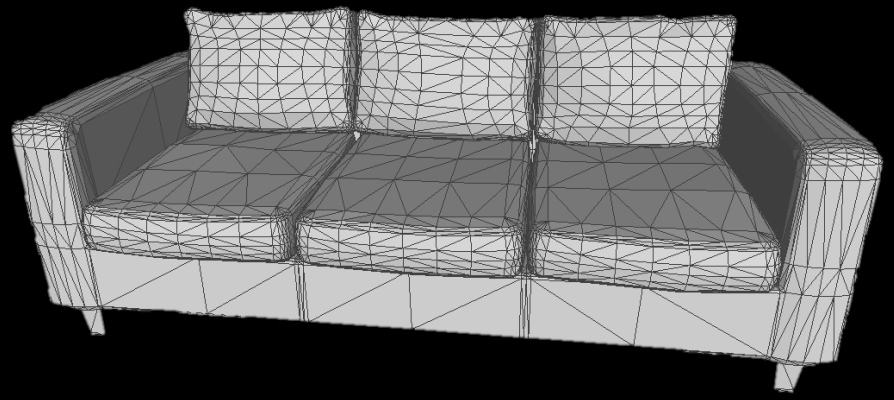
Deep Inverse Rendering for High-resolution SVBRDF Estimation from an Arbitrary Number of Images

Duan Gao [1,3], Xiao Li [2,3], Yue Dong [3], Pieter Peers [4], Kun Xu [1], Xin Tong [3]

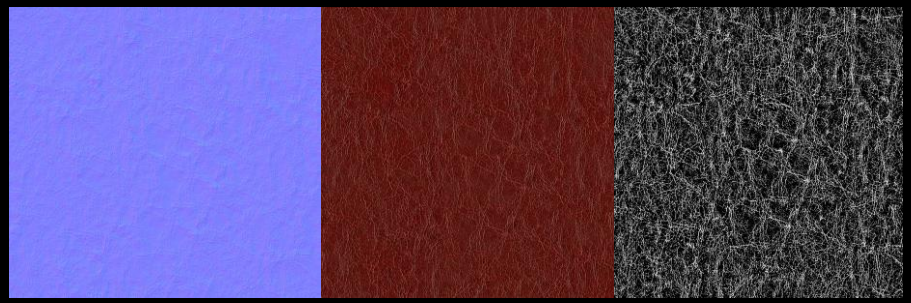
[1] Tsinghua University
[2] University of Science and Technology of China
[3] Microsoft Research Asia
[4] College of William & Mary

Rendering

Materials

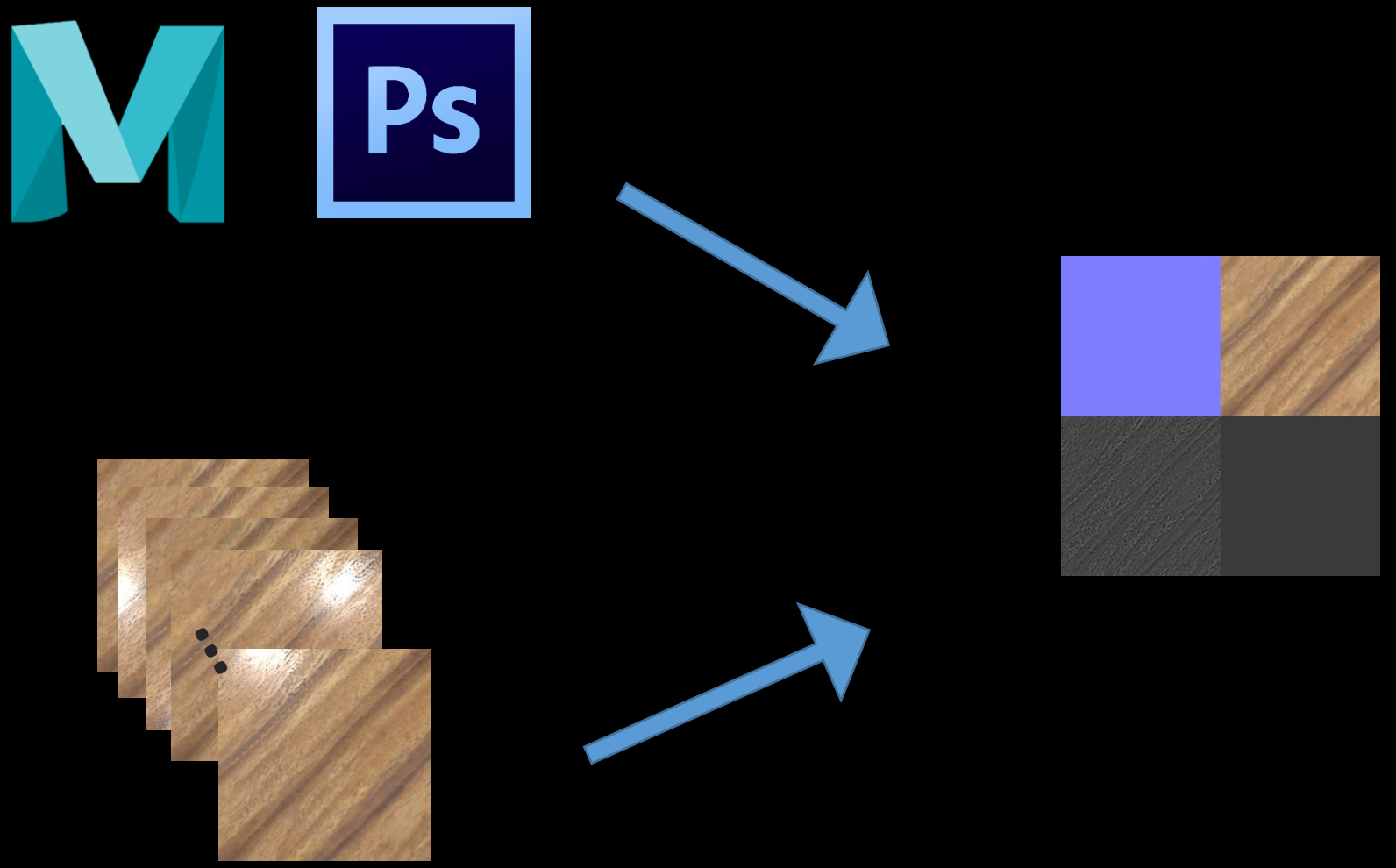


Geometry



Material

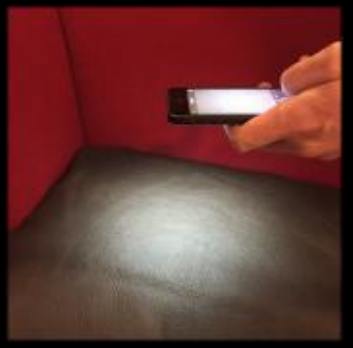
Appearance Estimation



Related Work

- Classic Inverse Rendering
  - Multi-Image Heuristics-based Appearance Modeling.
  - Single/Few Image Reflectance Modeling
- Learning-based Appearance Modeling

[Dong et al. 2014]



[Aittala et al. 2015]

[Deschaintre et al. 2018]

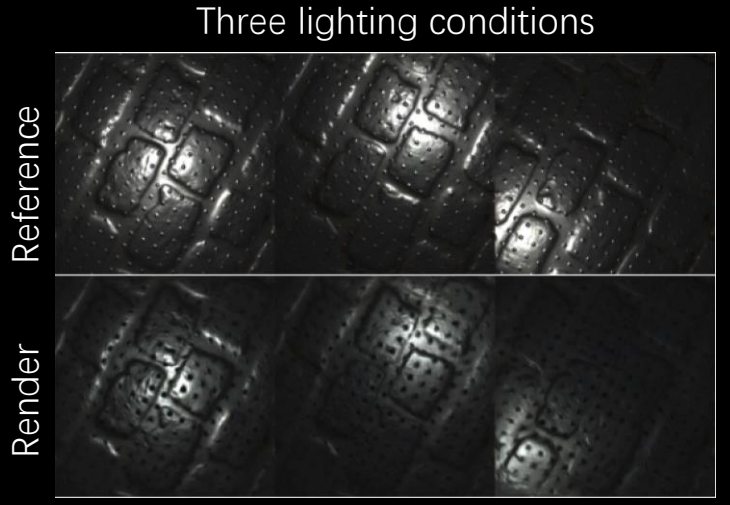
[Li et al. 2018]

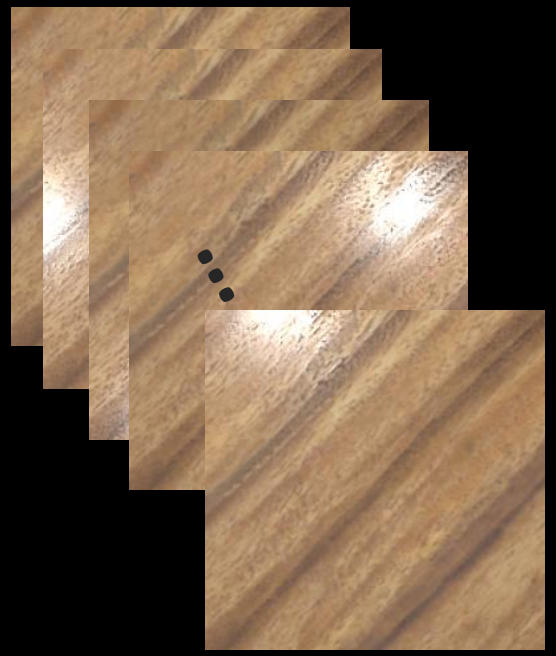
Our goal

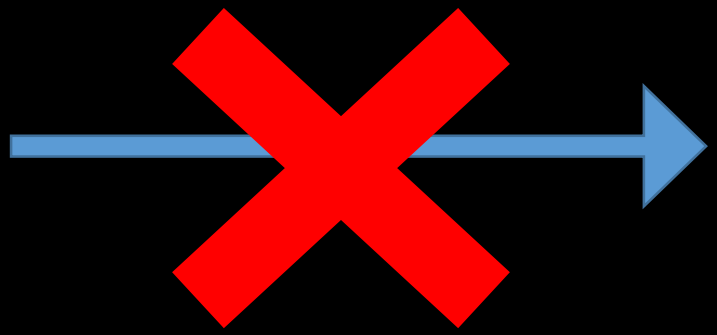
Challenges

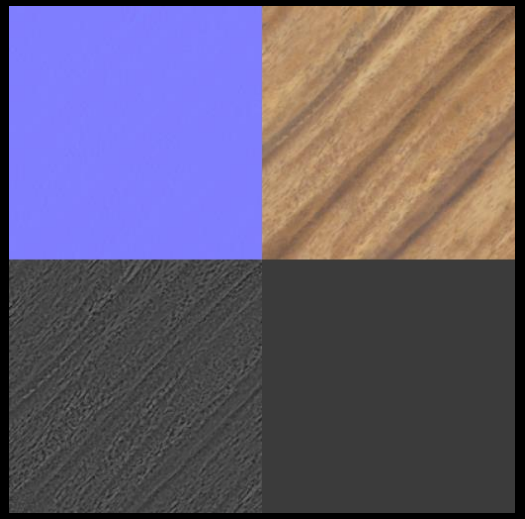
Non-trivial to combine current solutions

- Learning-based methods:
  hard to extend to arbitrary number of inputs





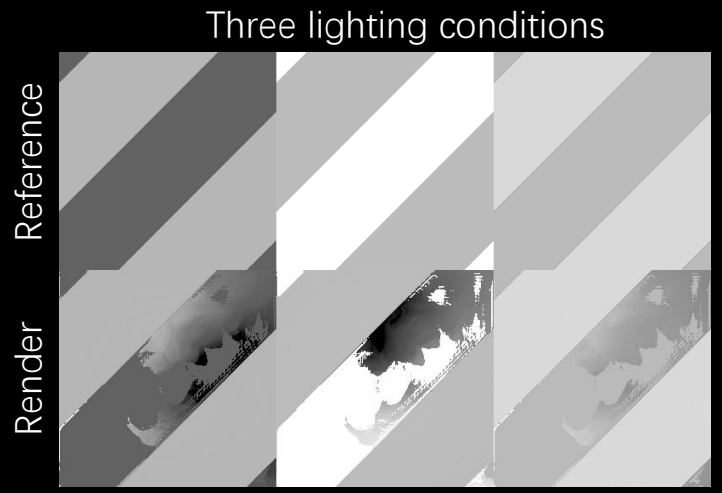


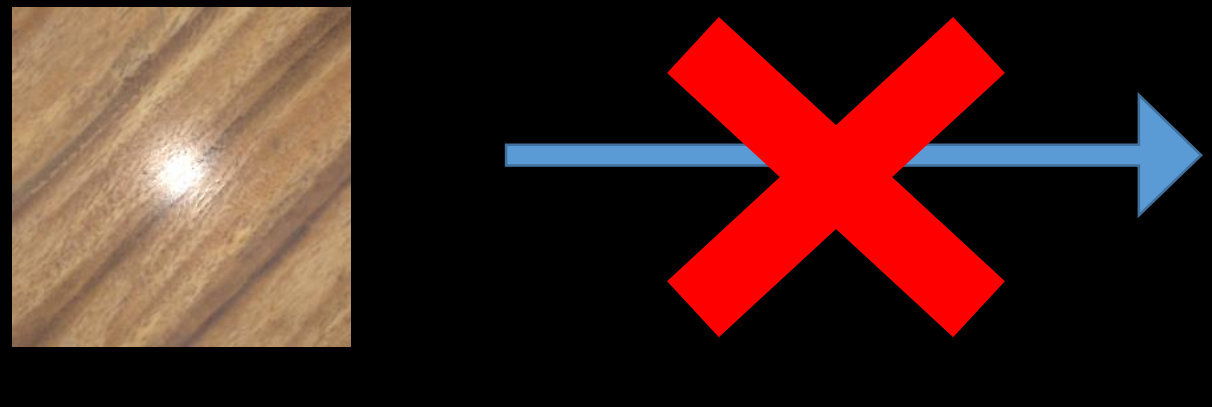


Challenges

Non-trivial to combine current solutions

- Classic Inverse Rendering:
  failed when input number is insufficient.

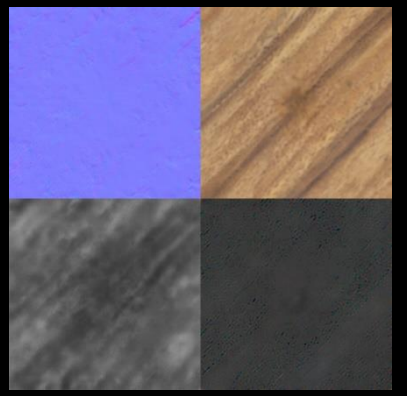


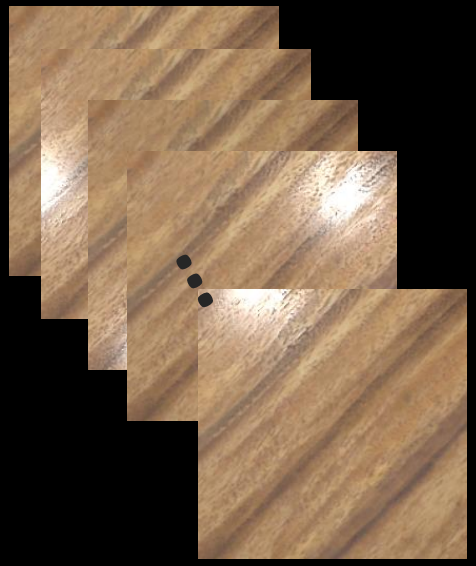


Our goal

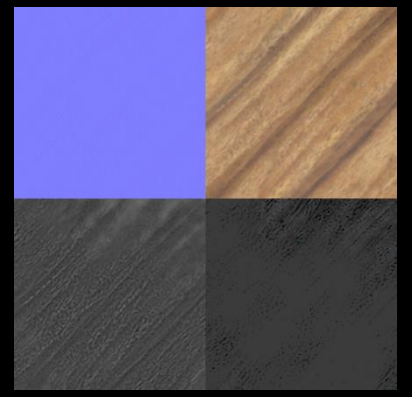
Single

Plausible





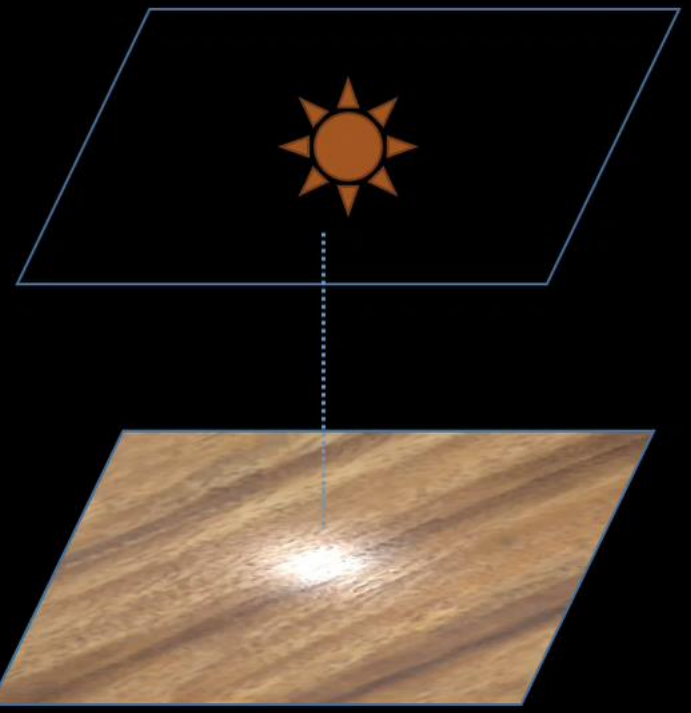
Multiple



Accurate

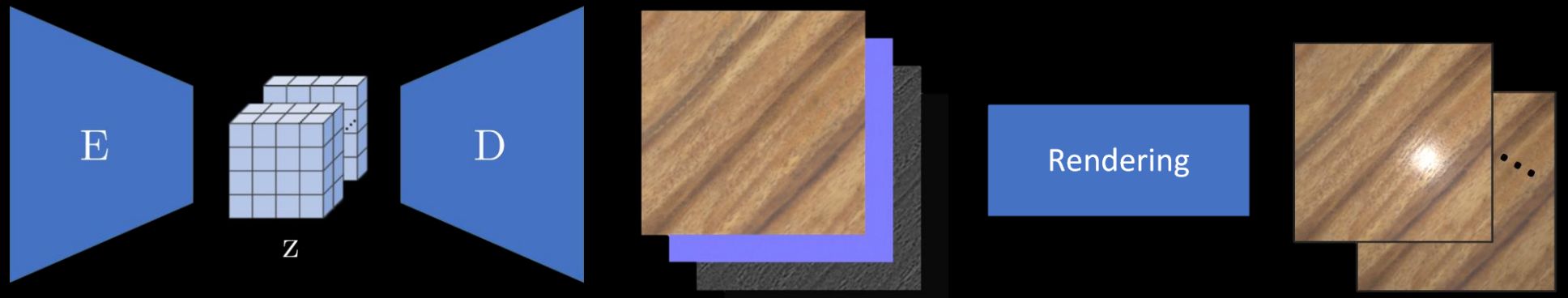
Assumptions

- Planar object
- Point light source collocated with the camera
- Fix distance between object plane and camera



Overview

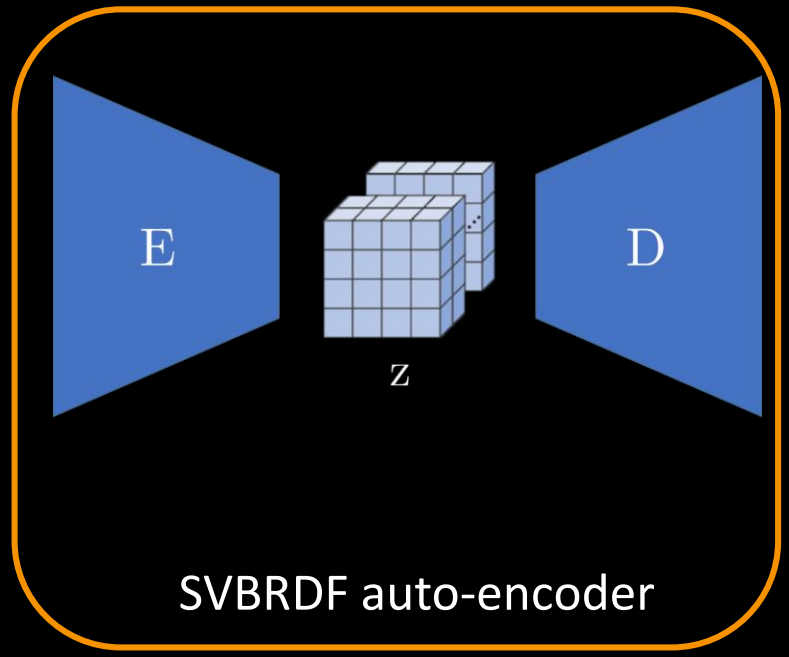
Key Idea: Deep Inverse Rendering

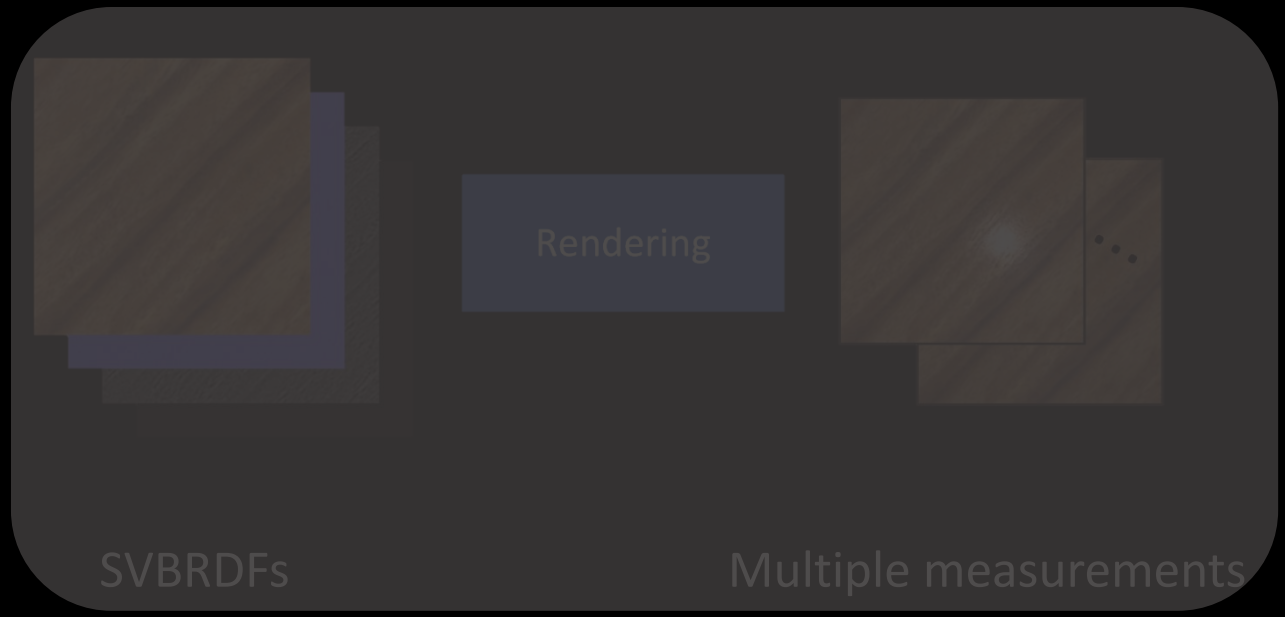


SVBRDF auto-encoder SVBRDFs Multiple measurements

Overview

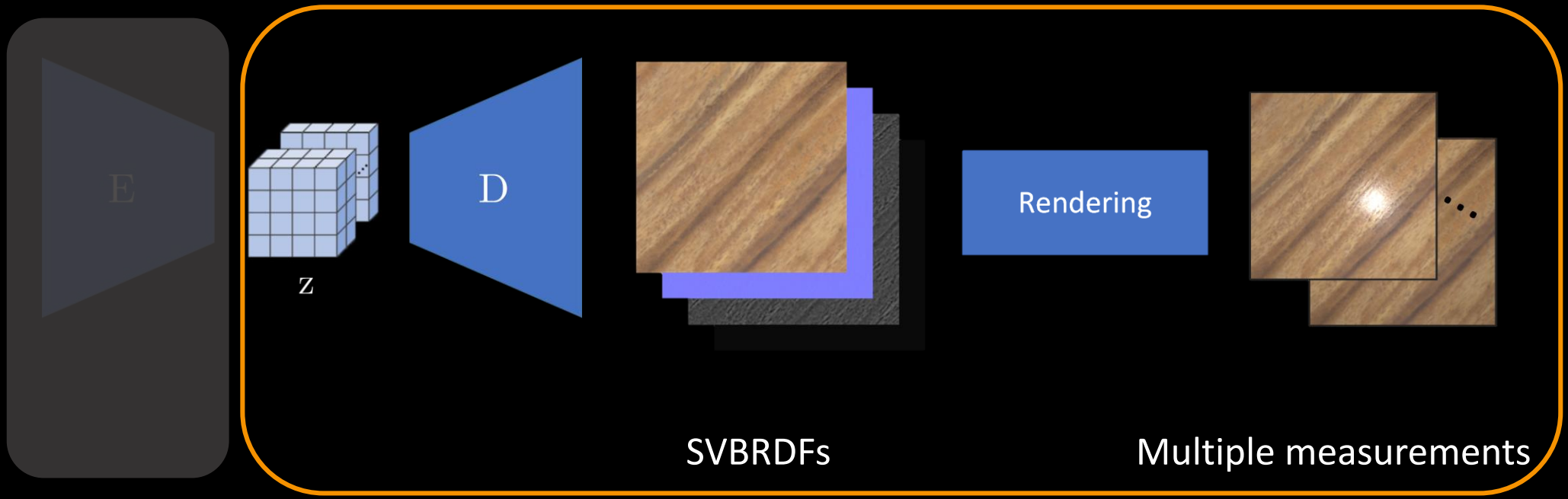
Key Idea: Deep Inverse Rendering



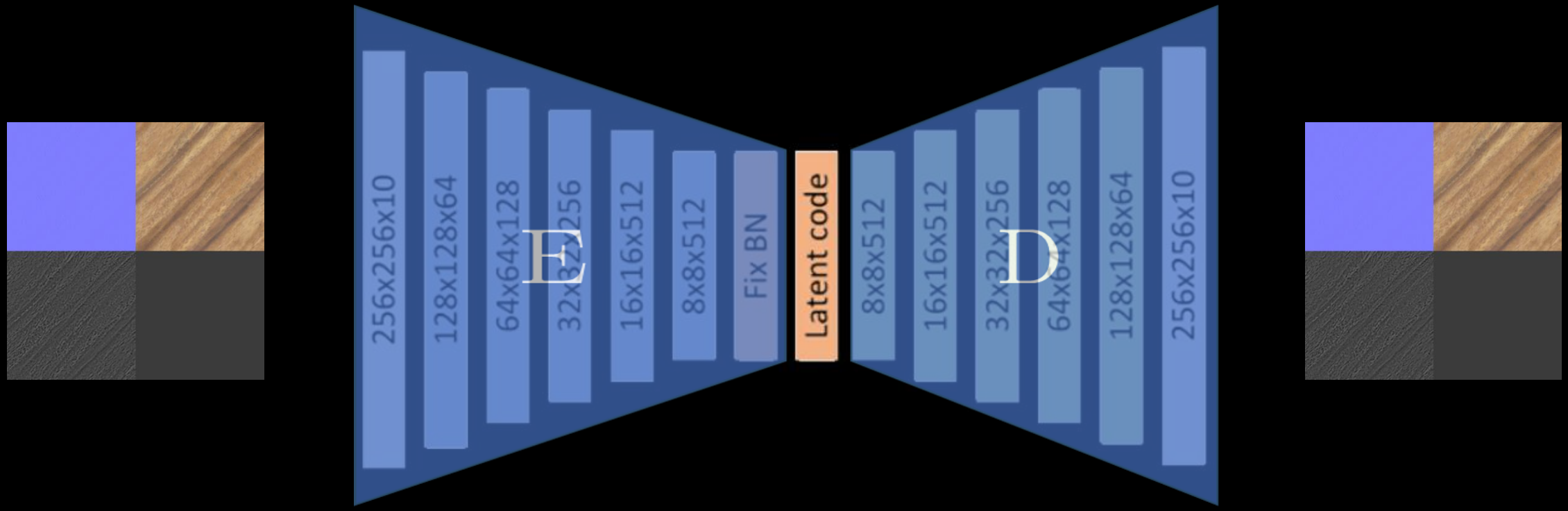


Overview

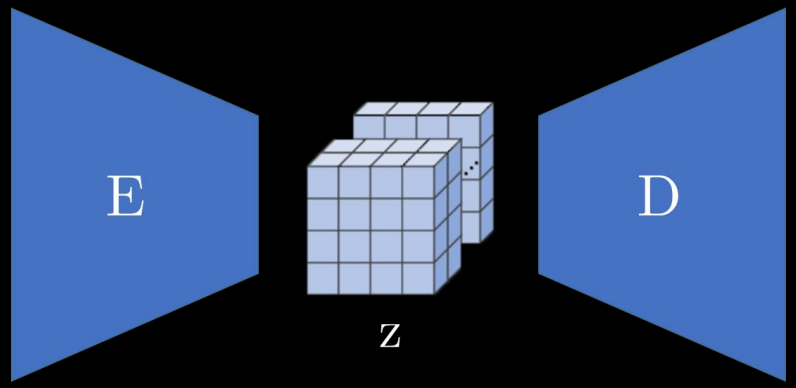
Key Idea: Deep Inverse Rendering



SVBRDF auto-encoder



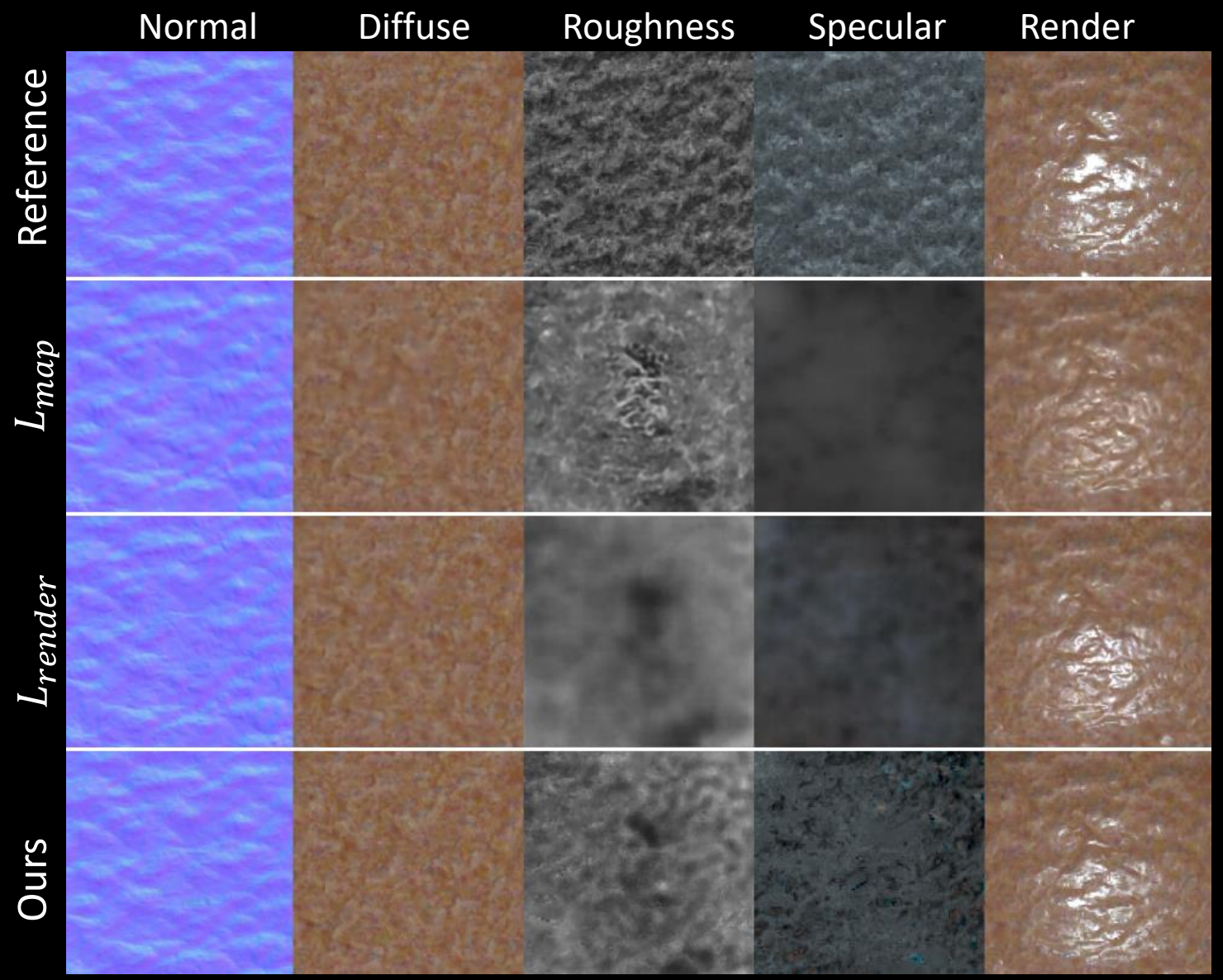
Training SVBRDF auto-encoder



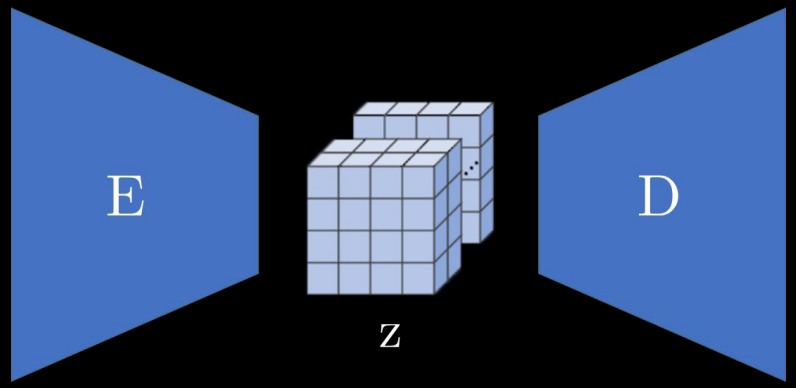
Training Loss:

$$\mathcal{L}_{train} = \mathcal{L}_{map} + \frac{1}{9}\mathcal{L}_{render}$$

Training SVBRDF auto-encoder



Training SVBRDF auto-encoder



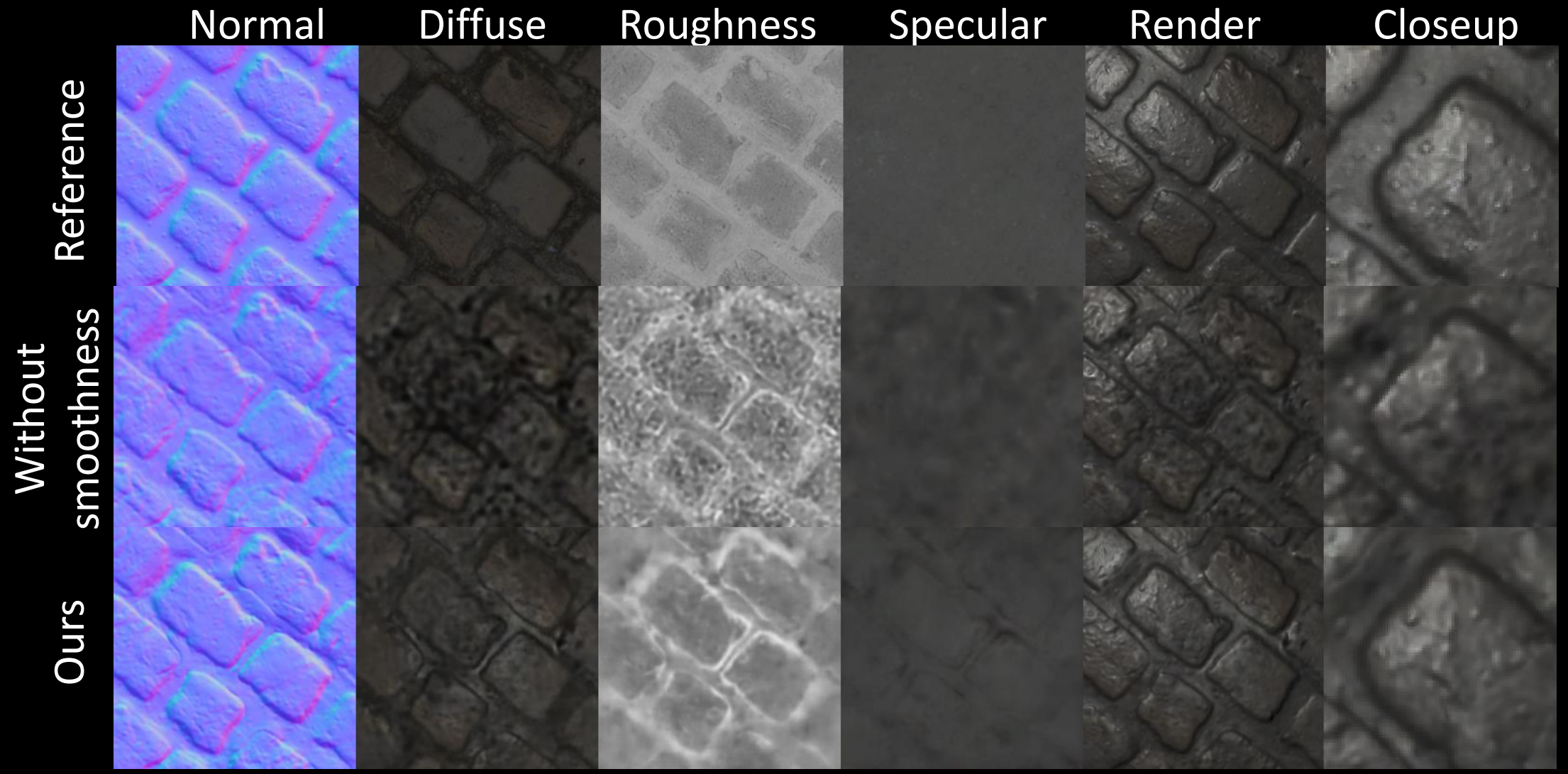
Training Loss:

$$\mathcal{L}_{train} = \mathcal{L}_{map} + \frac{1}{9}\mathcal{L}_{render}$$

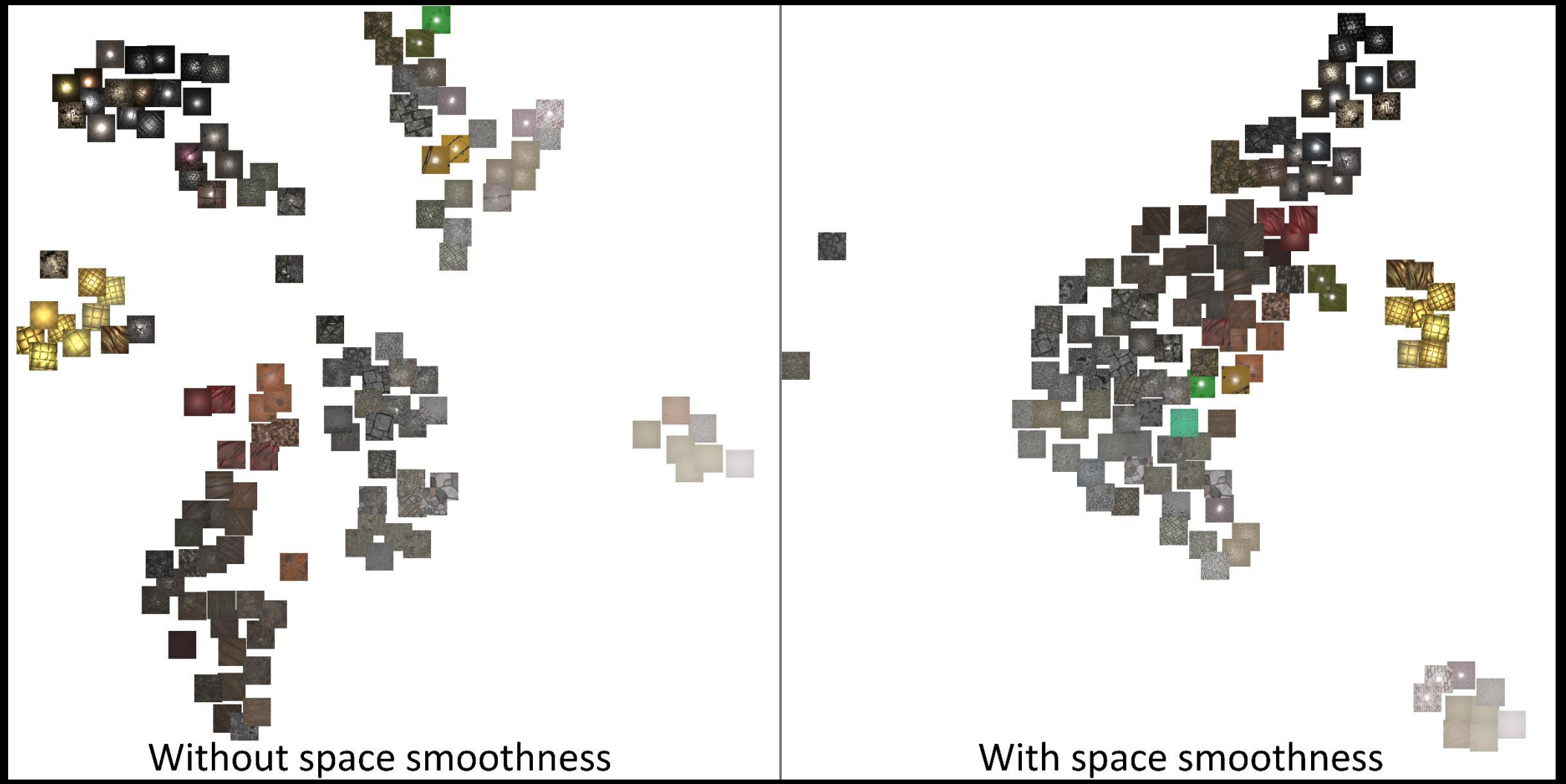
Latent space smoothness:

$$\mathcal{L}_{smooth} = \lambda_{smooth}||D(z) - D(z + \xi)||_1$$

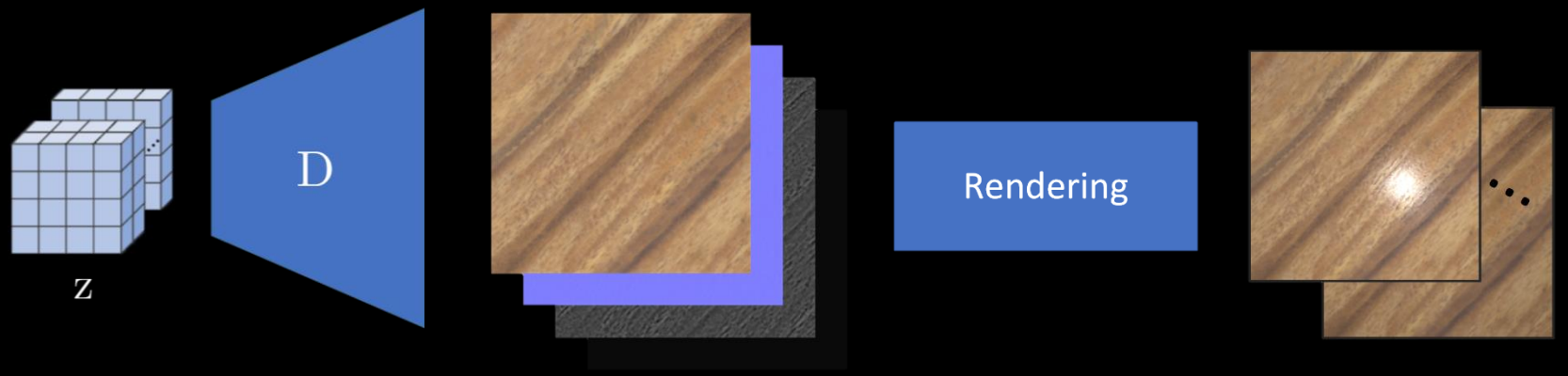
Training SVBRDF auto-encoder



t-SNE visualizations

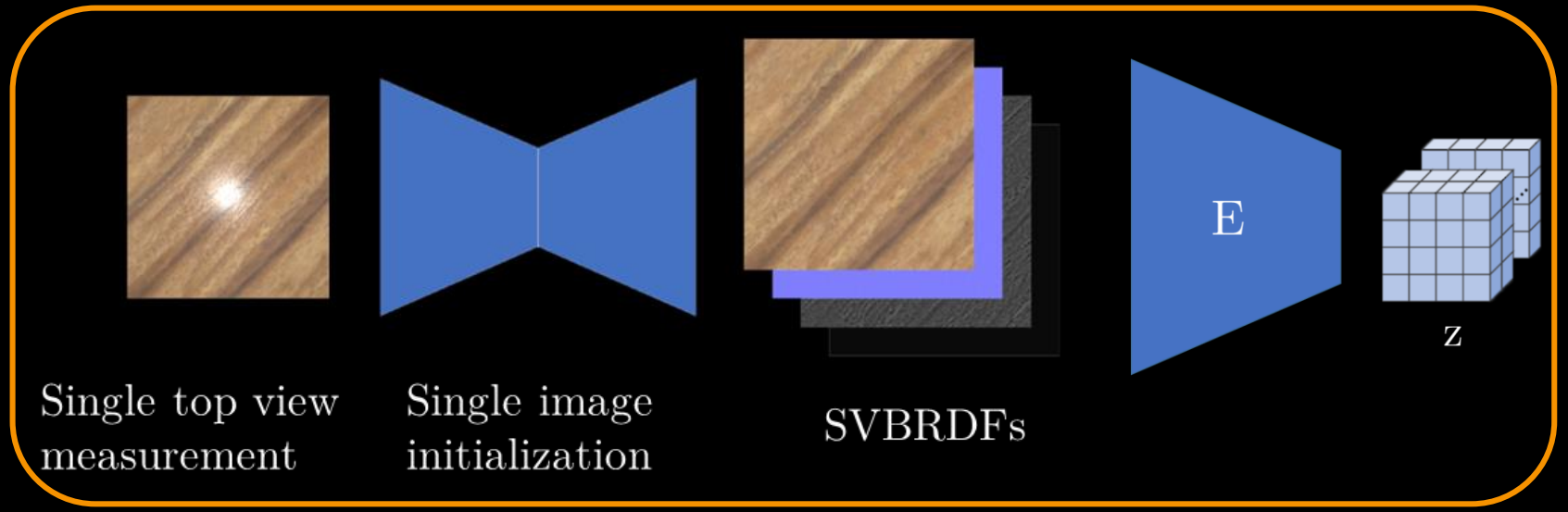


Optimize latent code from measurement(s)

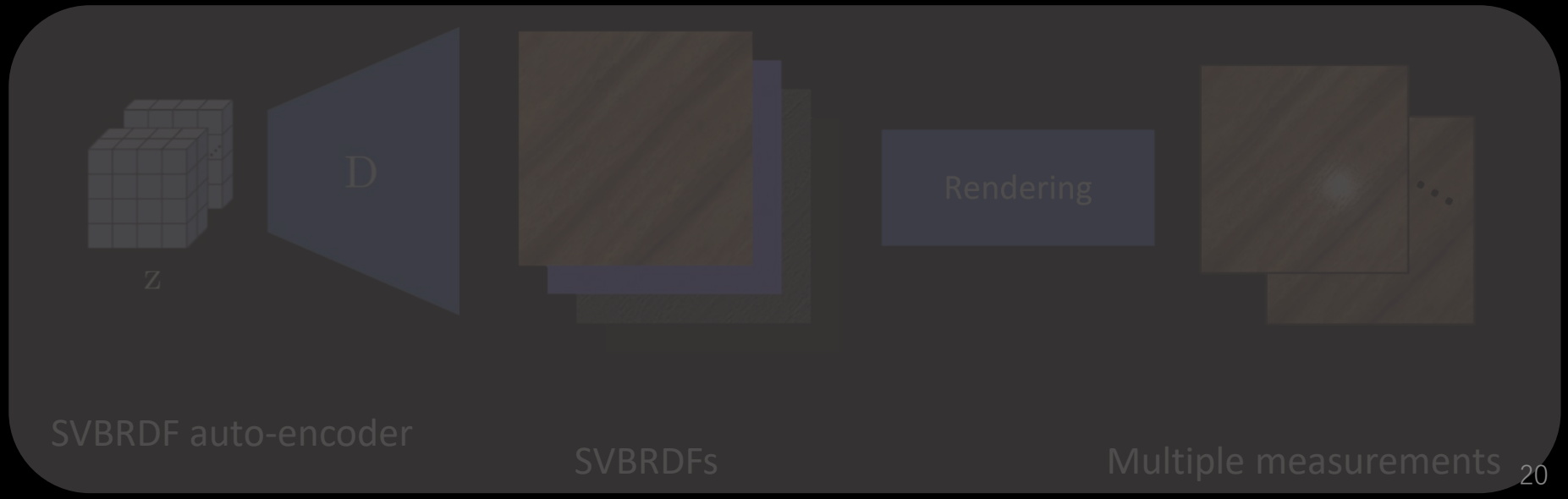


SVBRDF auto-encoder SVBRDFs Multiple measurements

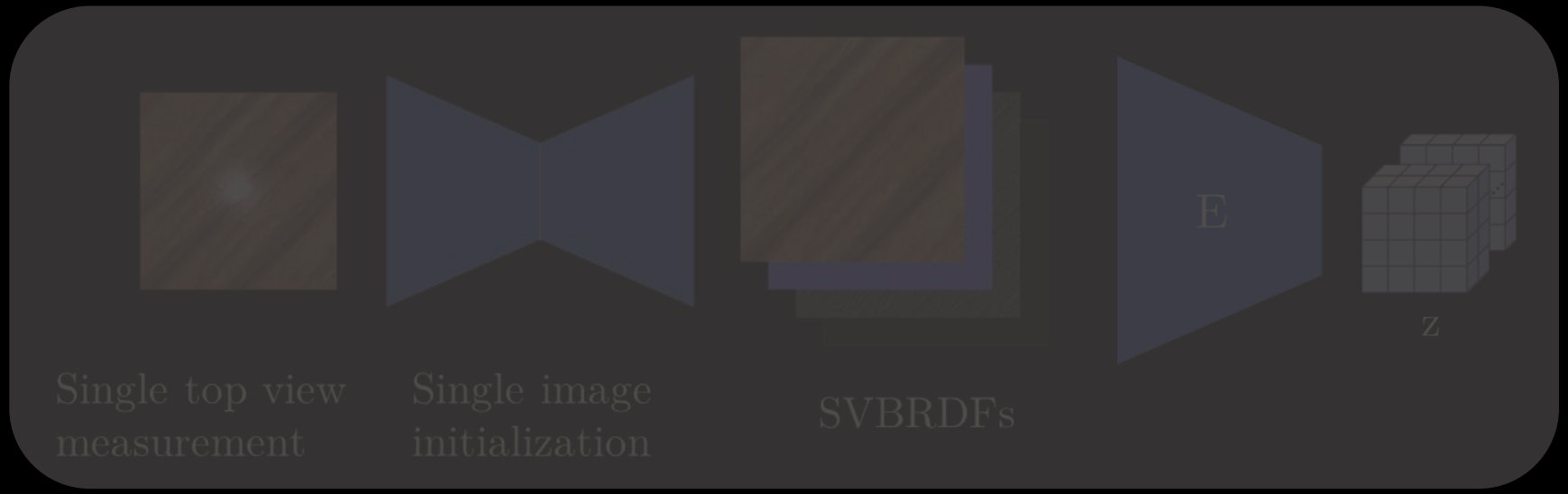
Bootstrap the optimization

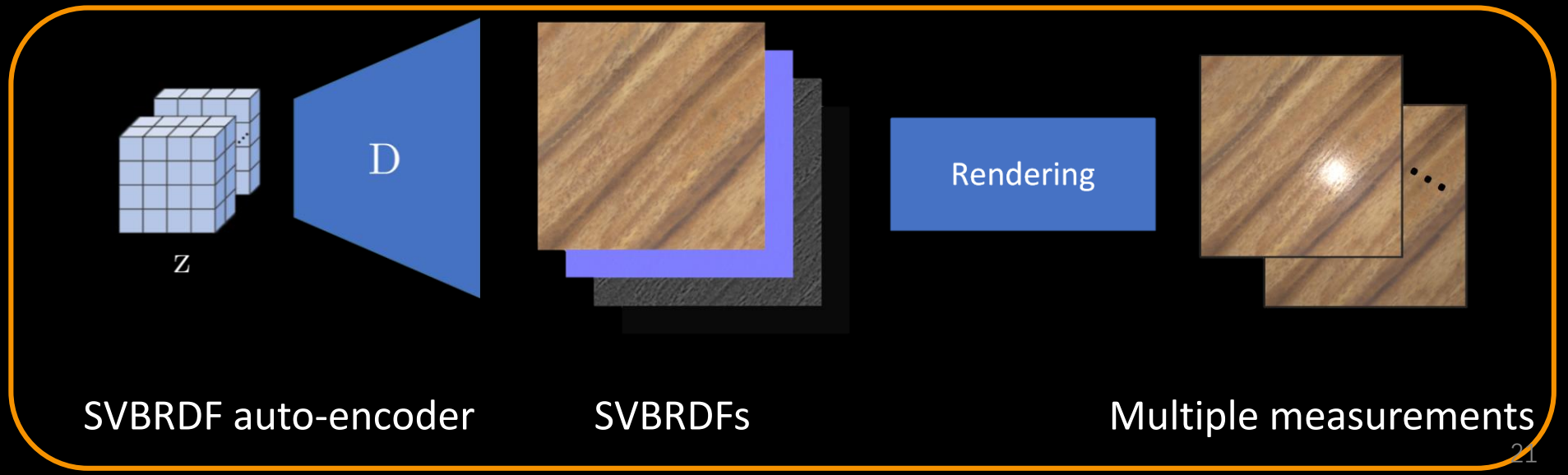


State-of-the art single input network [Deschaintre et al.]

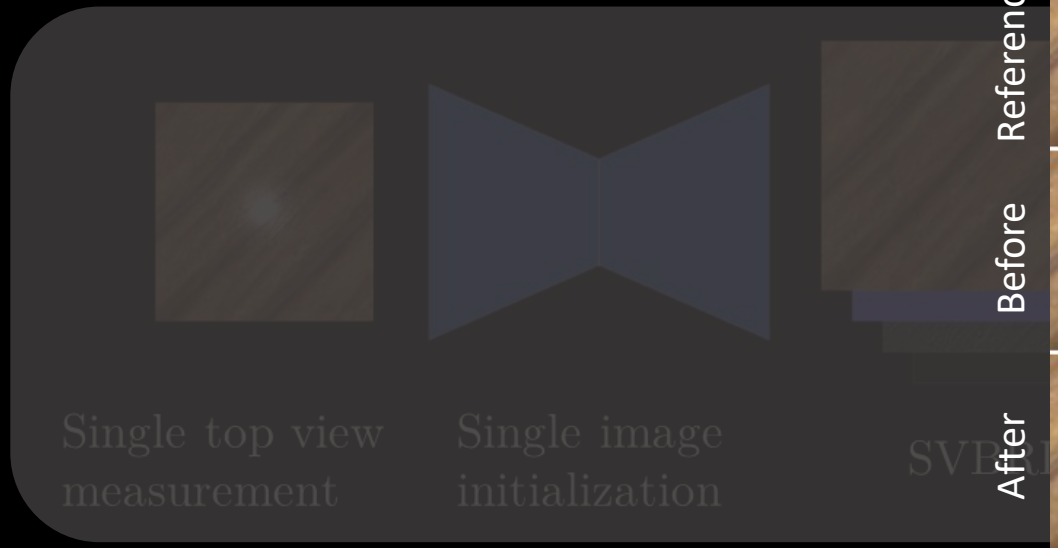


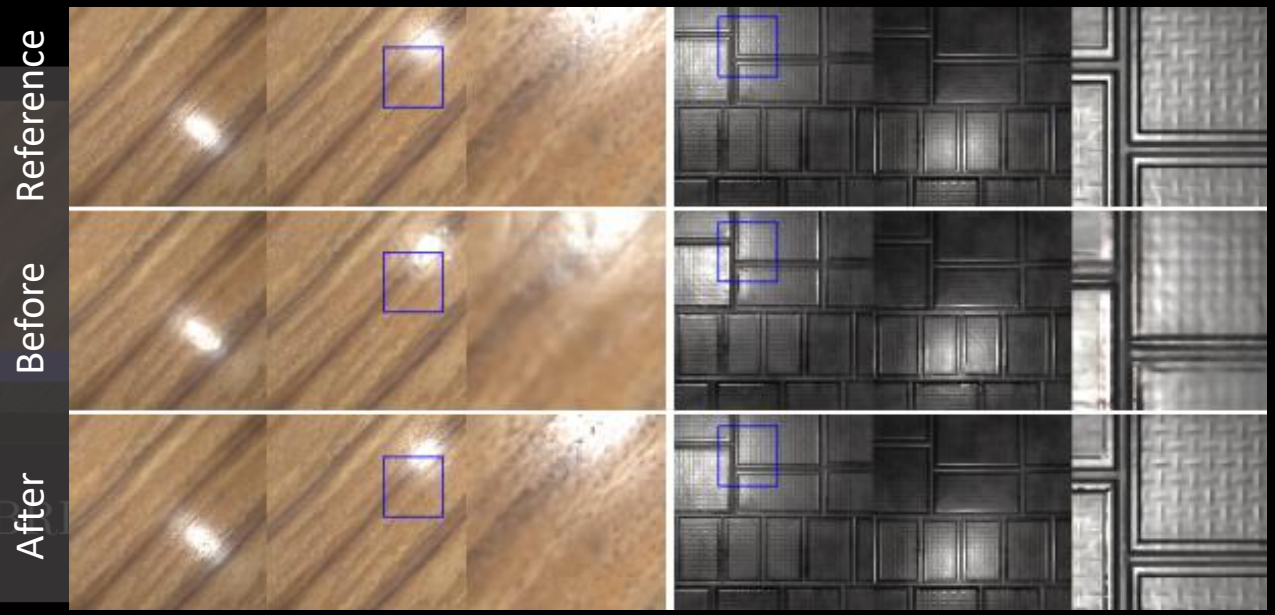
Optimize in latent space

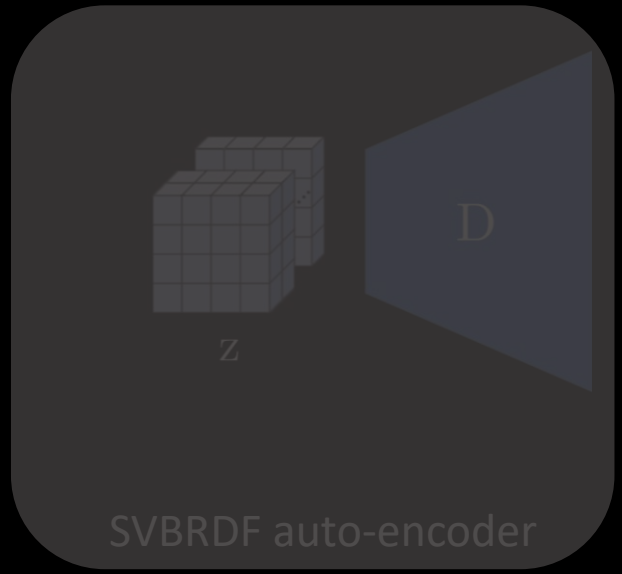


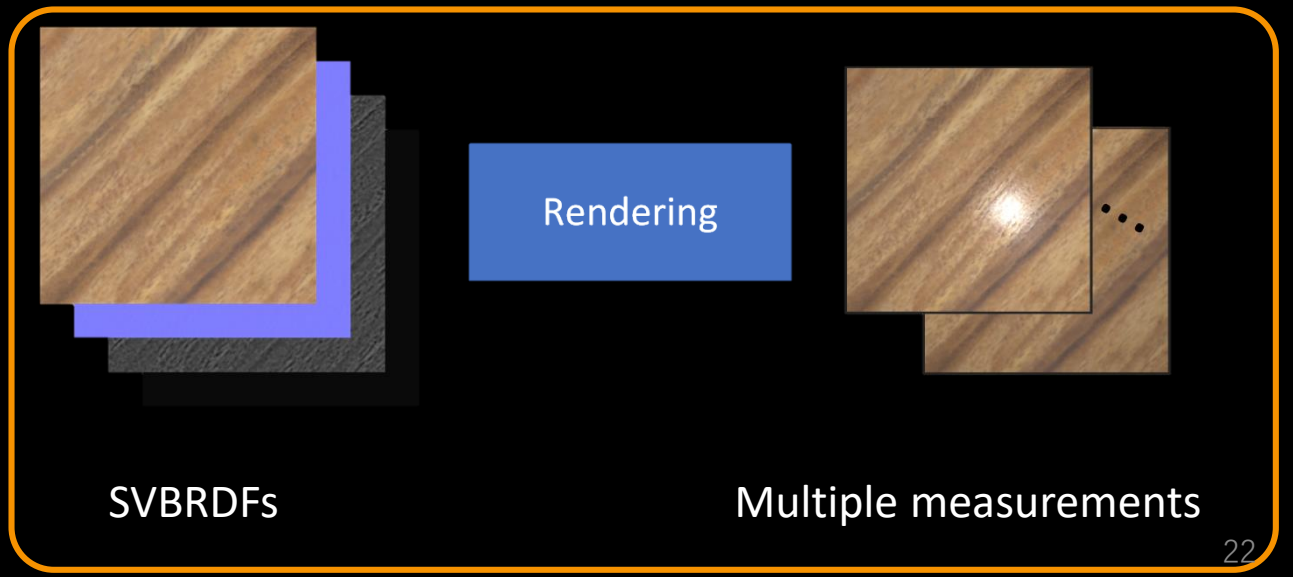


Detail refinement

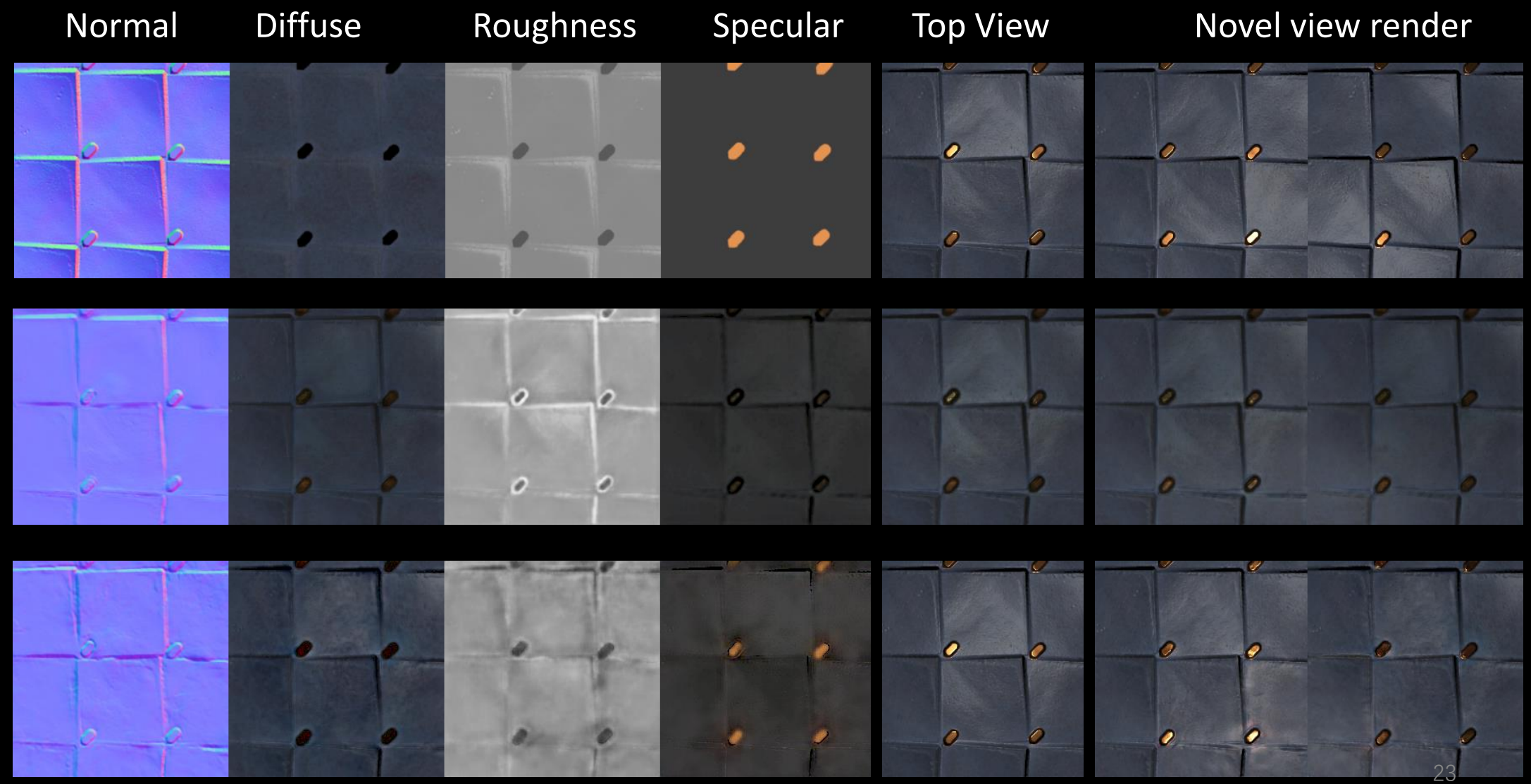




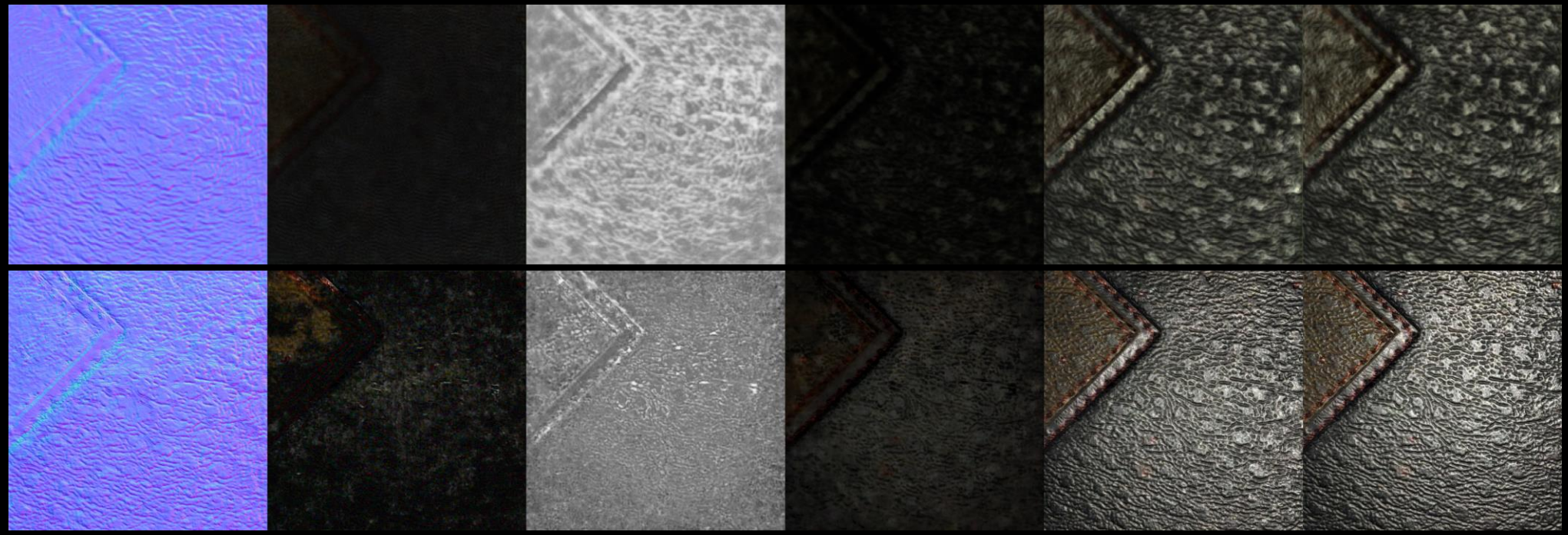


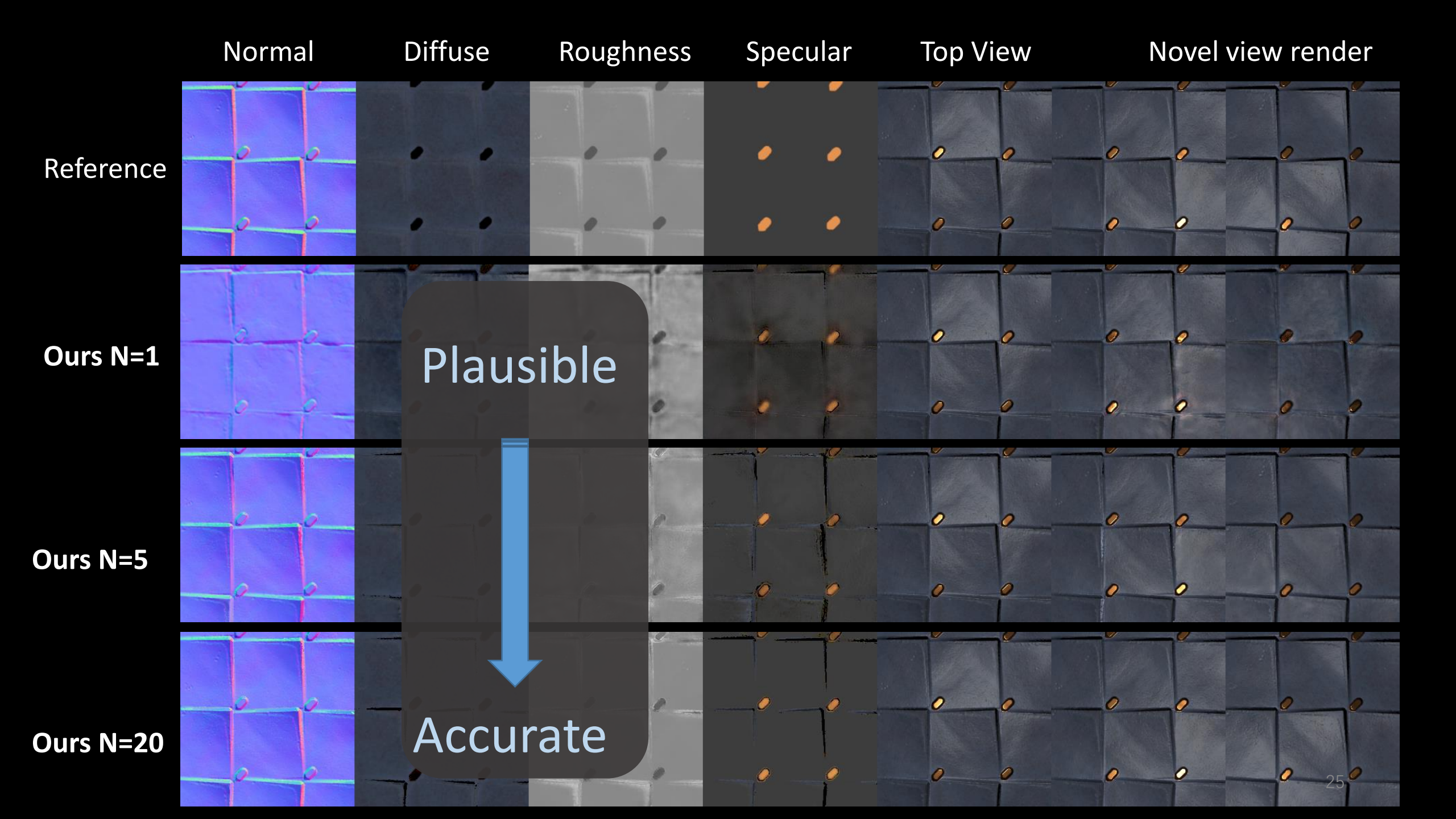


Improved quality with single input



Improved quality with single input

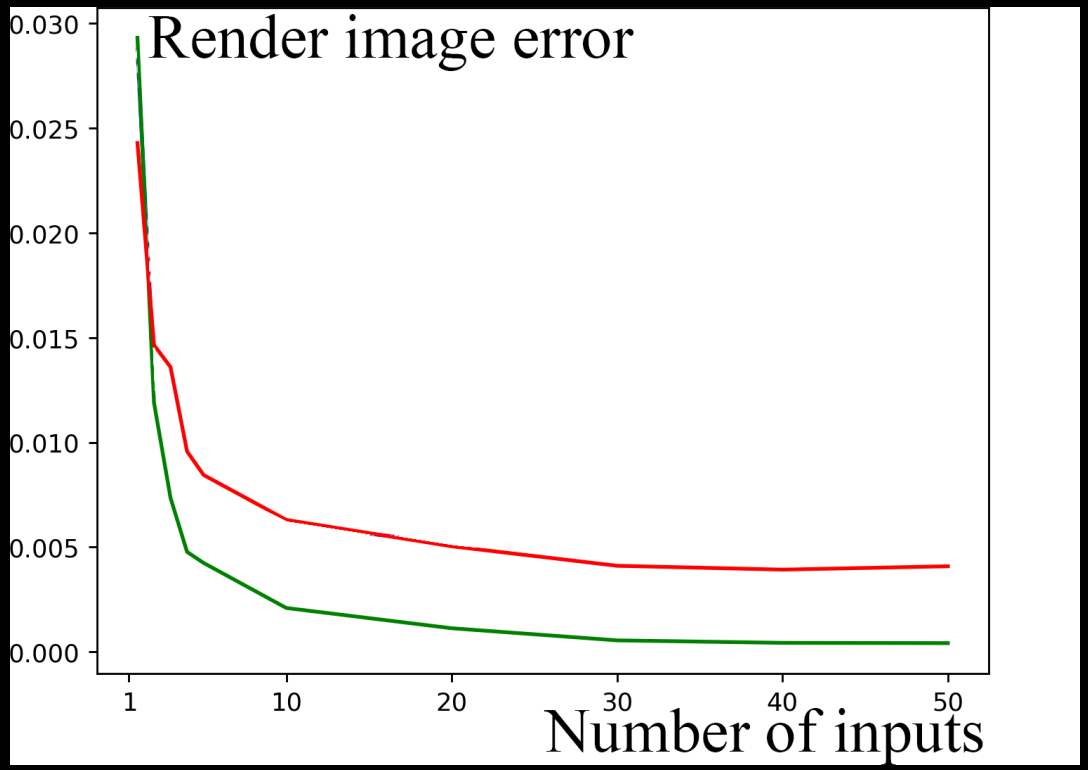


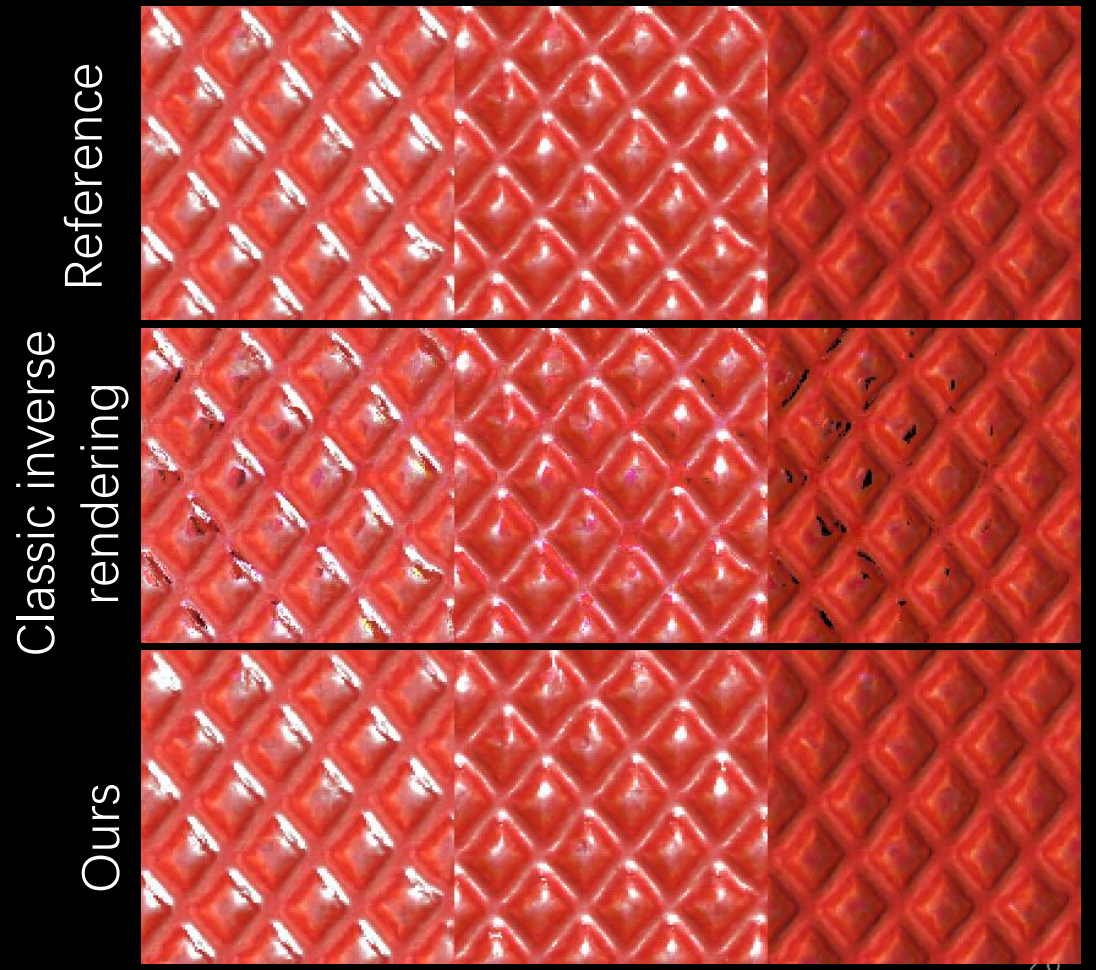

Normal
Diffuse
Roughness
Specular
Top View
Novel view render
Reference
Ours N=1
Ours N=5
Ours N=20
Plausible
Accurate

Comparison with class inverse rendering

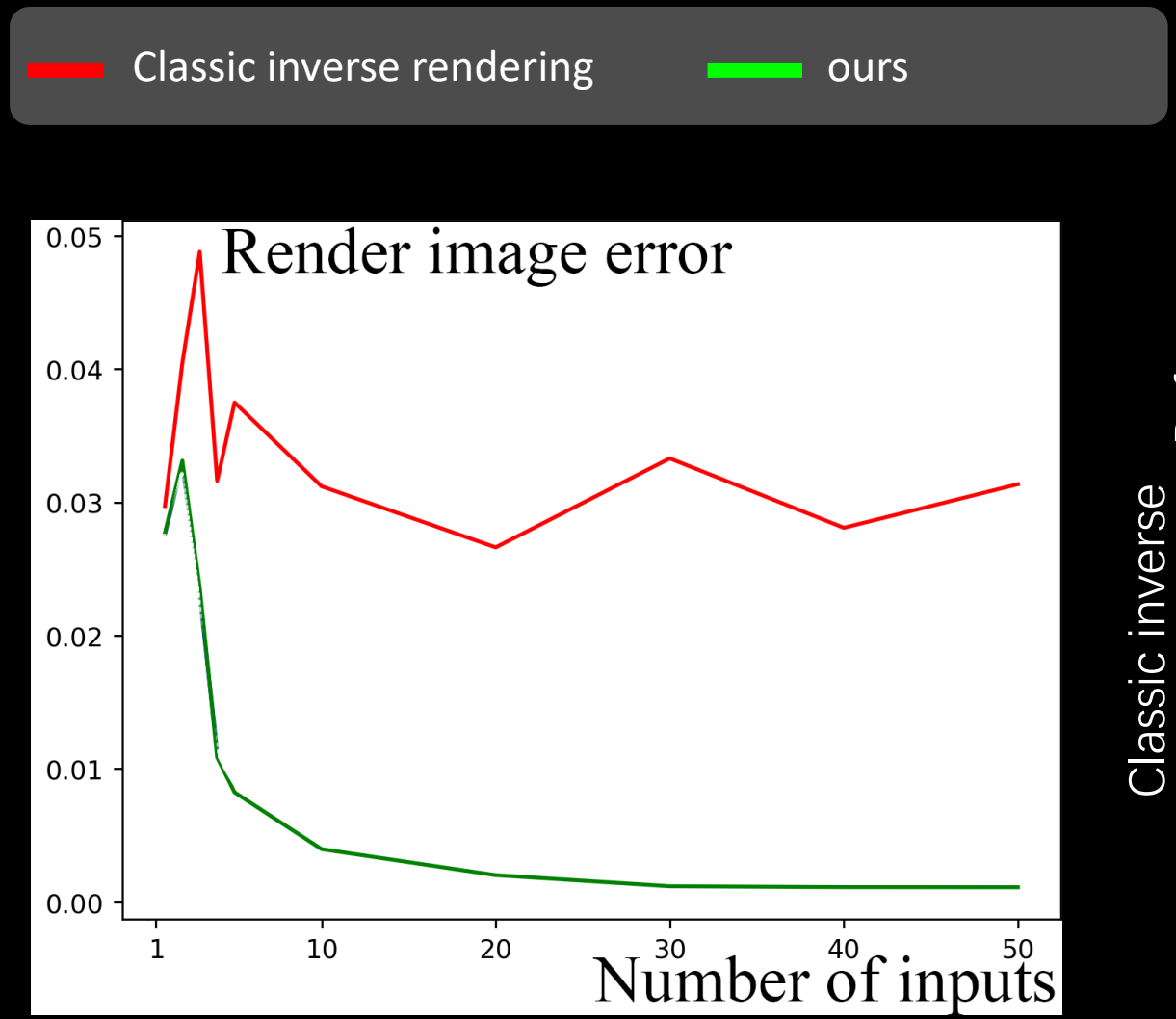
Classic inverse rendering

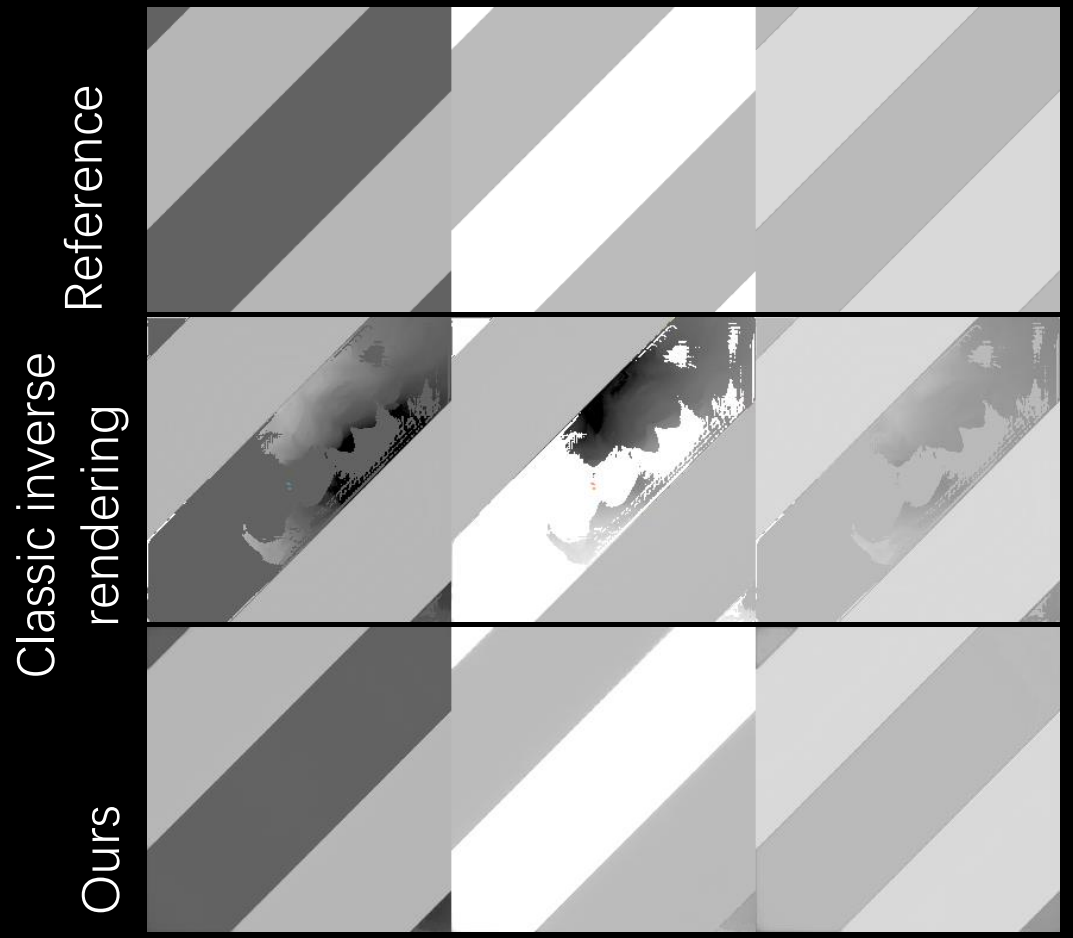
ours



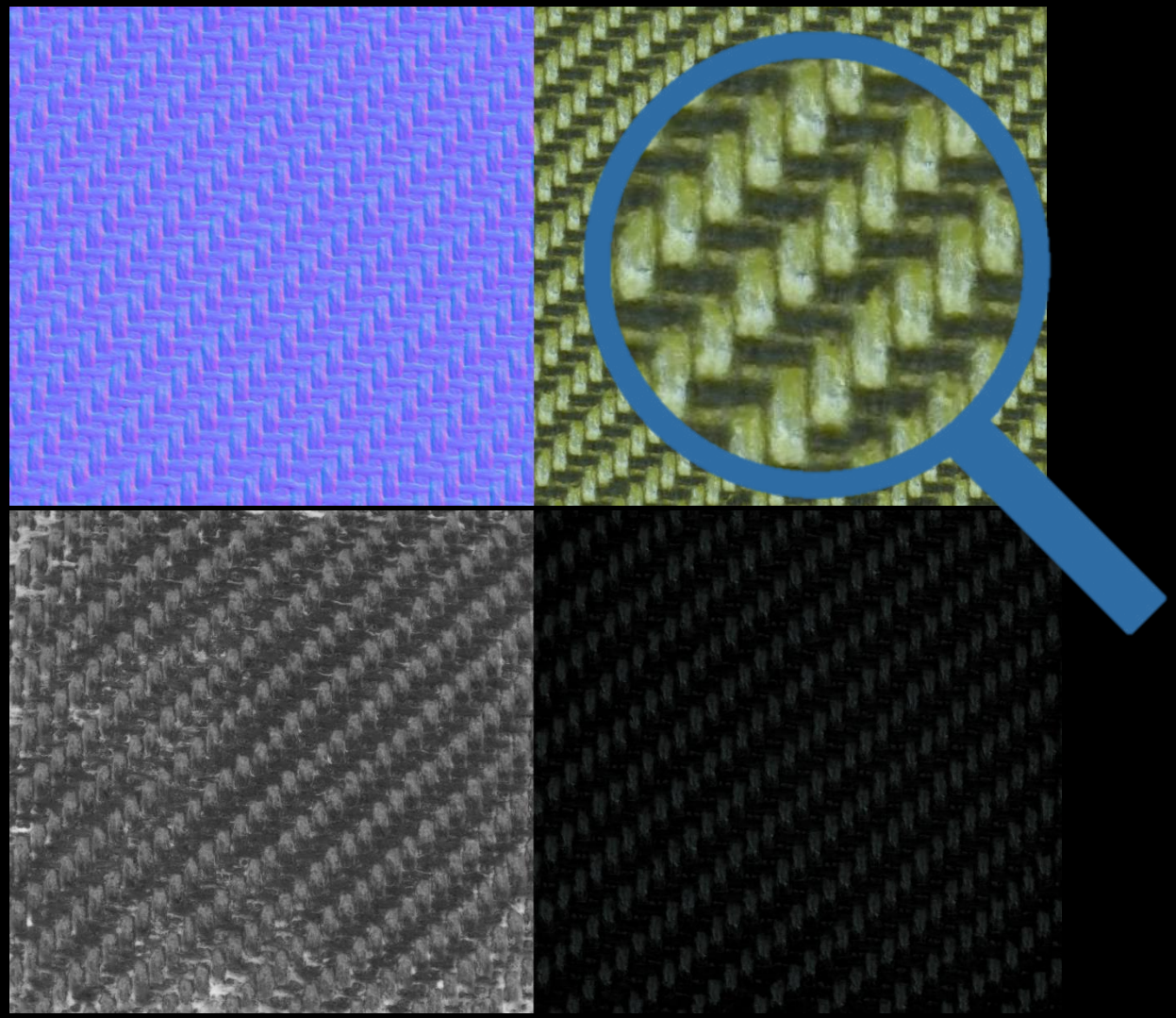


Comparison with class inverse rendering



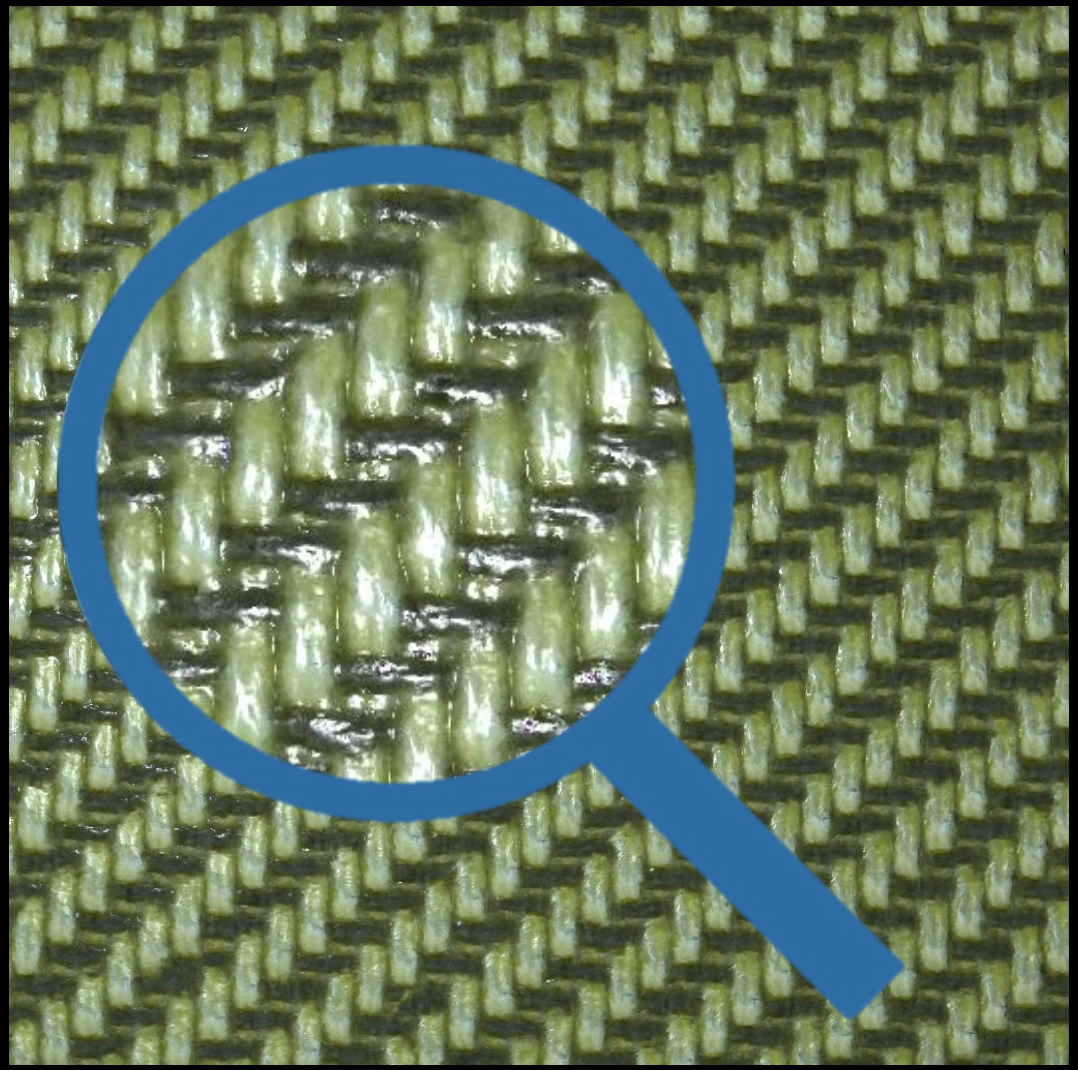


High resolution results



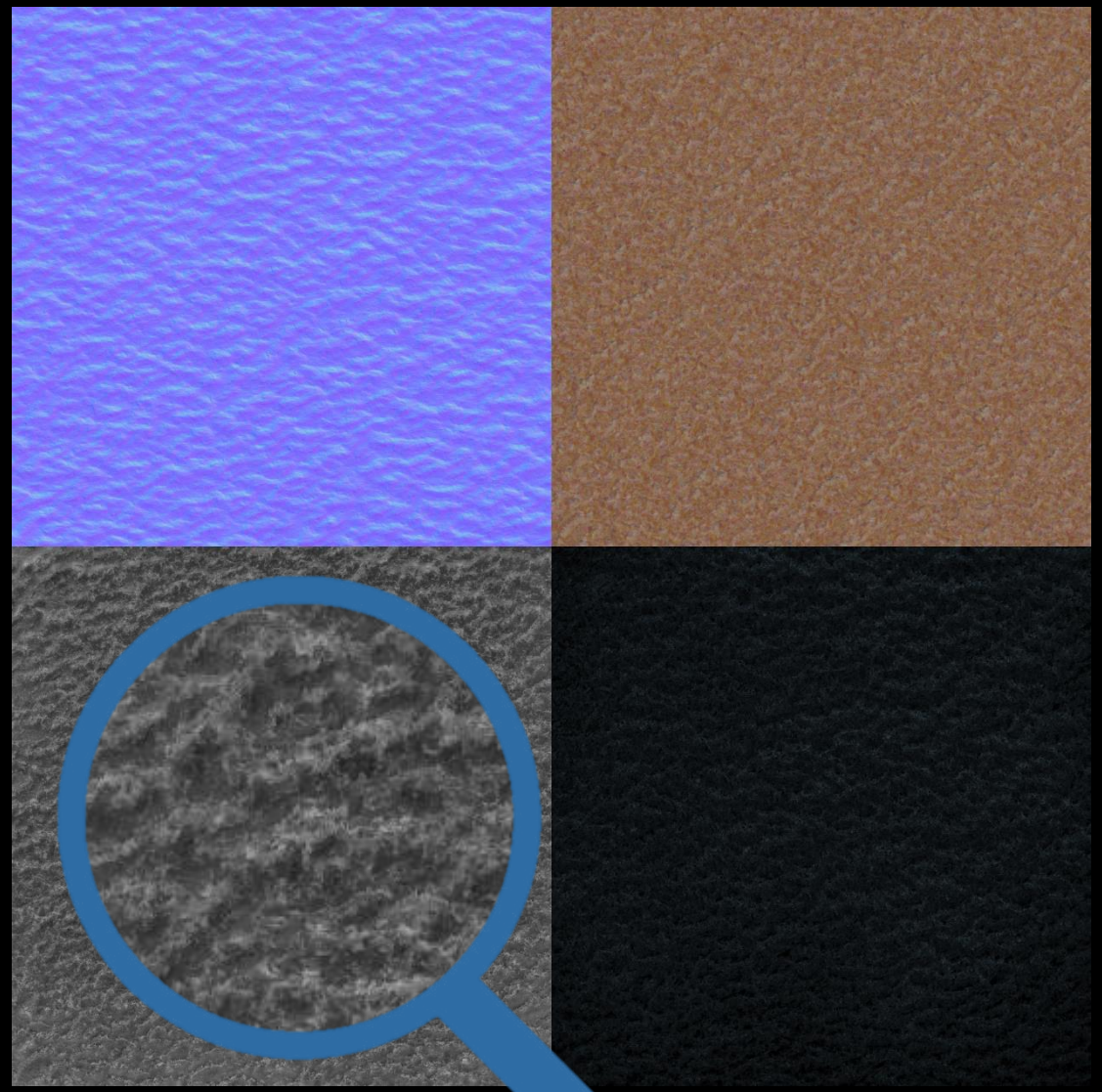
Estimated SVBRDF with 20 input photos

Support arbitrary resolution!



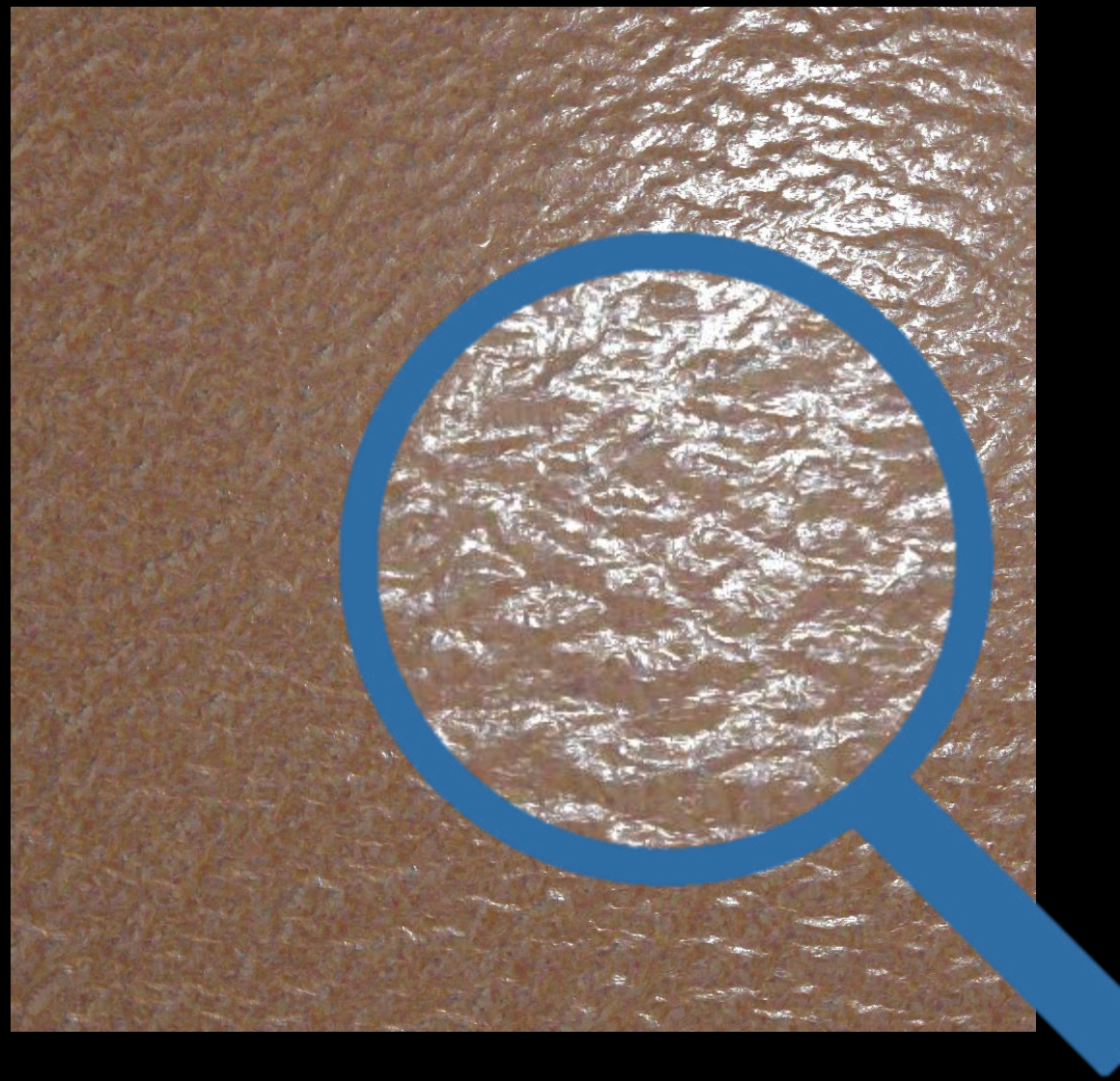
Novel view rendering

High resolution results



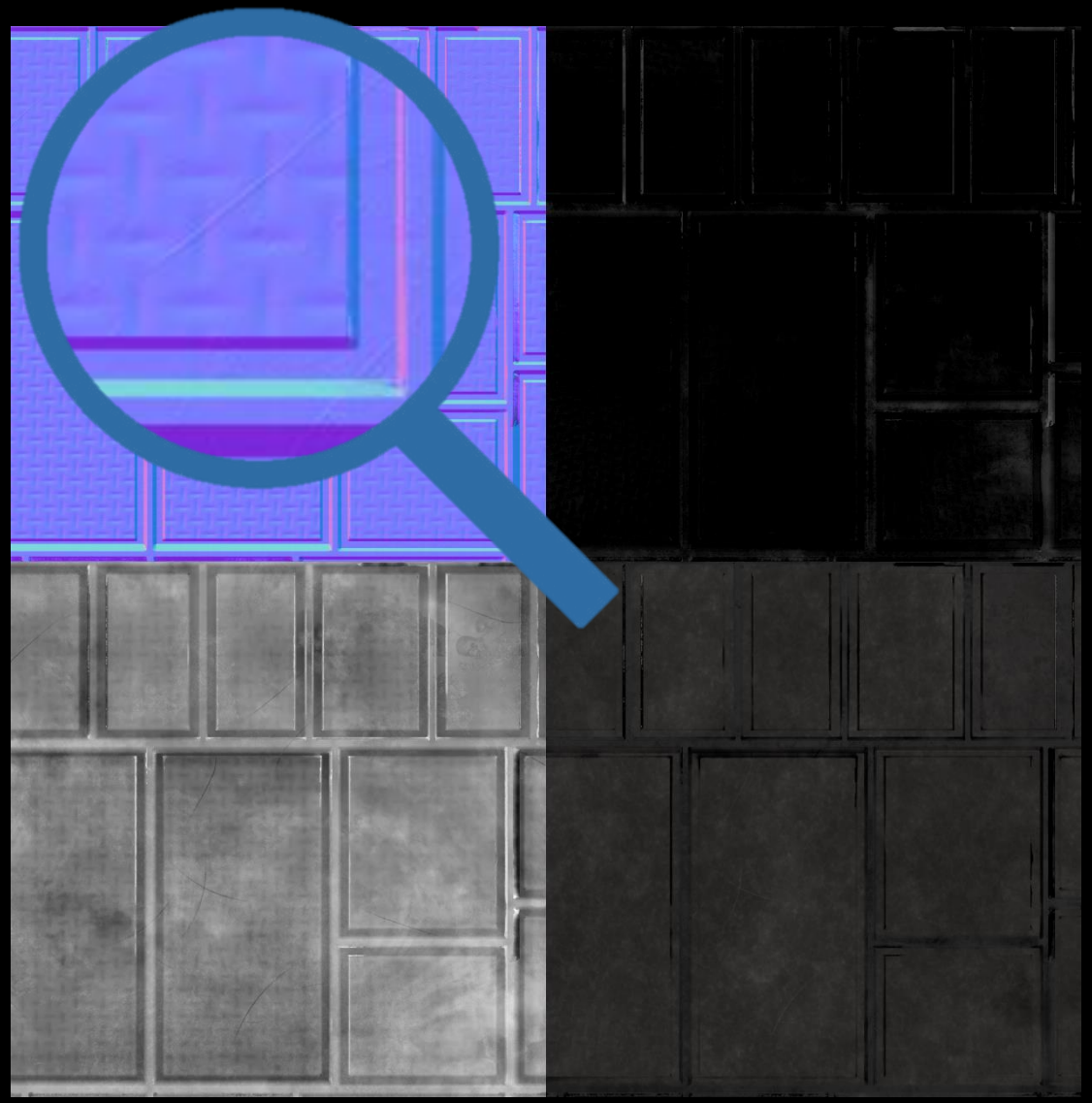
Estimated SVBRDF with 20 input photos

Support arbitrary resolution!



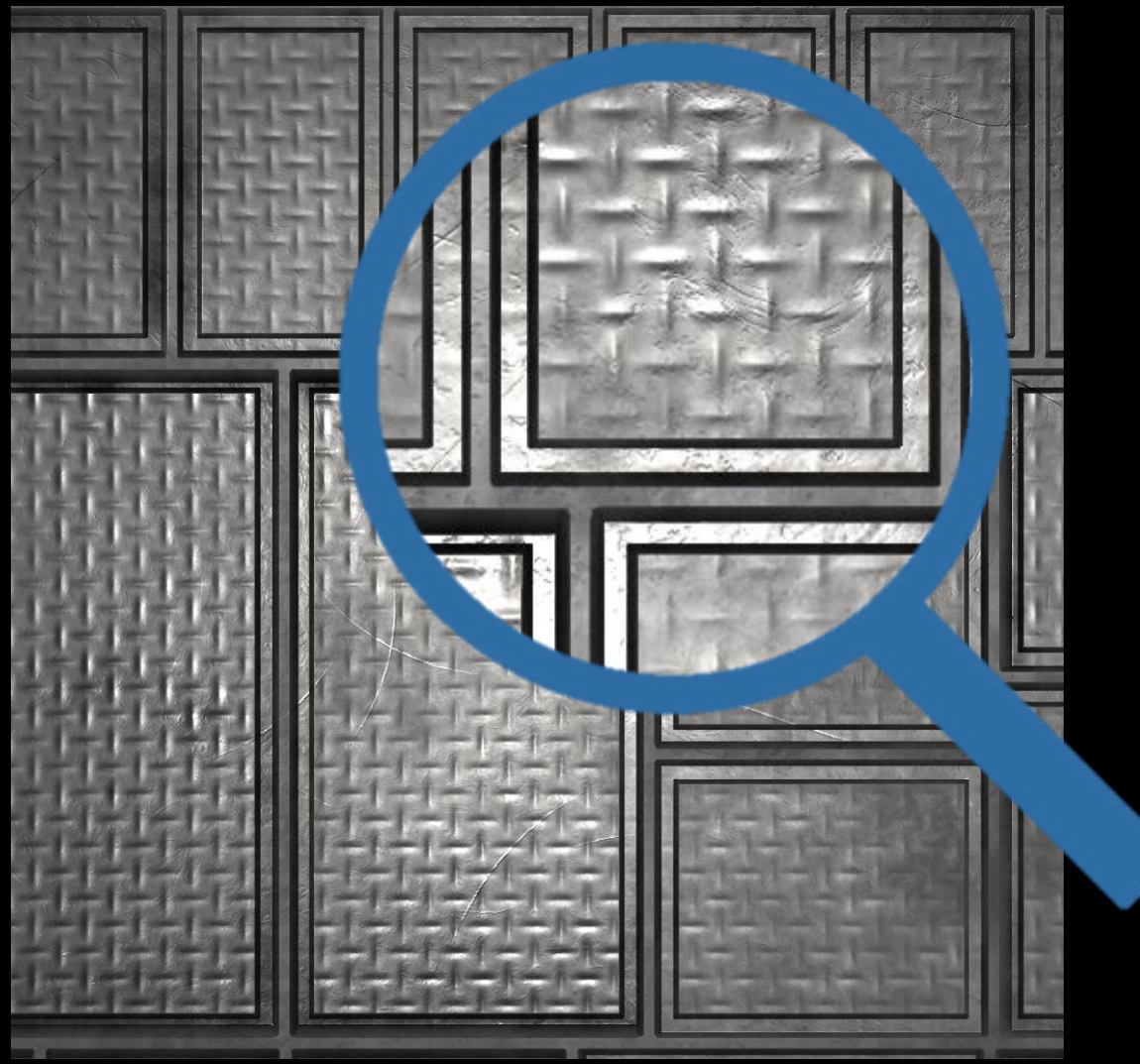
Novel view rendering

High resolution results



Estimated SVBRDF with 20 input photos

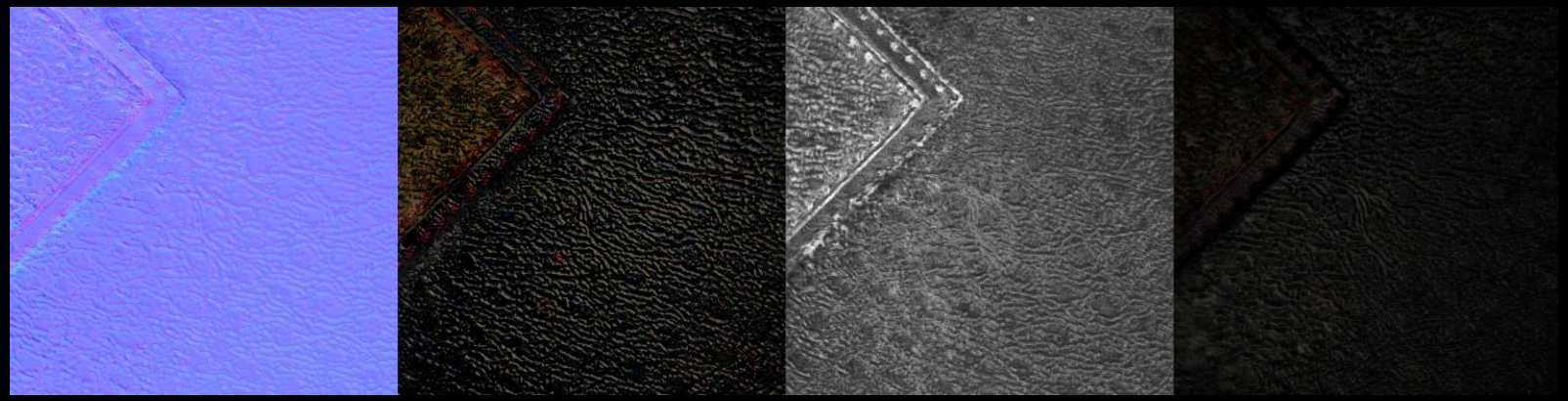
Support arbitrary resolution!



Novel view rendering

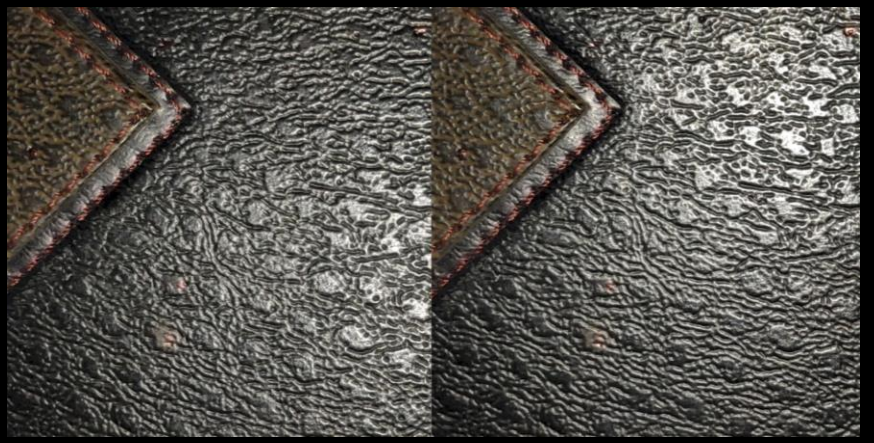
Real captured results

Leather, 1k resolution, 2 inputs

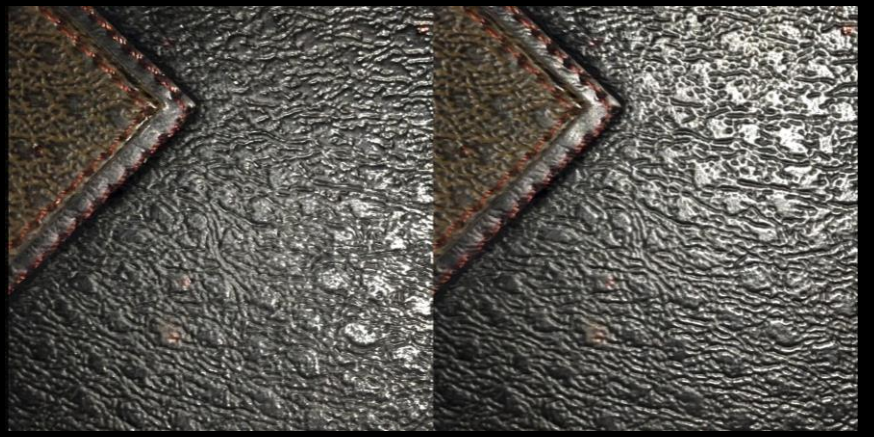


Normal **Diffuse** **Roughness** **Specular**

GT



Render



Novel view

Real captured results

Wood, 1k resolution, 10 inputs

Normal **Diffuse** **Roughness** **Specular**

GT

Render

Novel view

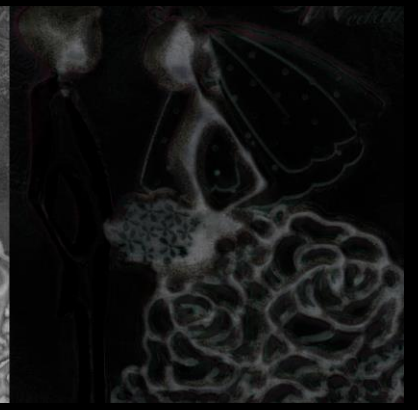
Real captured results

Metal Plate, 1k resolution, 30 inputs

Normal

Diffuse

Roughness



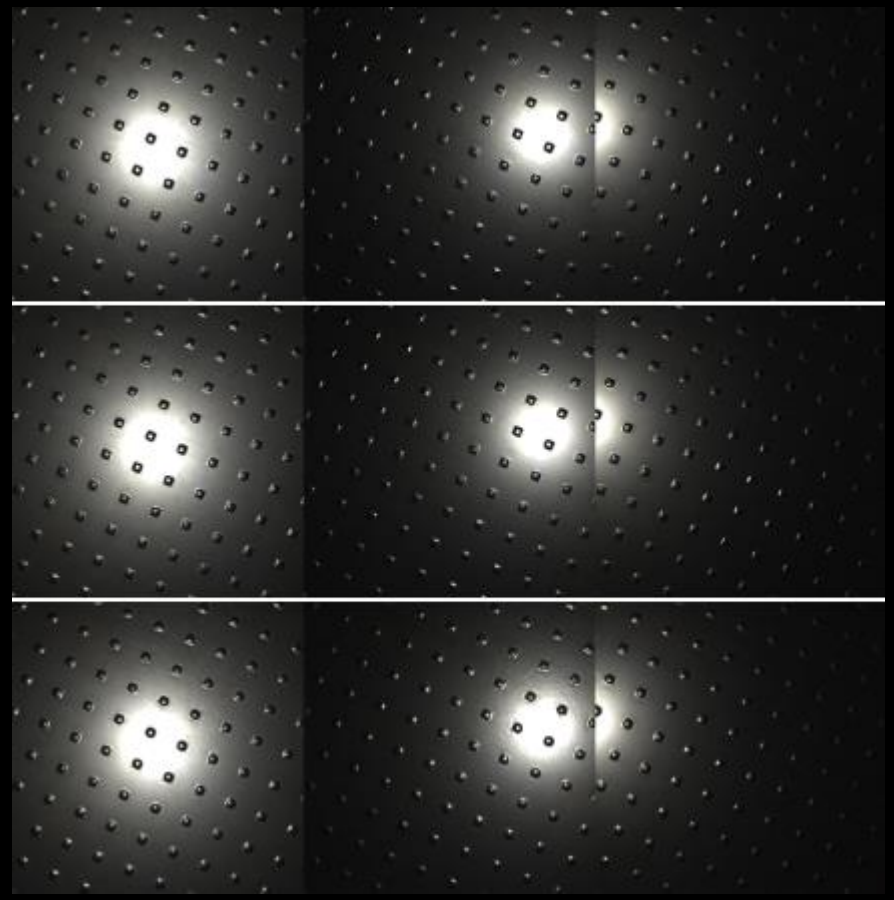
Specular

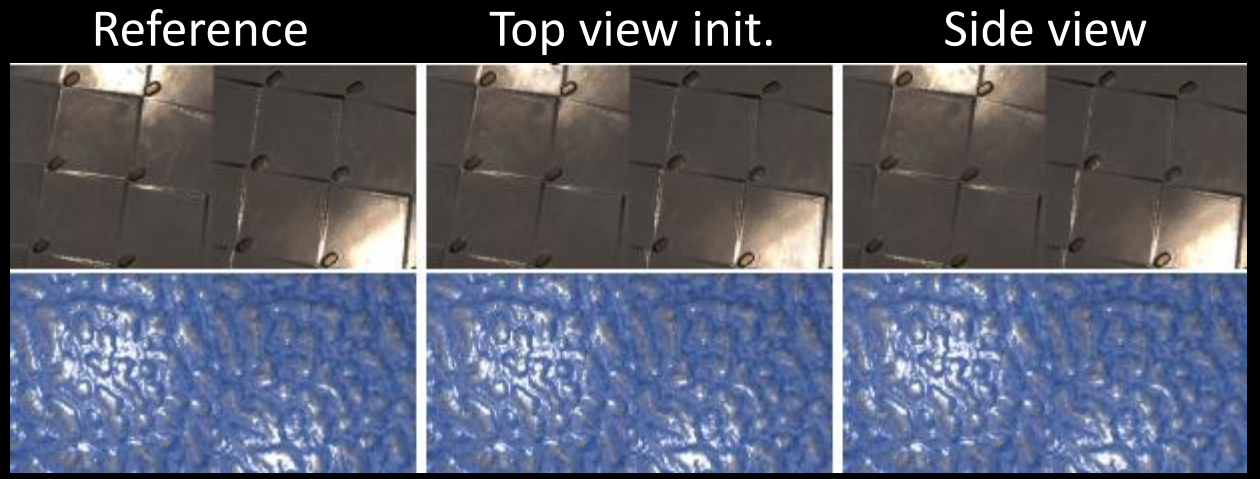
GT

Render

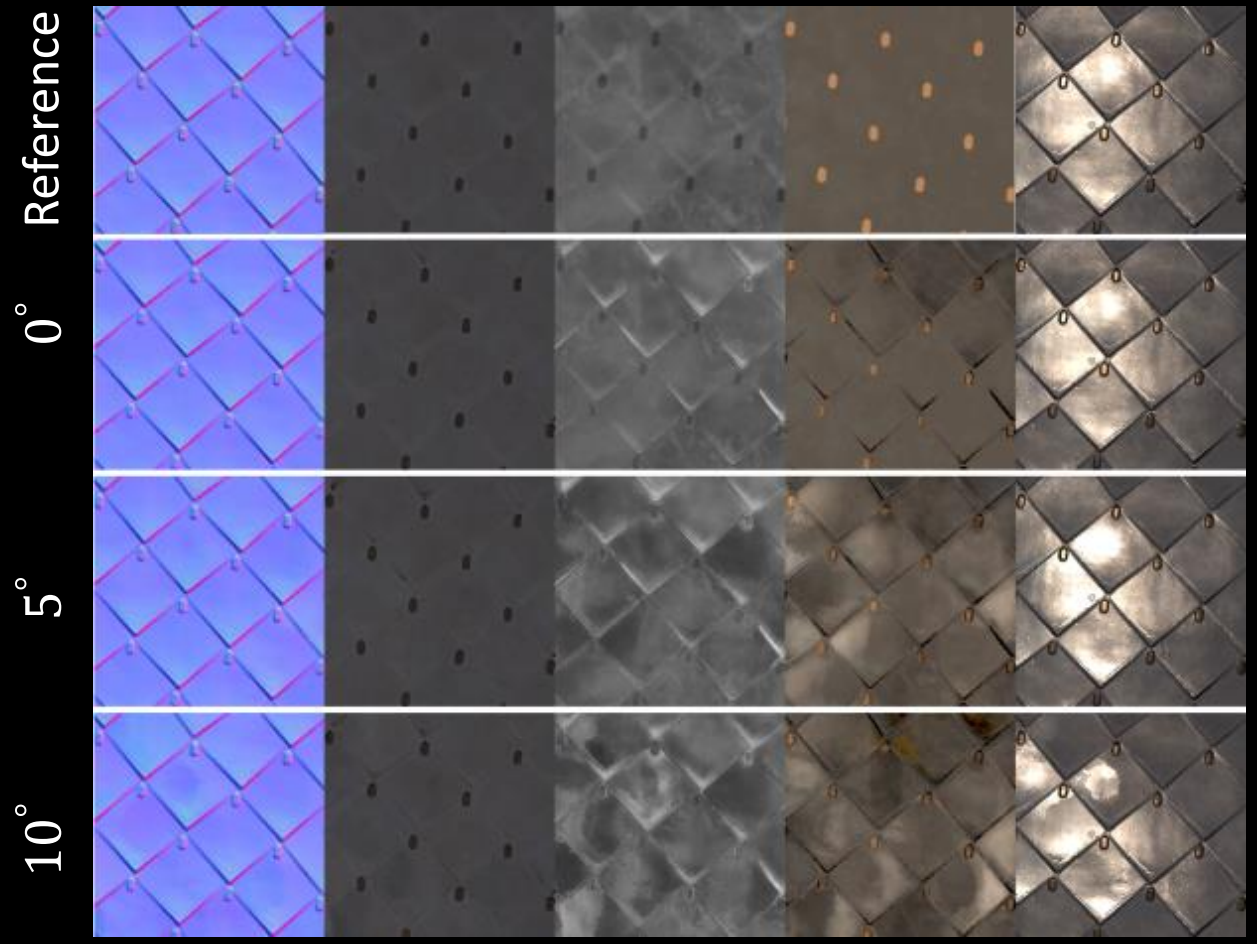
Novel view

Robustness

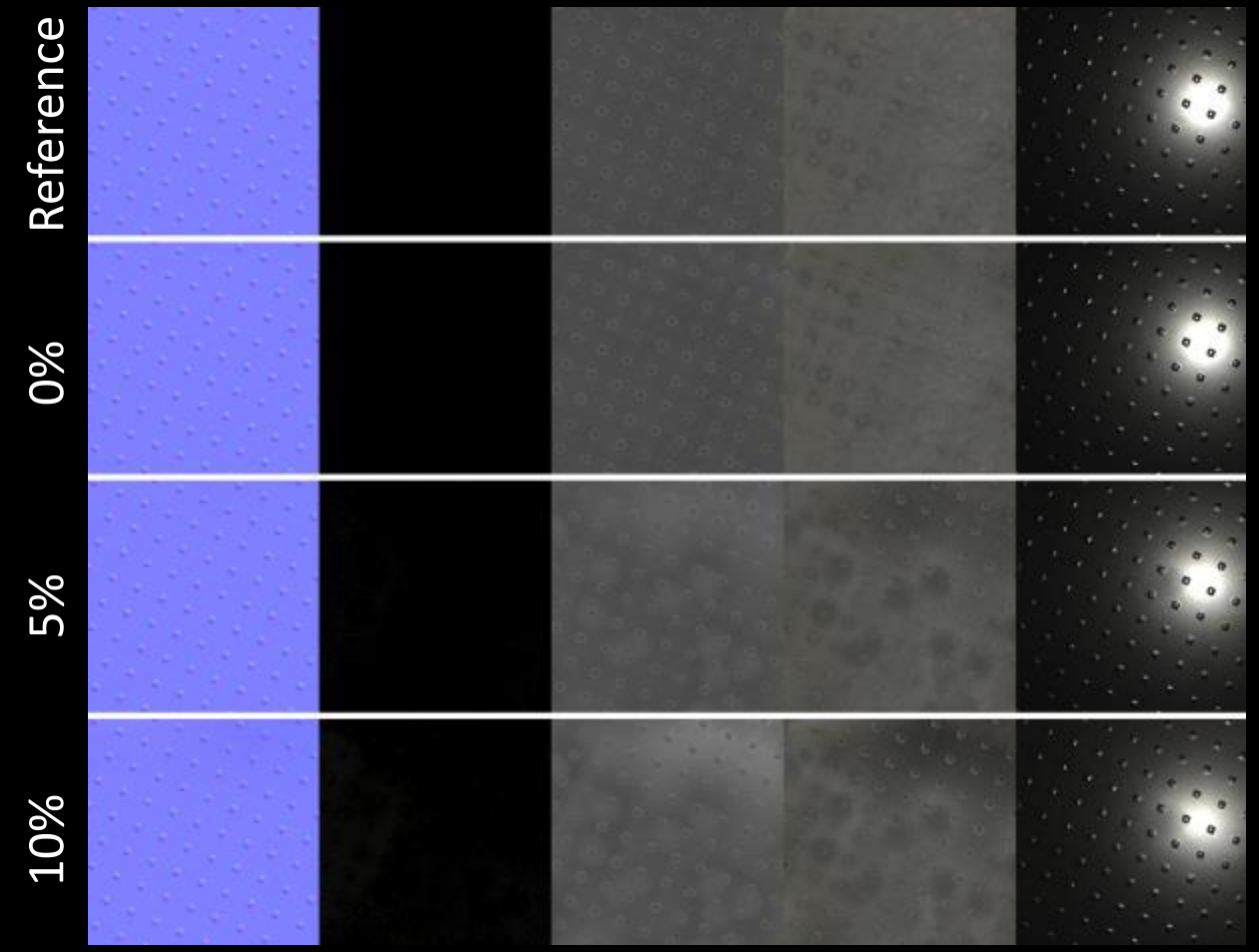




Robustness



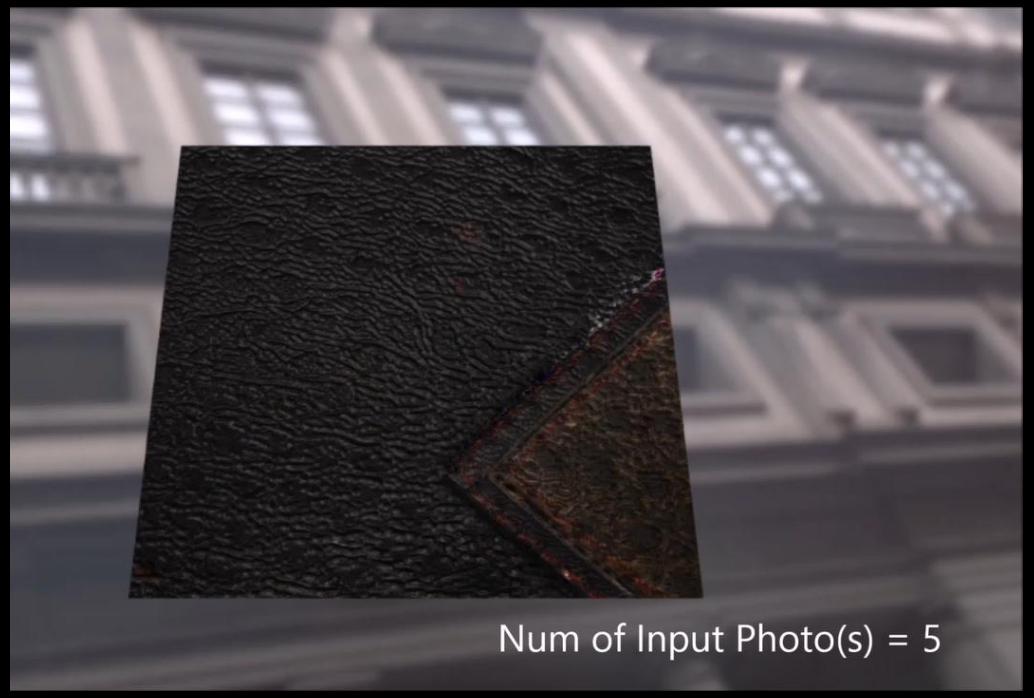
Robustness



Conclusion

- Unified deep inverse rendering framework for estimating SVBRDF from arbitrary number of input photographs.

- Learned latent space + optimization in latent space



Thanks