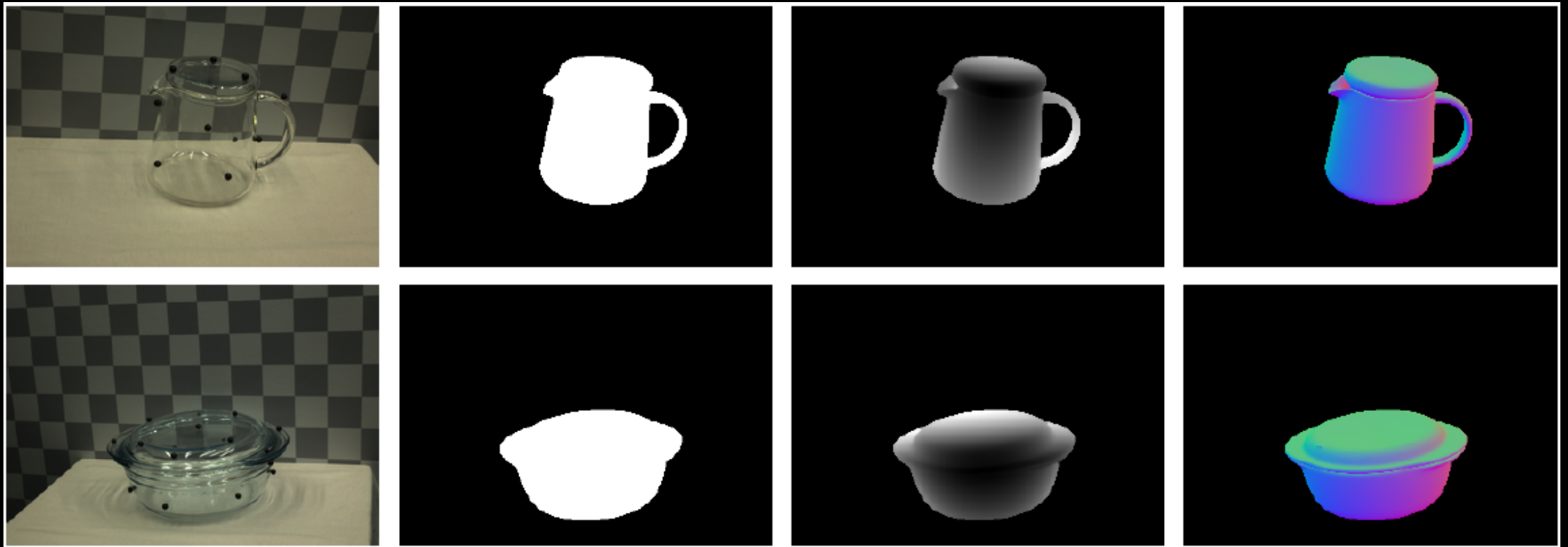




Small dataset with reference data for real images for evaluation



Perhaps useful in future work (three objects: sphere, teapot, bowl, 30 views of each)



Thank you for your attention