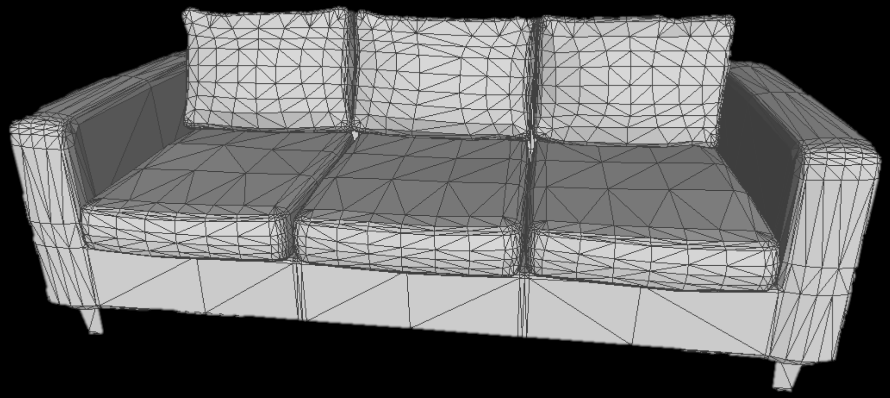
Deep Inverse Rendering for High-resolution SVBRDF Estimation from an Arbitrary Number of Images

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

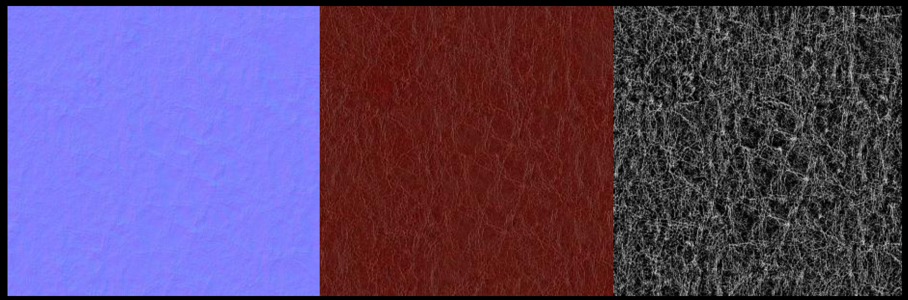
[1] Tsinghua University
[2] University of Science and Technology of China
[3] Microsoft Research Asia
[4] College of William & Mary

RENDERING

MATERIAL APPEARANCE

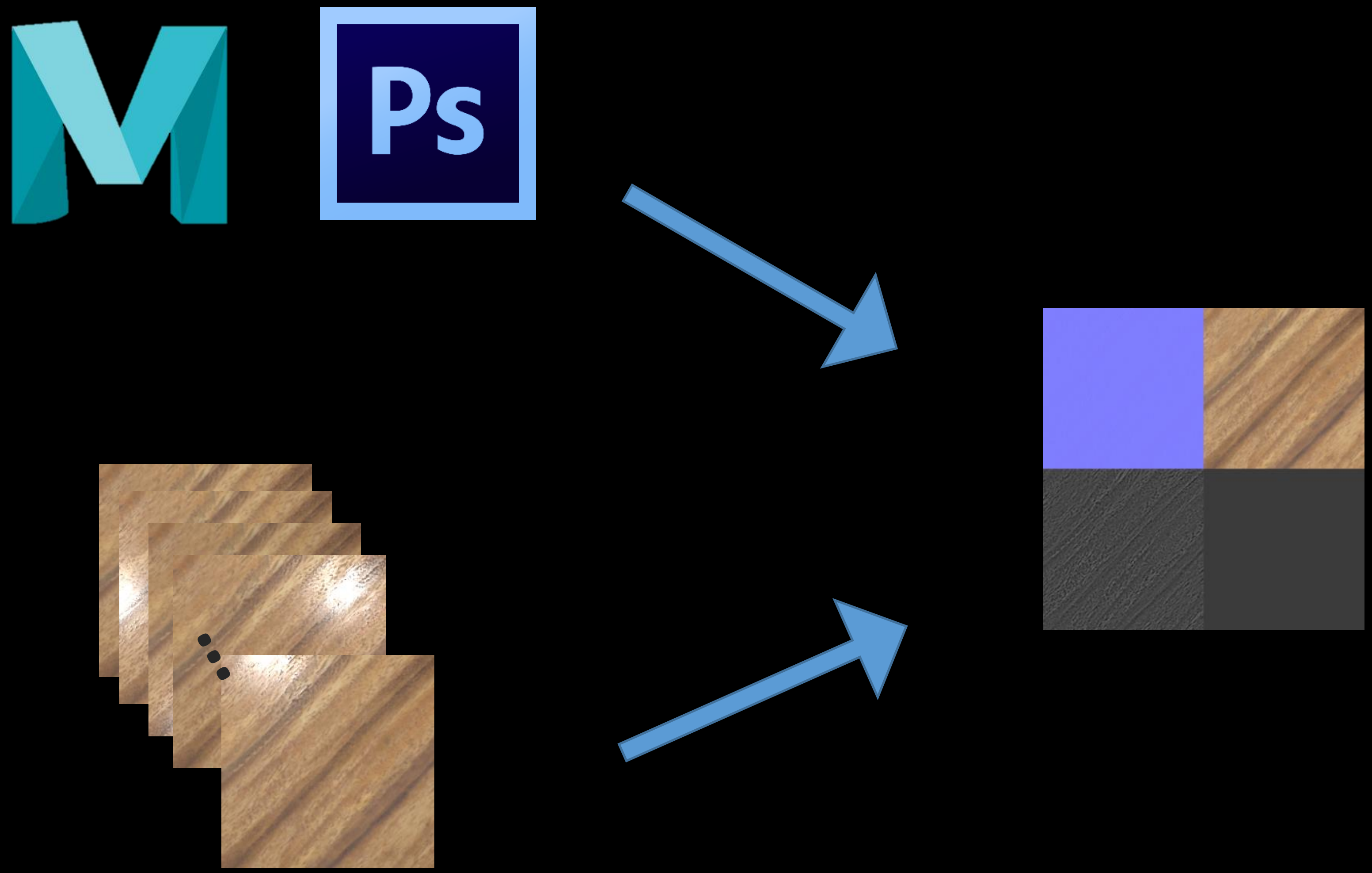


Geometry



Material

APPEARANCE ESTIMATION



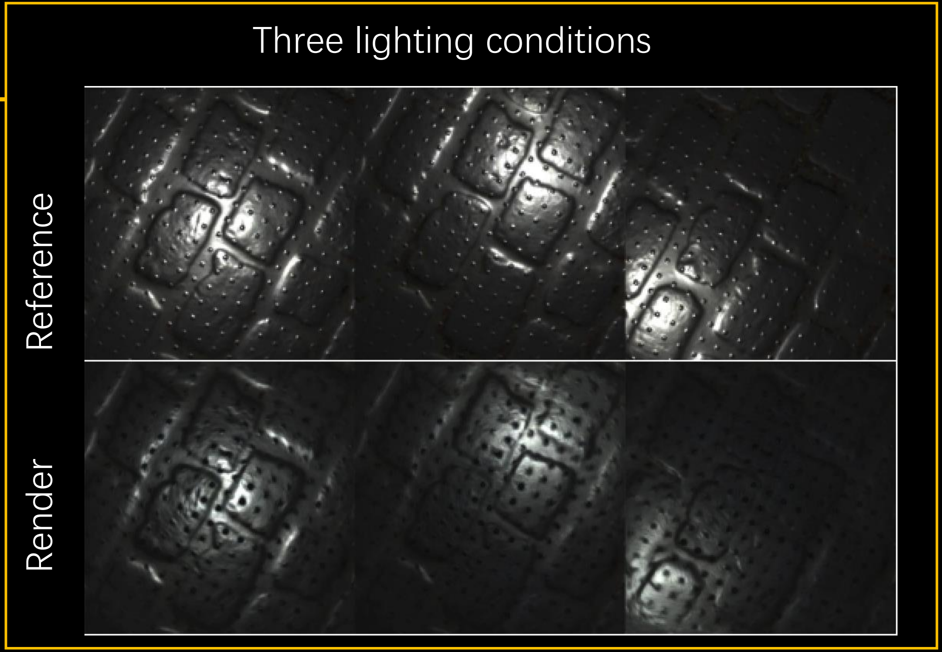
OUR GOAL

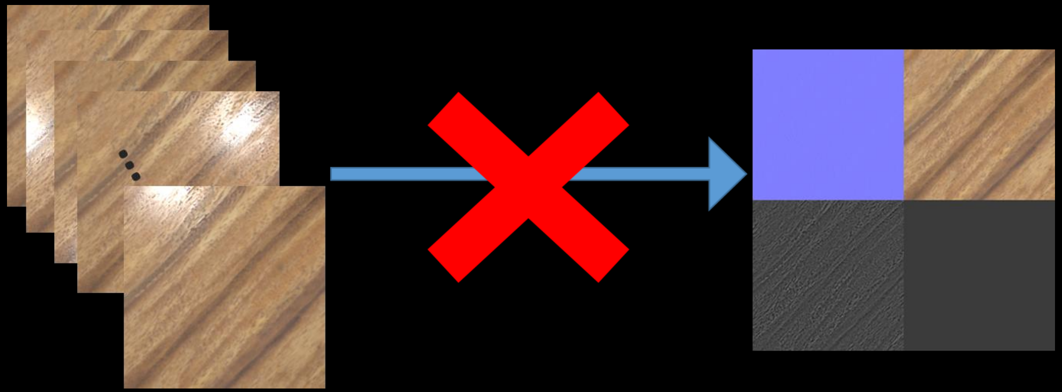
RELATED WORK

[Deschaintre et al. 2018]

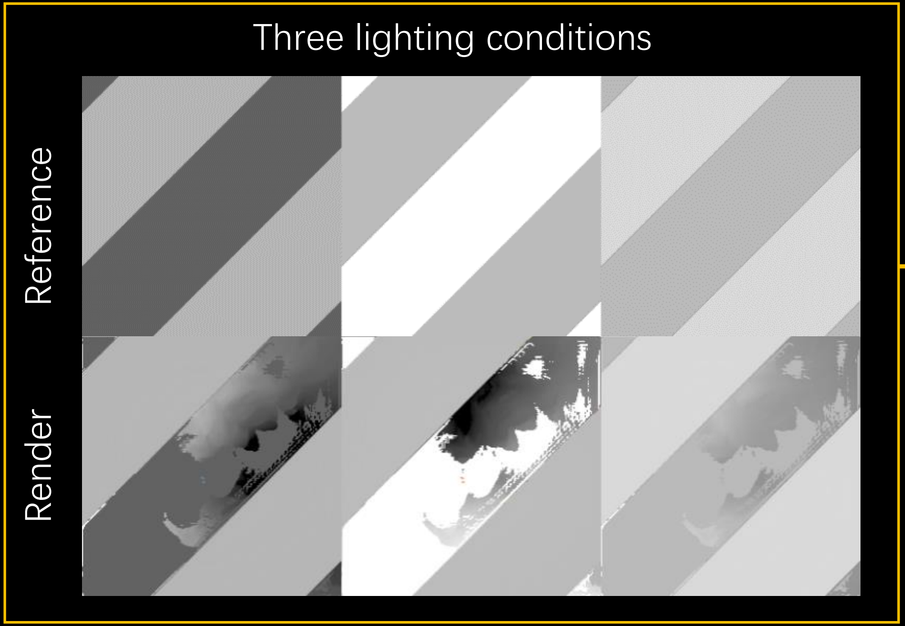
Learning-based methods

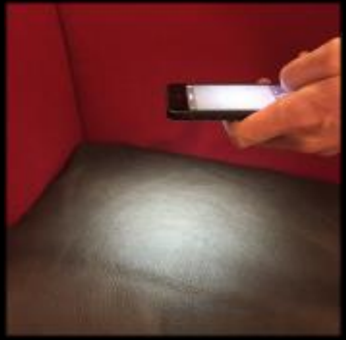
- Single input image
- Plausible





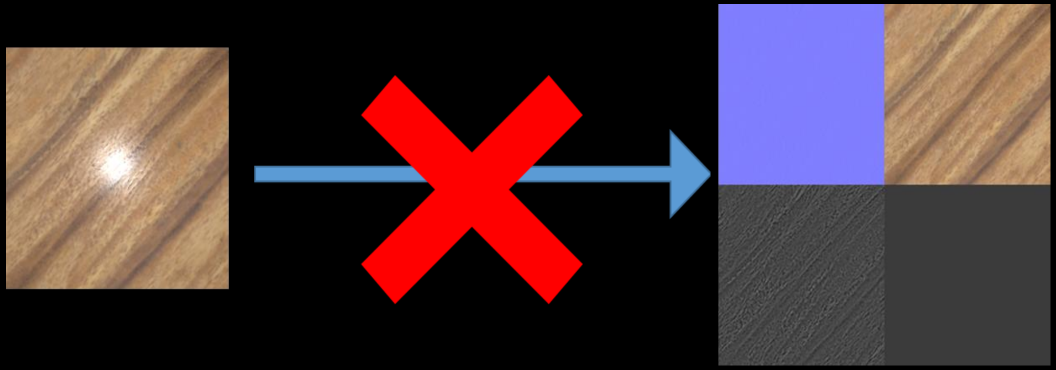
RELATED WORK





[Aittala et al. 2015]

[Dong et al. 2014]

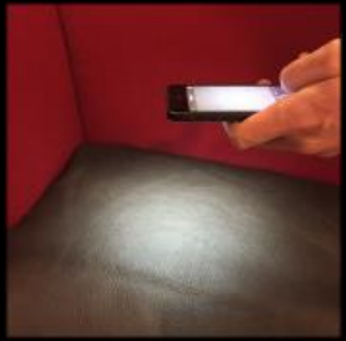


Classic Inverse Rendering

- Many input images
  (or strong assumptions)
- Accurate

RELATED WORK

[Deschaintre et al. 2018]



[Aittala et al. 2015]

[Dong et al. 2014]

single

few

many

Number of input images

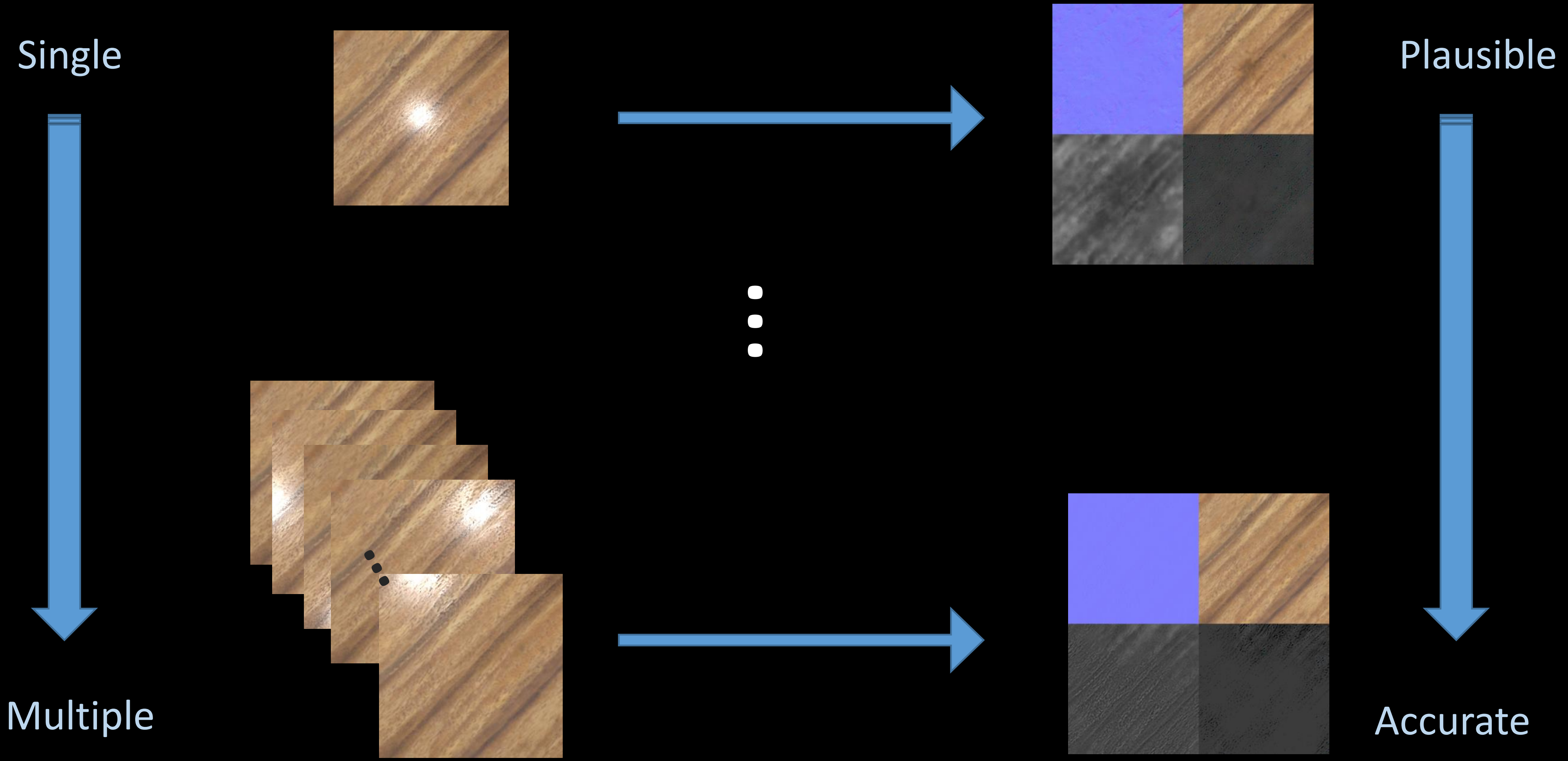
Learning-based methods

- Single input image
- Plausible

Classic Inverse Rendering

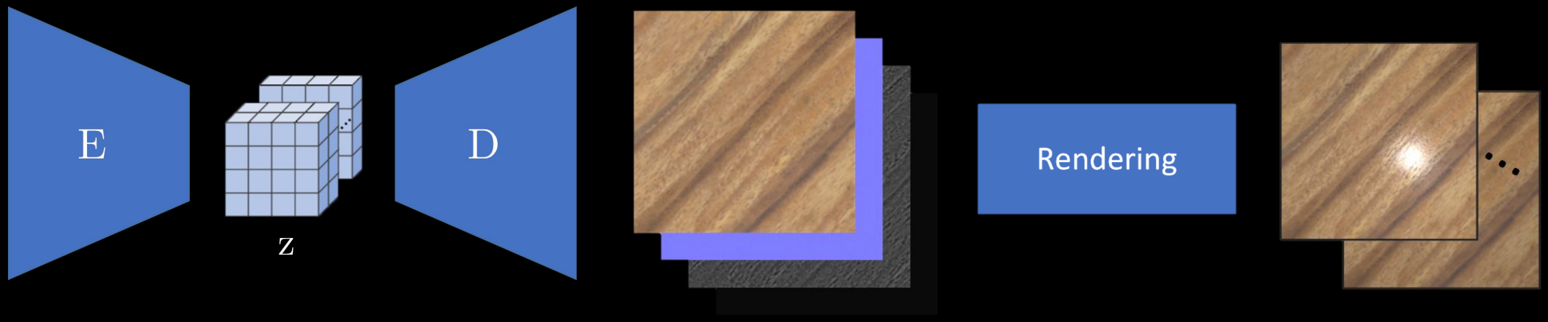
- Many input images
  (or strong assumptions)
- Accurate

OUR CONTRIBUTION



OUR METHOD

Key Idea: Deep Inverse Rendering



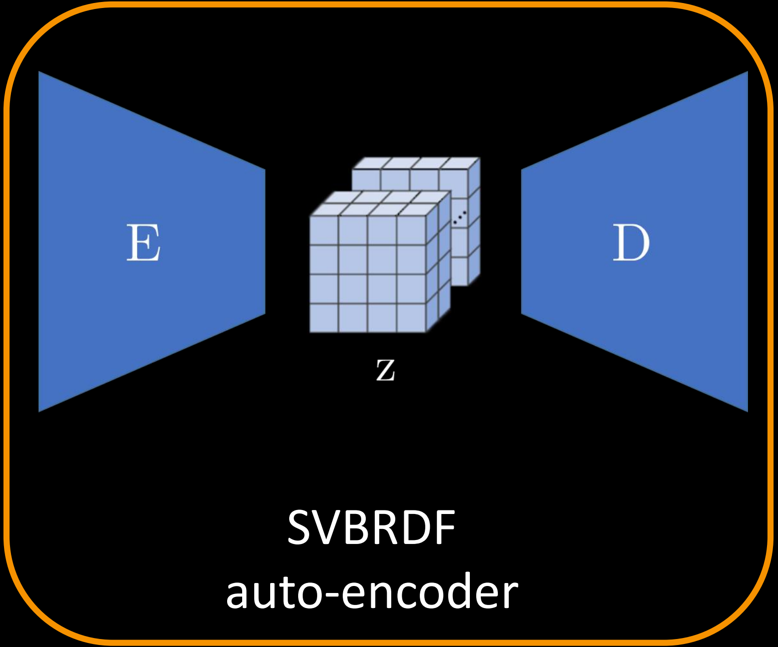
SVBRDF auto-encoder

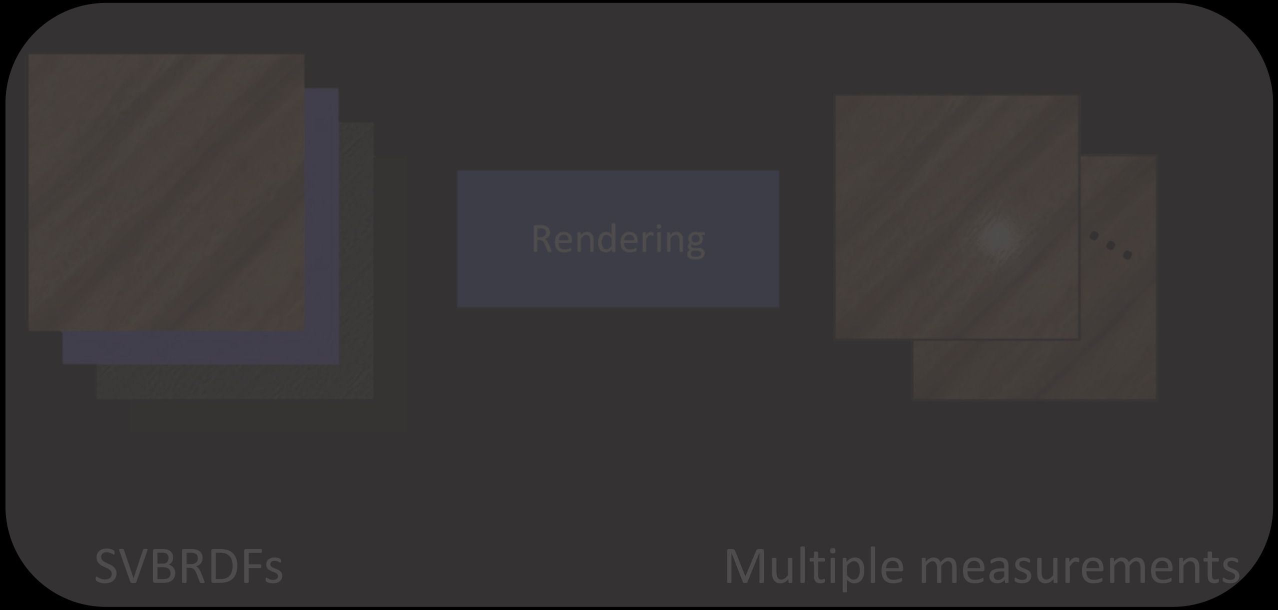
SVBRDFs

Multiple measurements

OUR METHOD

Key Idea: Deep Inverse Rendering

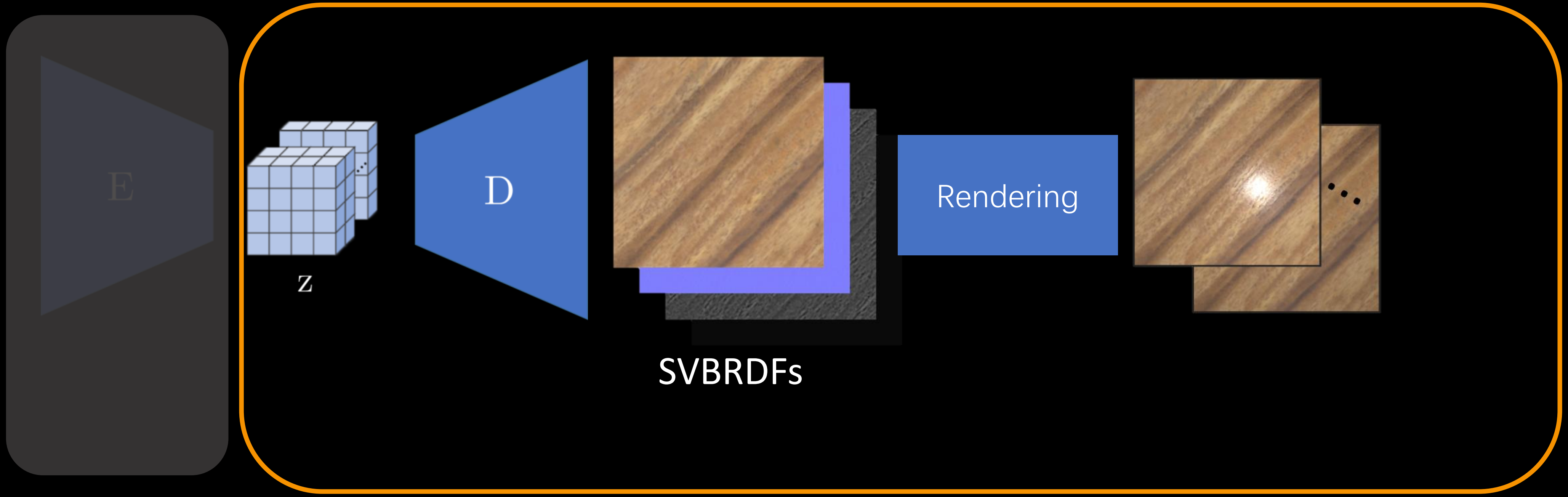




OUR METHOD

Key Idea: Deep Inverse Rendering

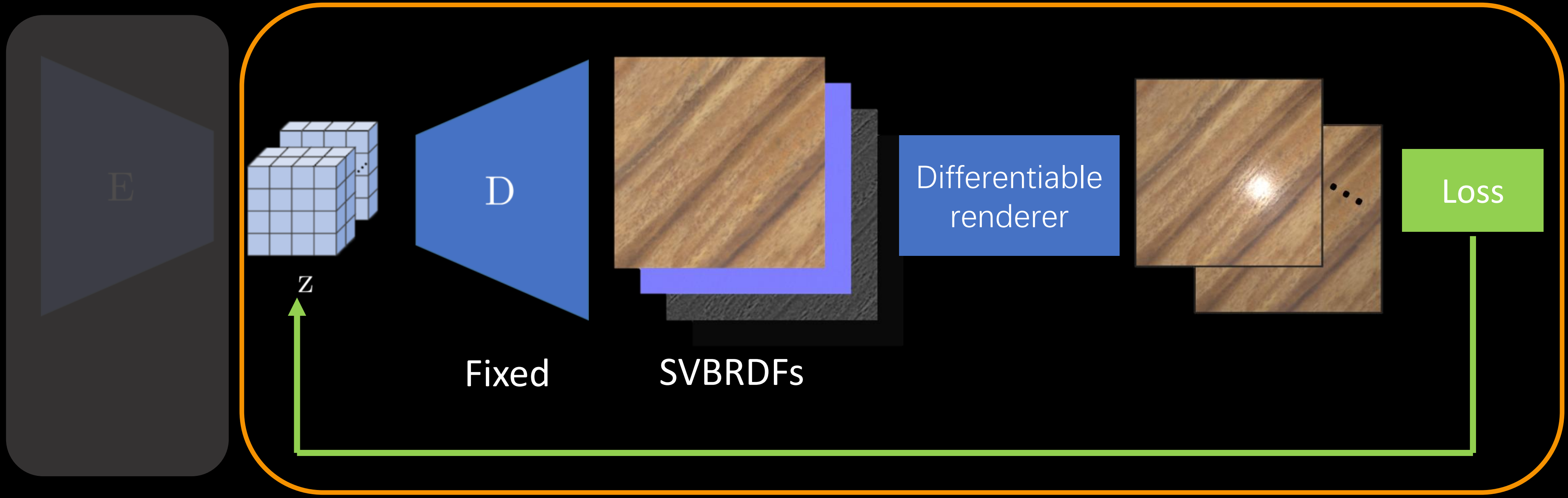
- Optimize in learned latent space



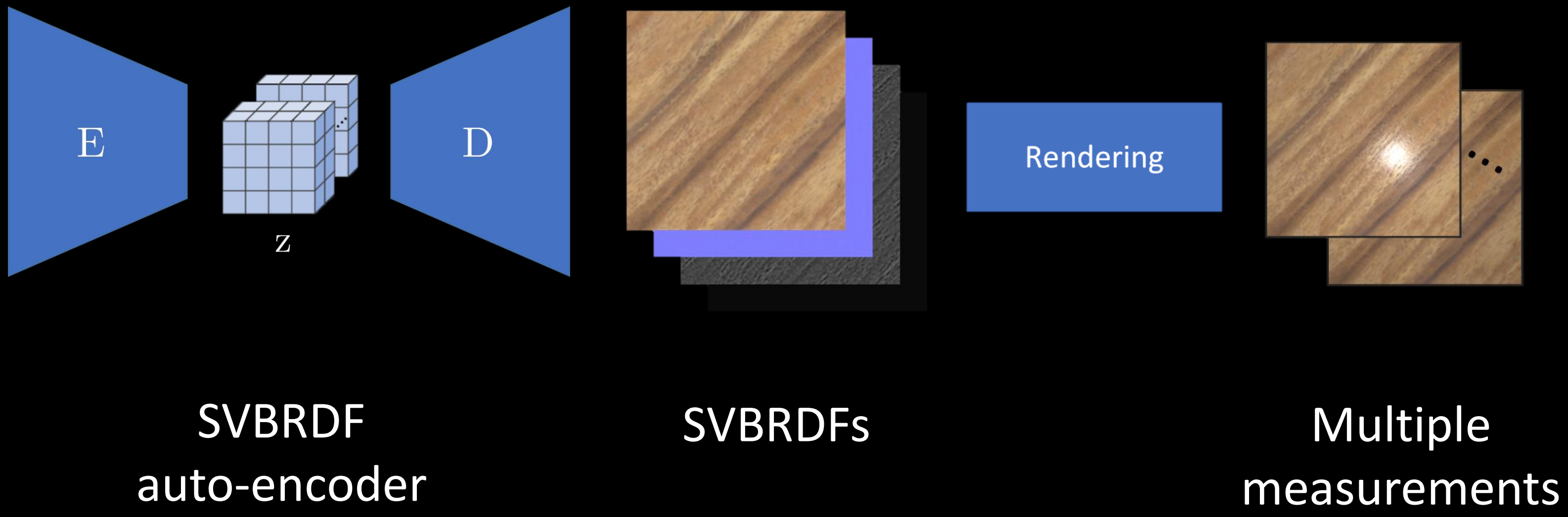
OUR METHOD

Key Idea: Deep Inverse Rendering

- Optimize in learned latent space



KEY CHALLENGES



KEY CHALLENGES

- **How to set correct error metric to preserve quality coherence of different maps?**

- **How to construct a smooth space suitable for optimization?**

- **How to get a good initialization?**

KEY CHALLENGES

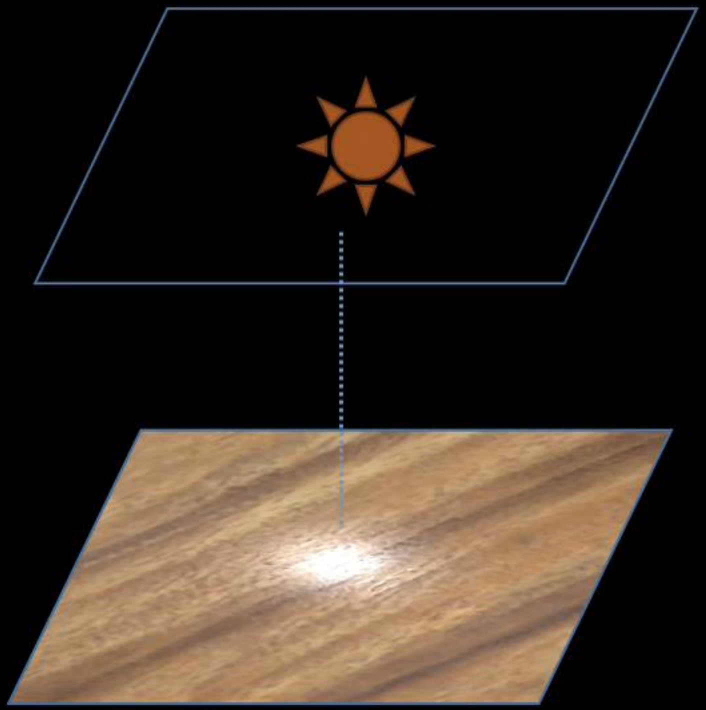
- Training Loss

- Smoothness regularization

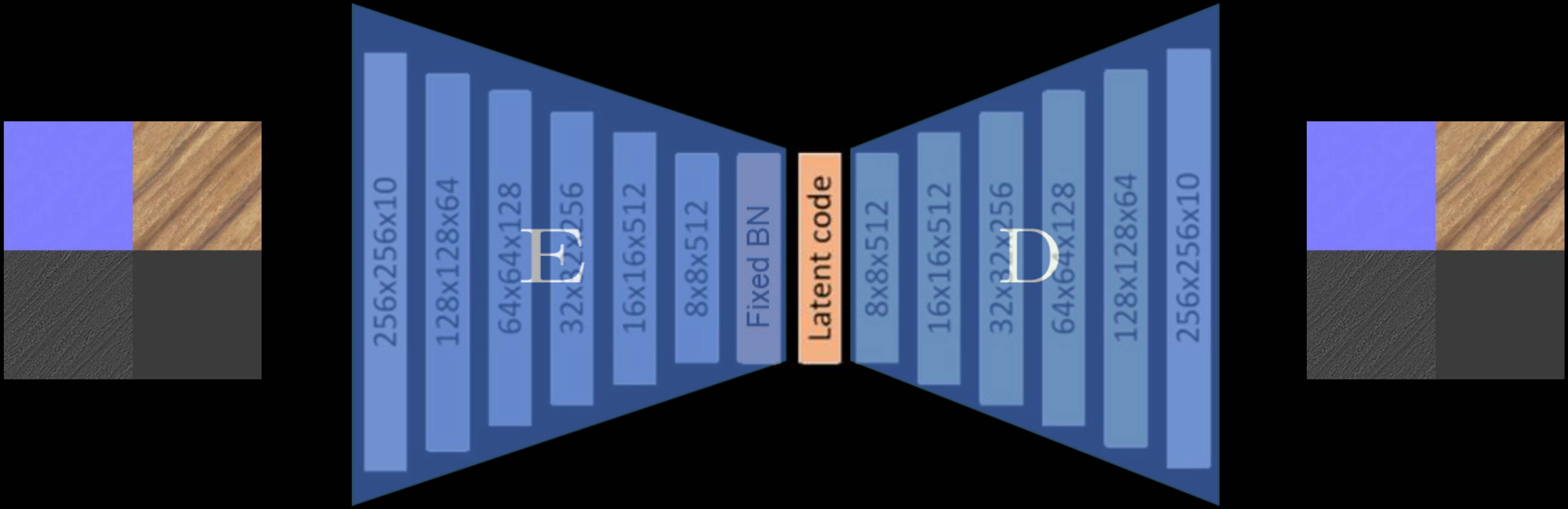
- Initialization strategy

ASSUMPTIONS

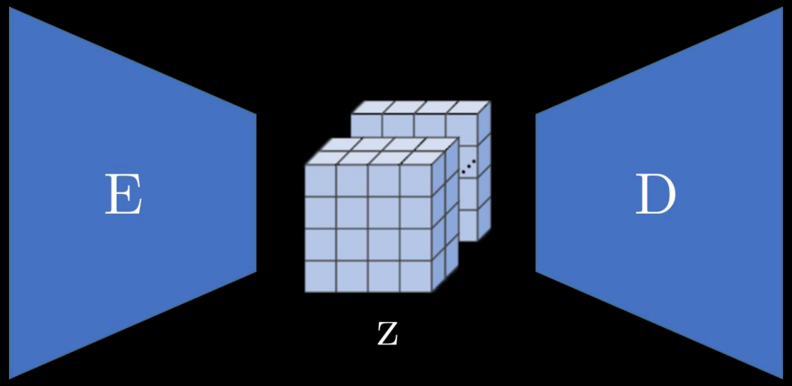
- Planar object
- Point light source collocated with the camera
- Fix distance between object plane and camera



SVBRDF AUTO-ENCODER



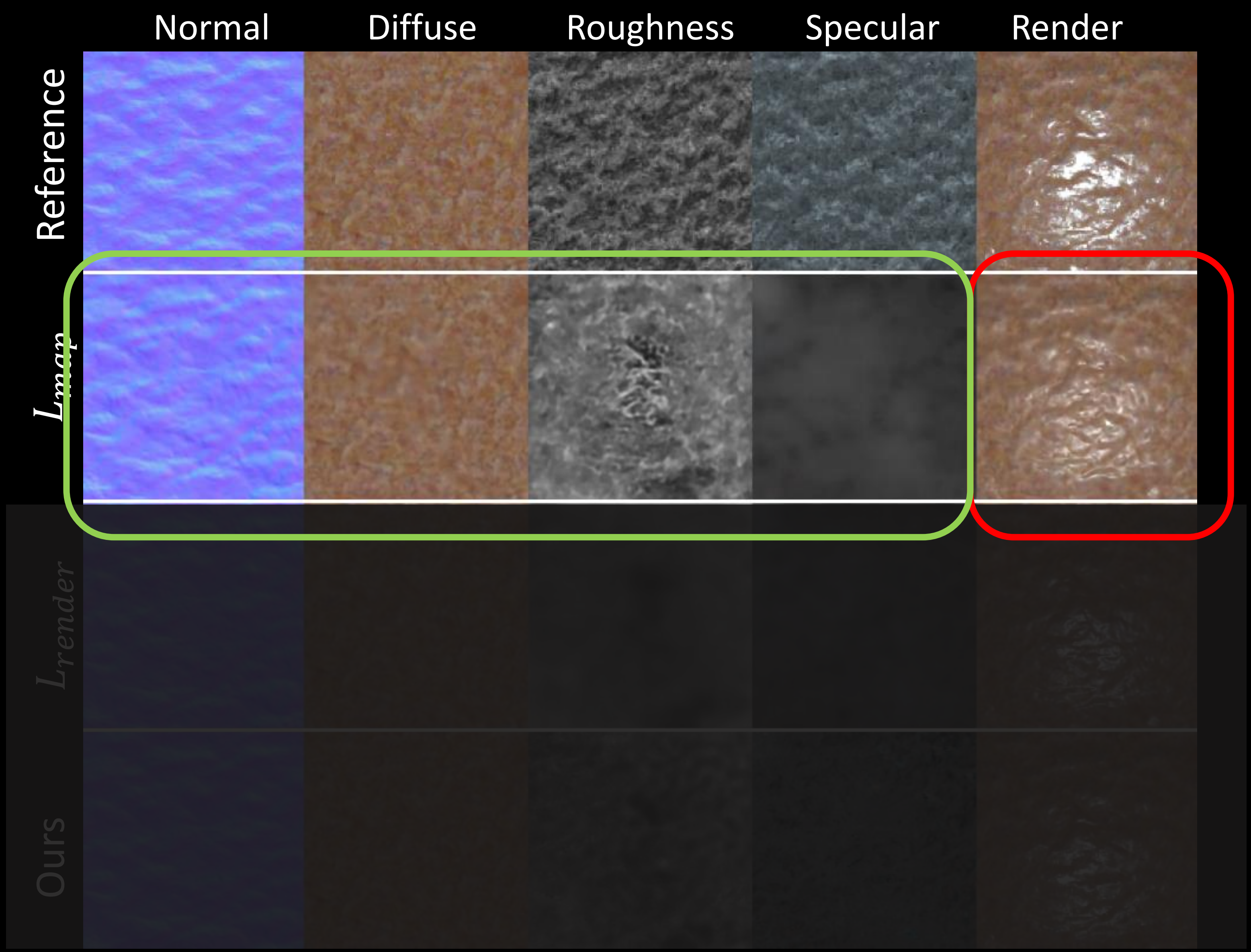
TRAINING SVBRDF AUTO-ENCODER



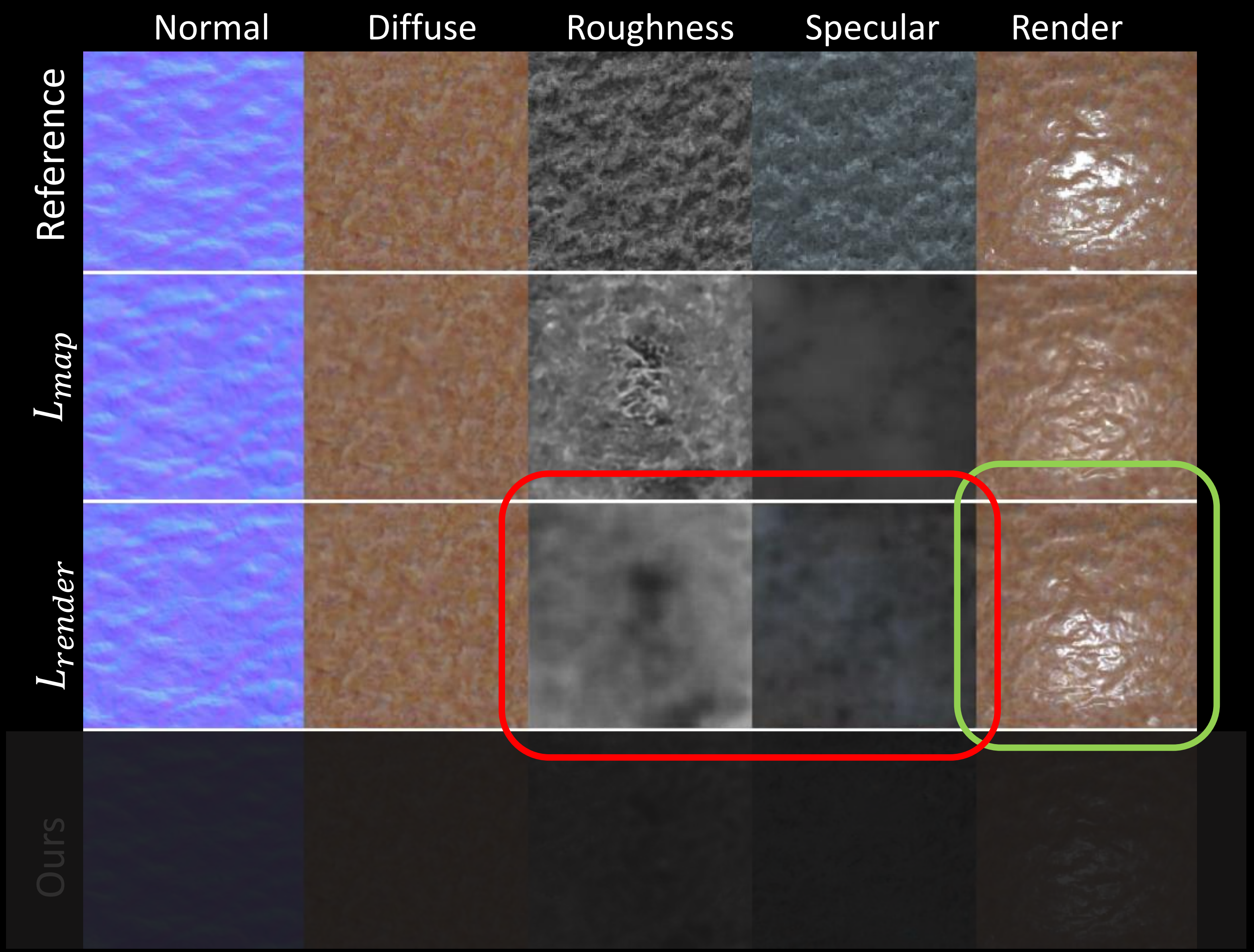
Training Loss:

$$\mathcal{L}_{train} = \mathcal{L}_{map} + \lambda_{render} \mathcal{L}_{render}$$

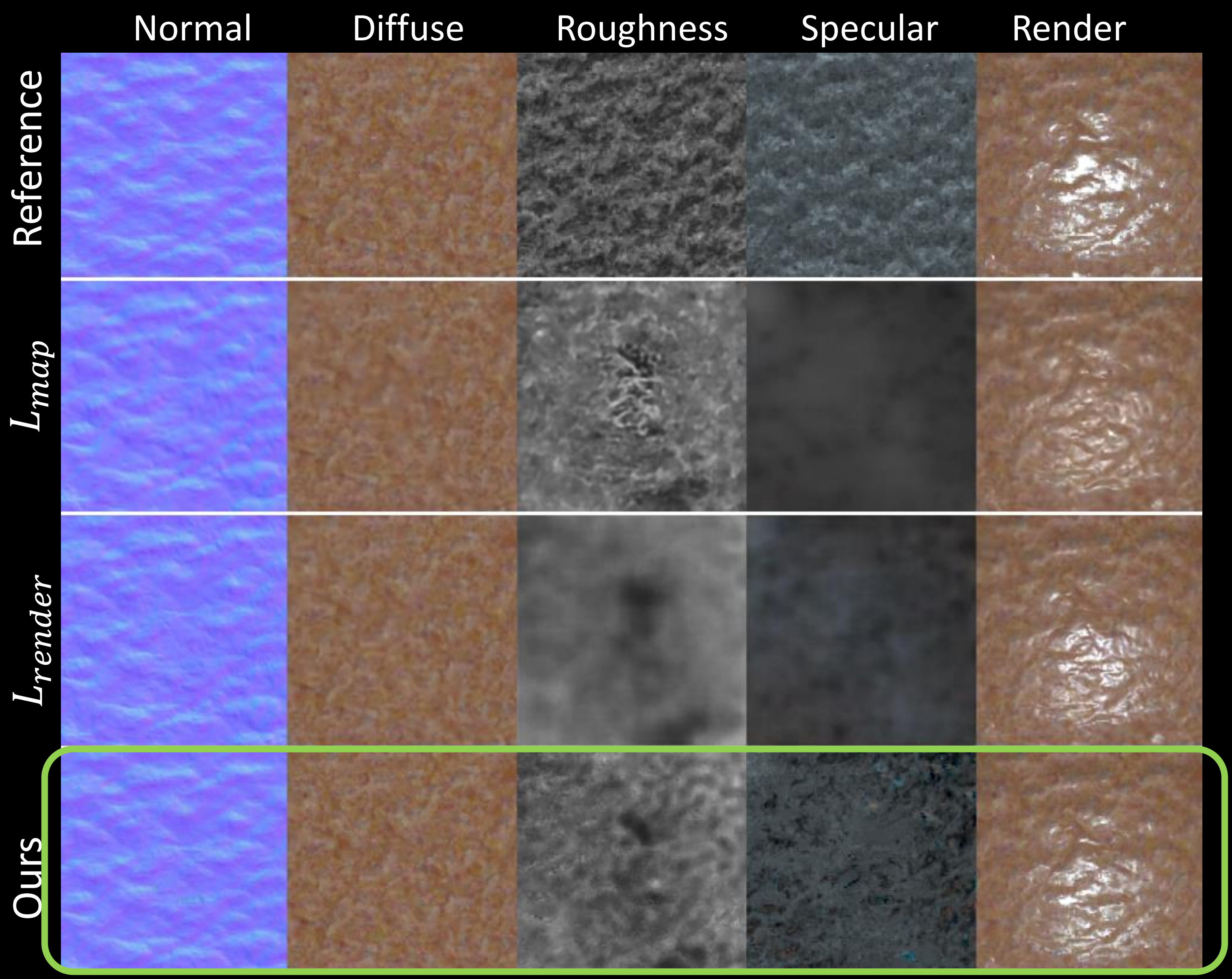
TRAINING SVBRDF AUTO-ENCODER



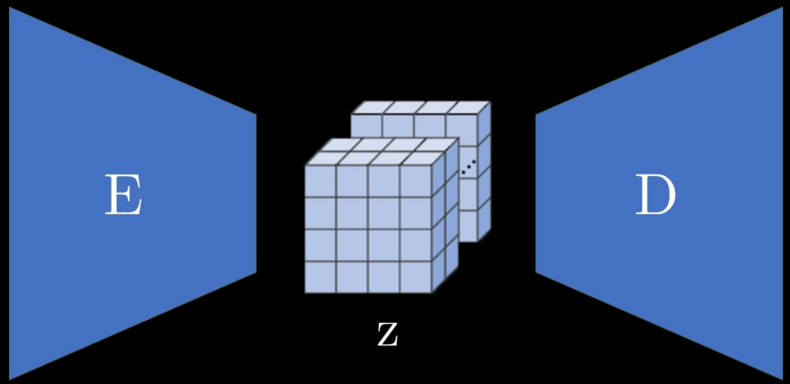
TRAINING SVBRDF AUTO-ENCODER



TRAINING SVBRDF AUTO-ENCODER



TRAINING SVBRDF AUTO-ENCODER



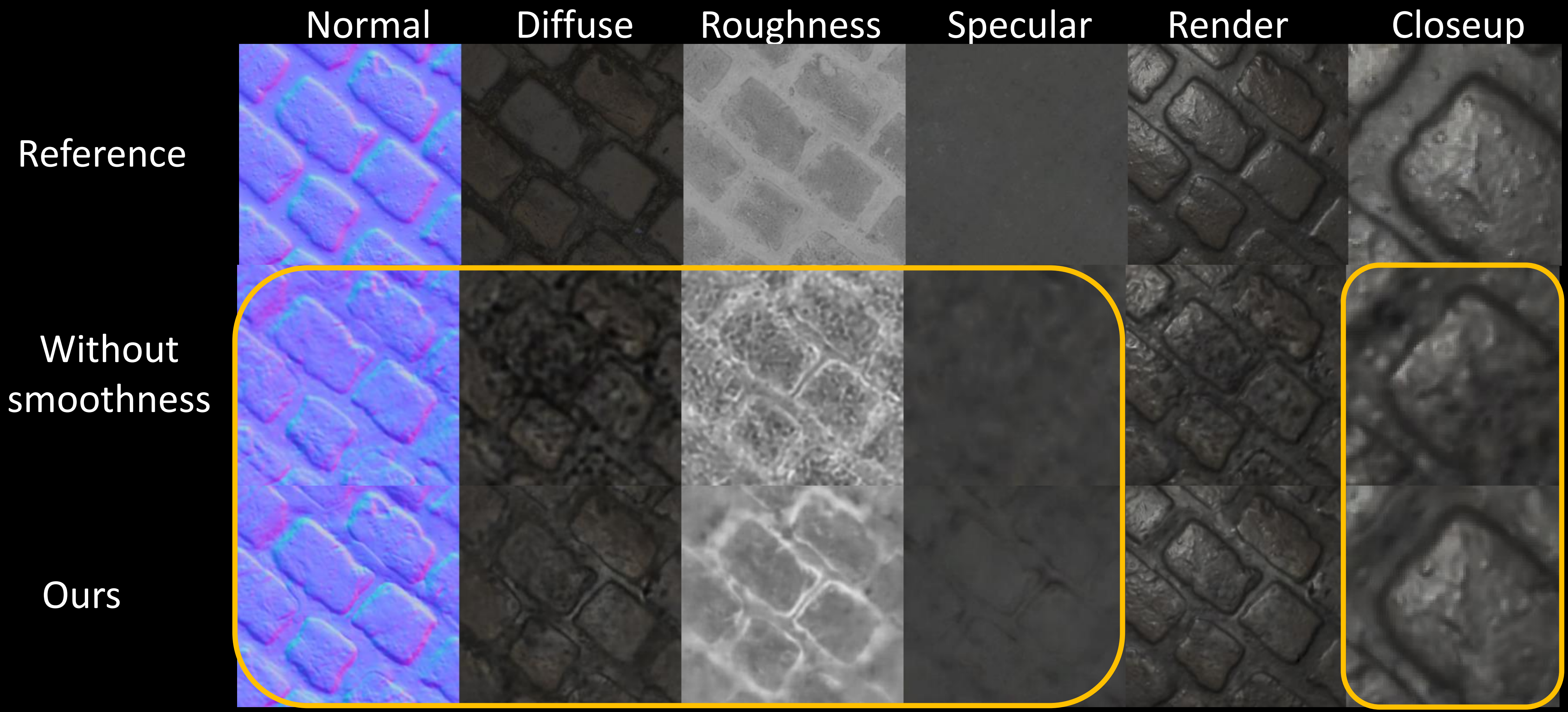
Training Loss:

$$\mathcal{L}_{train} = \mathcal{L}_{map} + \lambda_{render}\mathcal{L}_{render}$$

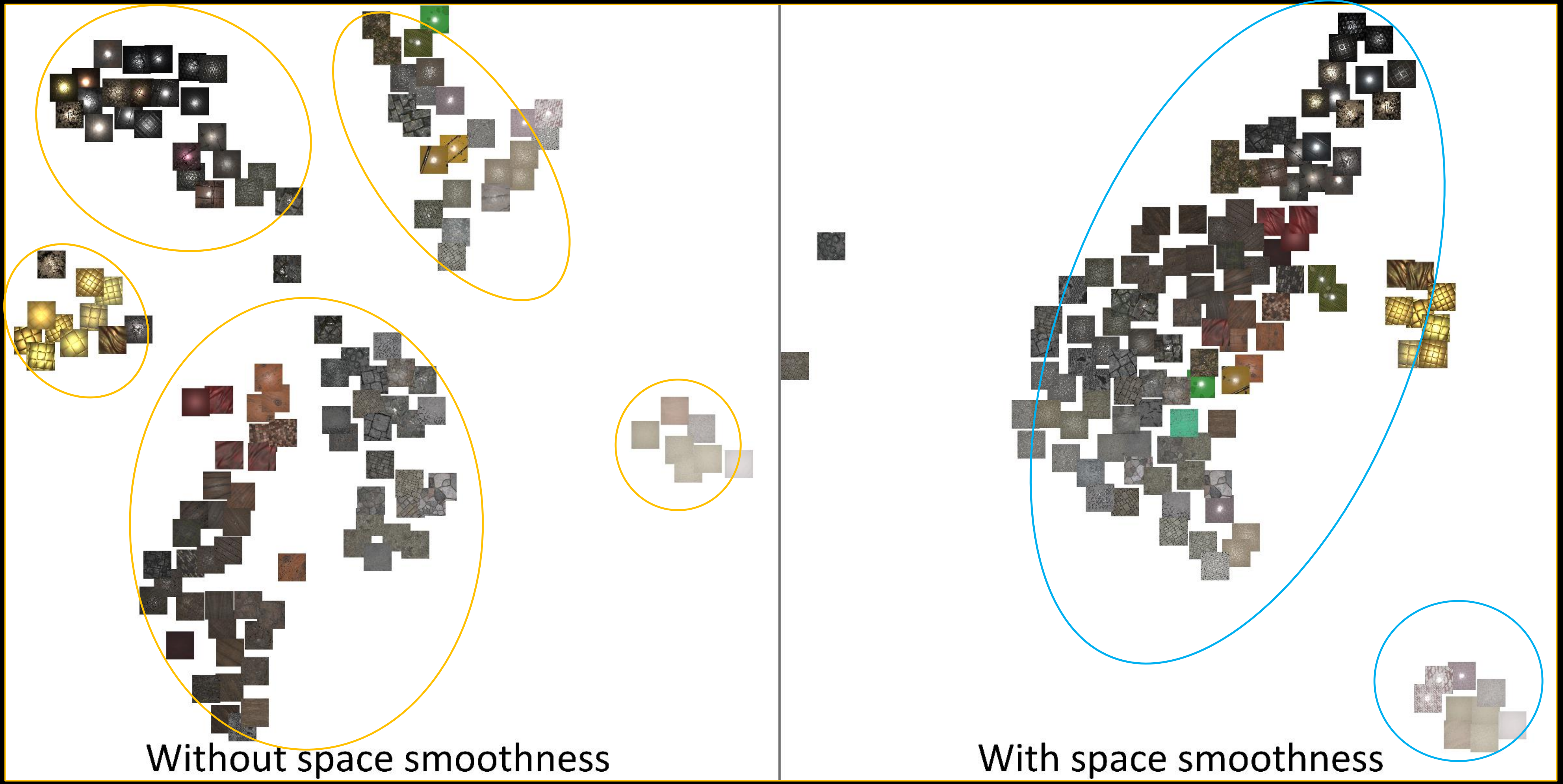
Latent space smoothness:

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

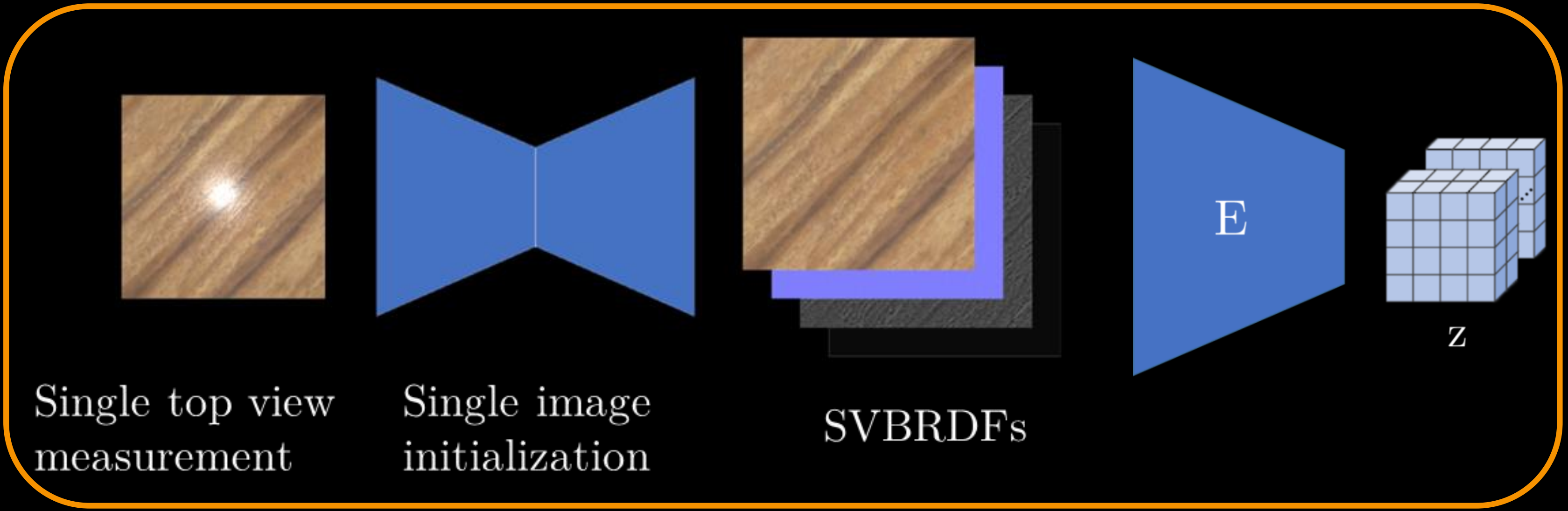
TRAINING SVBRDF AUTO-ENCODER



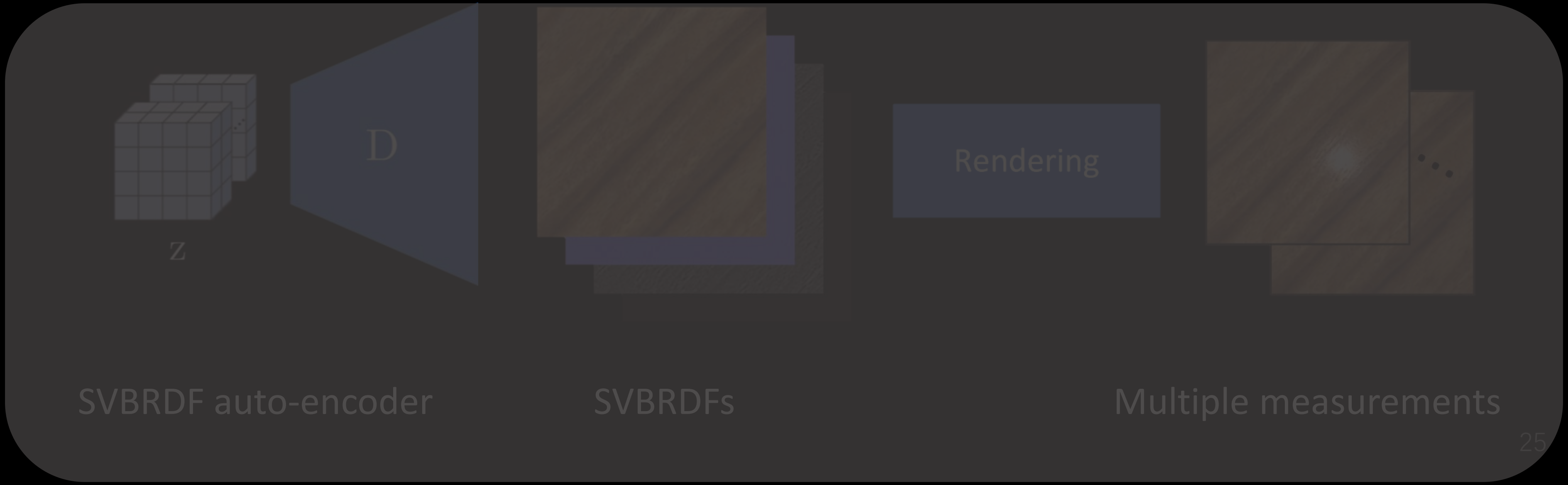
TRAINING SVBRDF AUTO-ENCODER



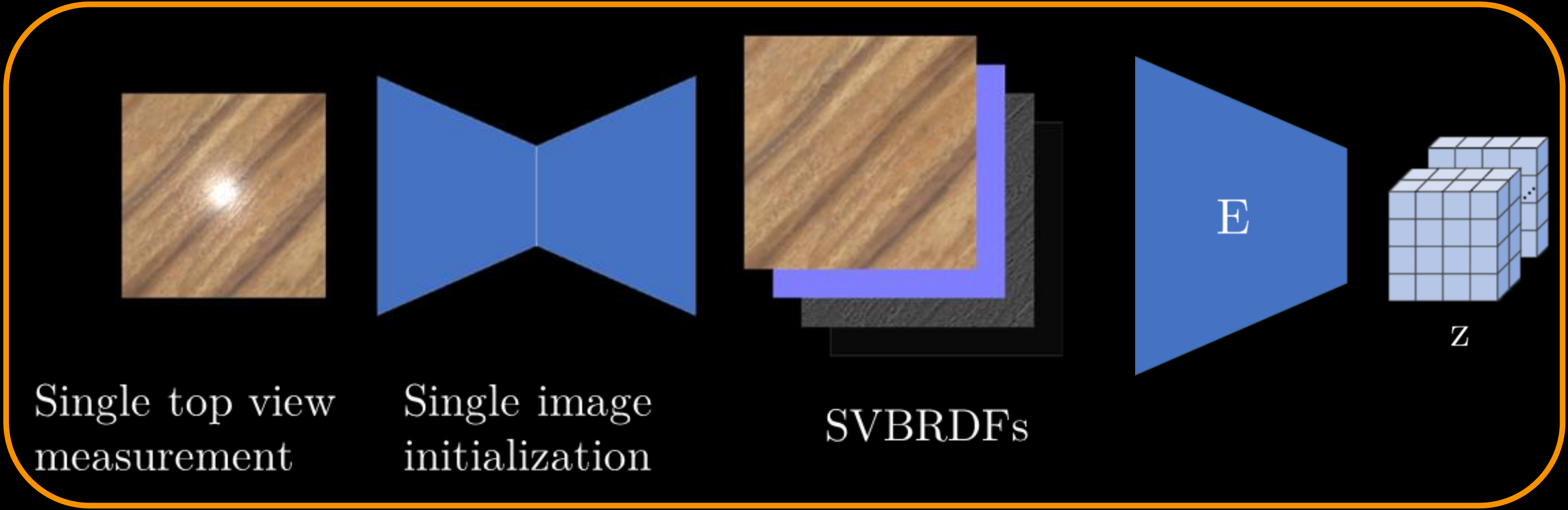
BOOTSTRAP THE OPTIMIZATION



State-of-the art single input network [Deschaintre et al. 2018]

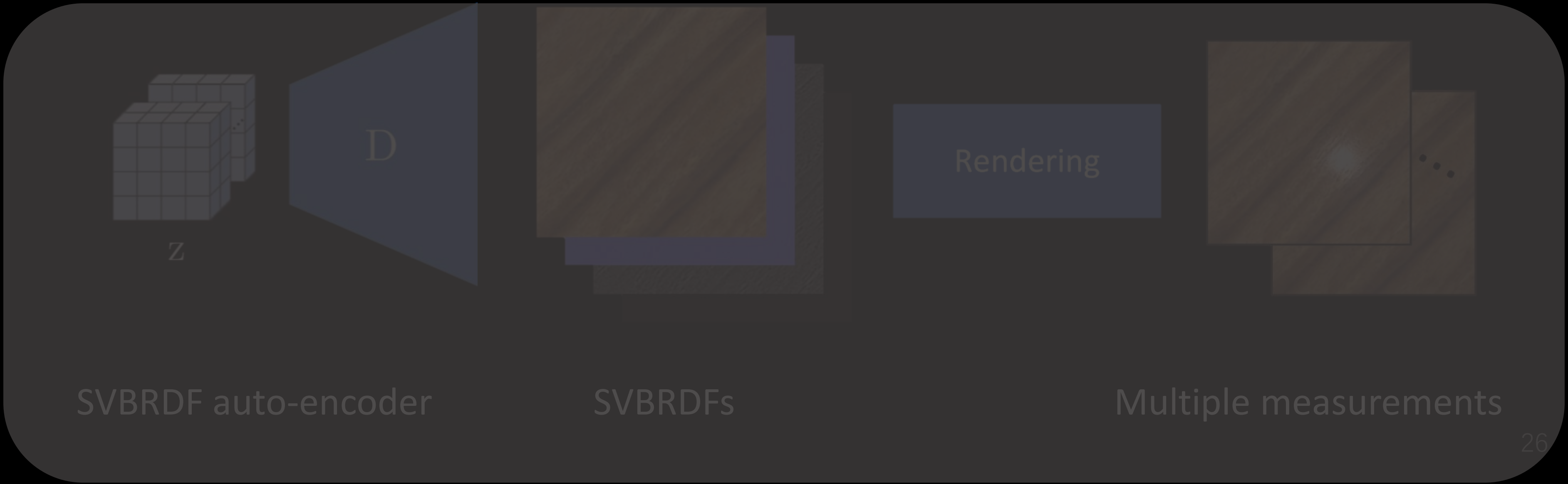


BOOTSTRAP THE OPTIMIZATION

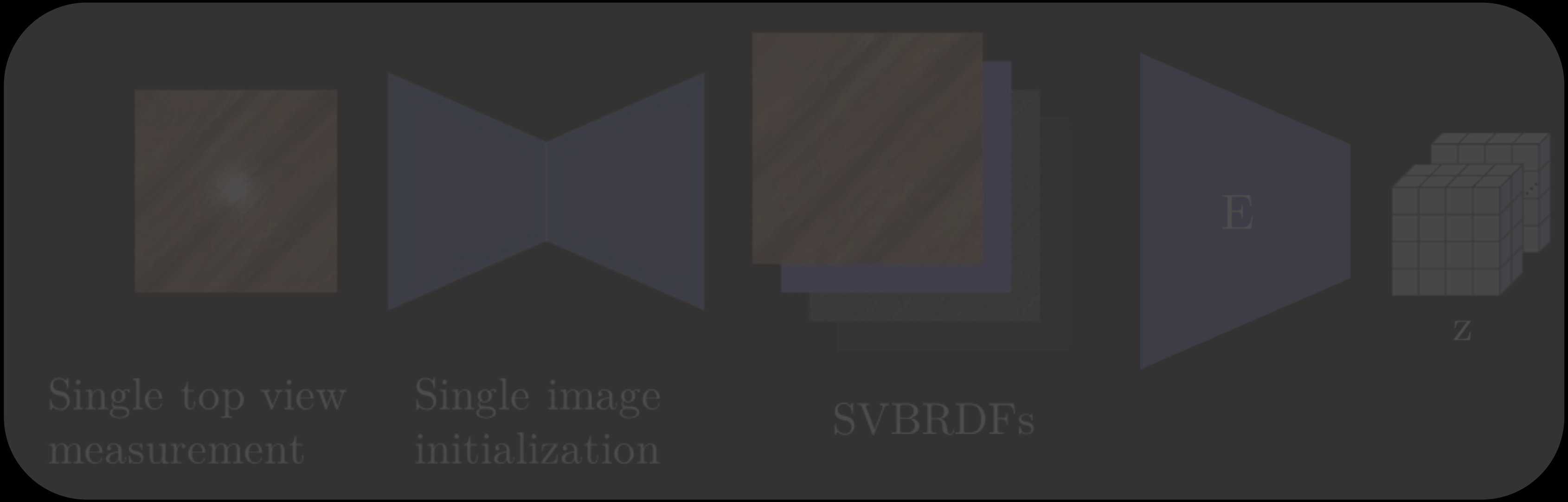


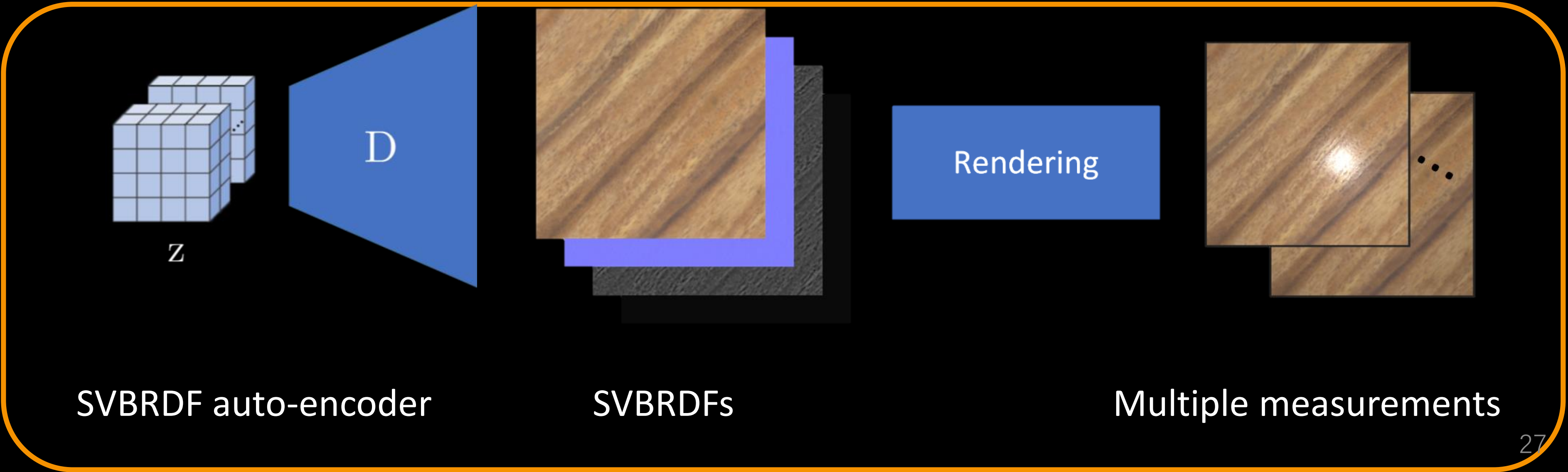
State-of-the art single input network [Deschaintre et al. 2018]

Or any other state-of-the art methods!

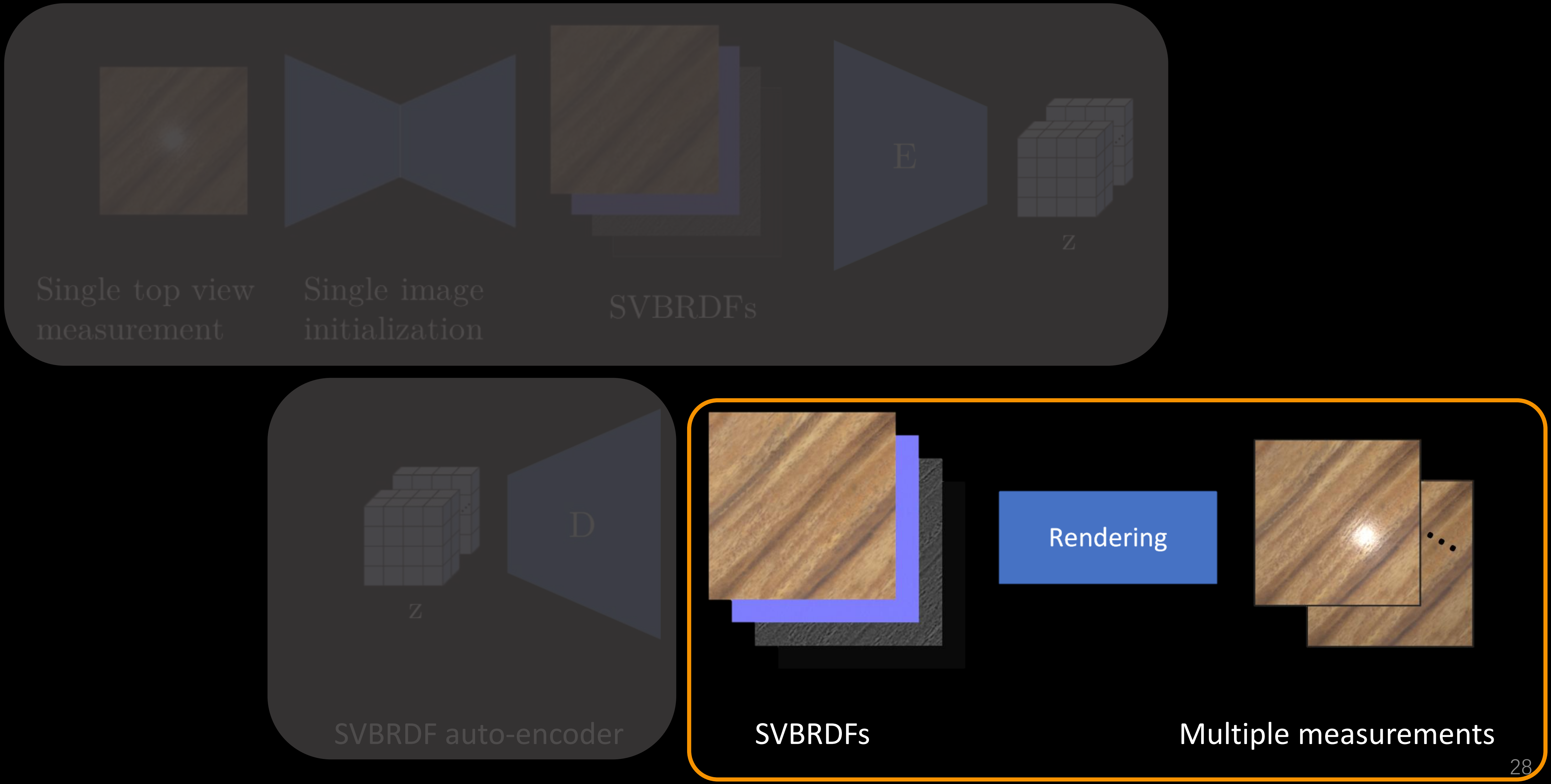


OPTIMIZE IN LATENT SPACE

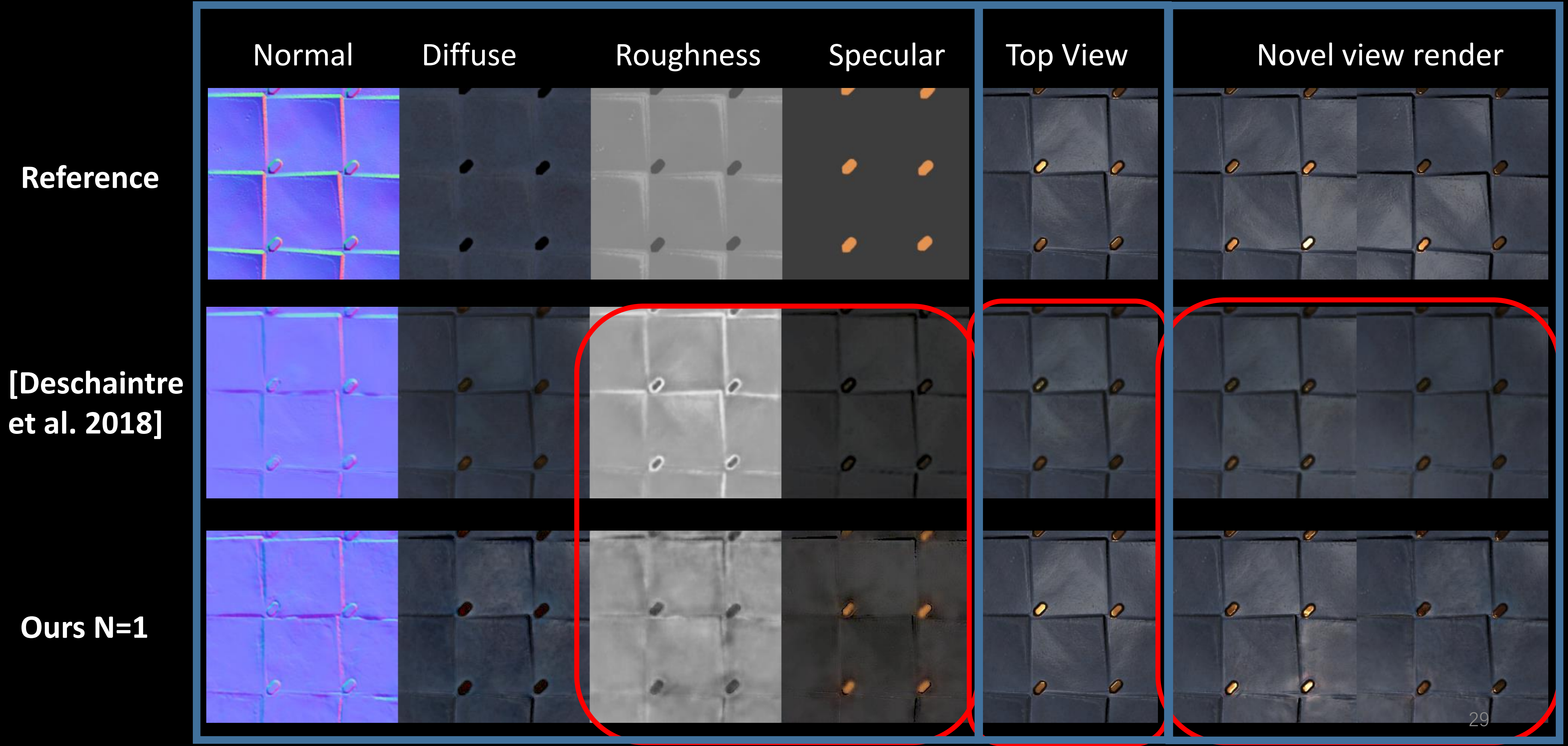


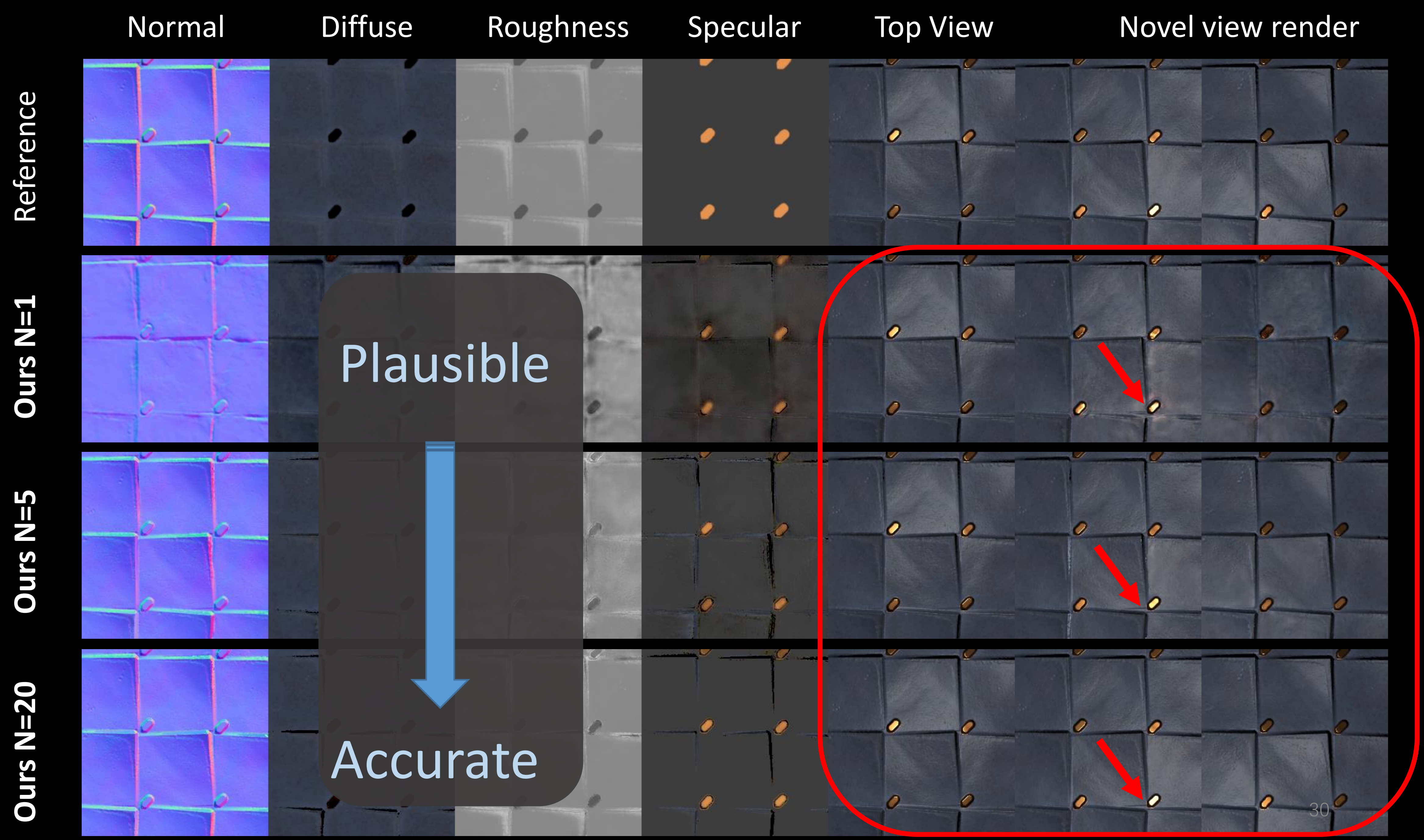


DETAIL REFINEMENT



IMPROVED QUALITY WITH SINGLE INPUT

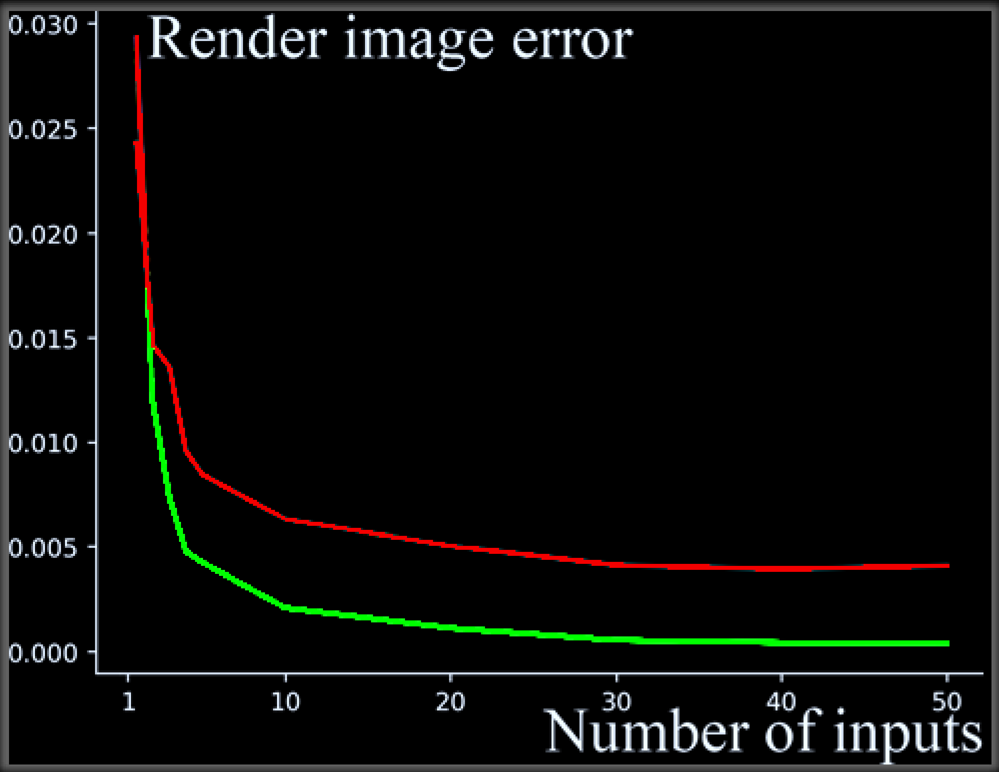


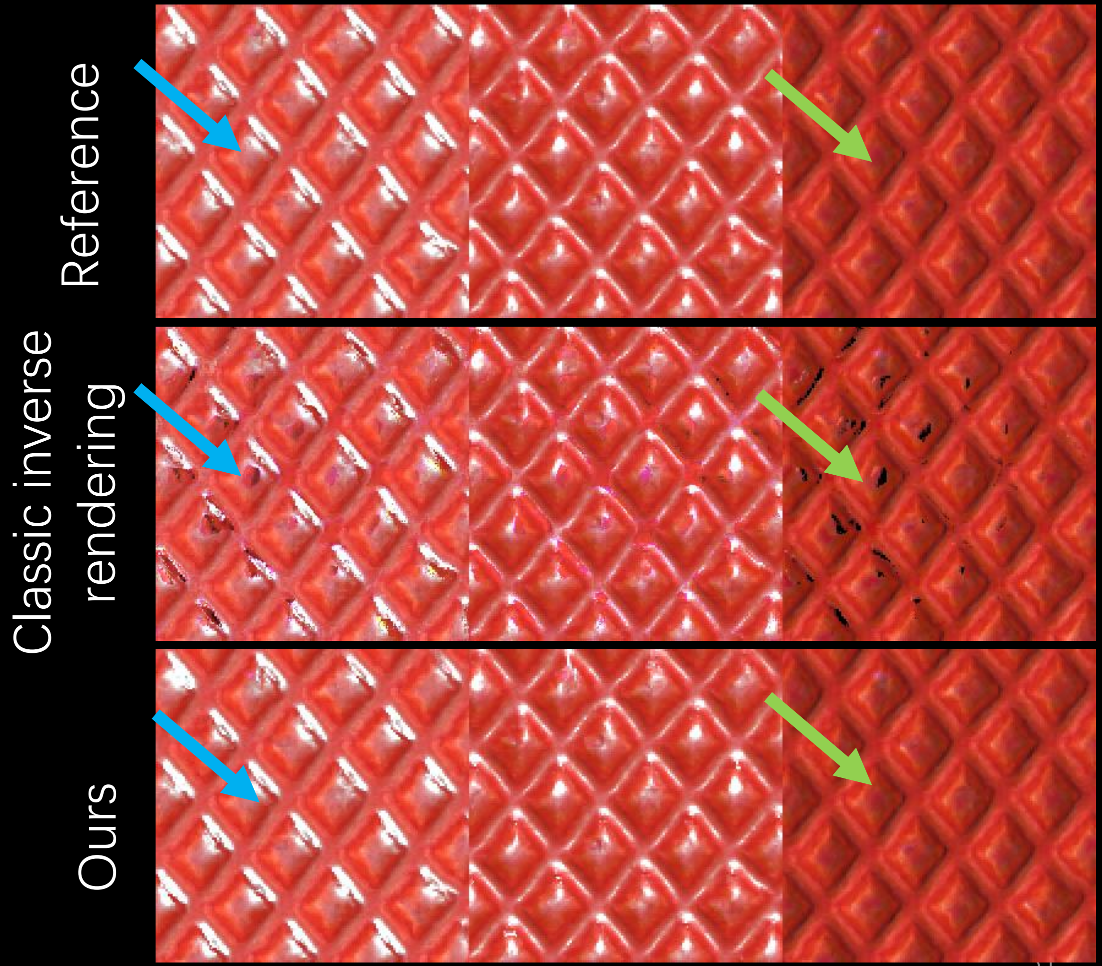


Normal
Diffuse
Roughness
Specular
Top View
Novel view render
Reference
Ours N=1
Ours N=5
Ours N=20
Plausible
Accurate

COMPARISON WITH CLASSIC INVERSE RENDERING

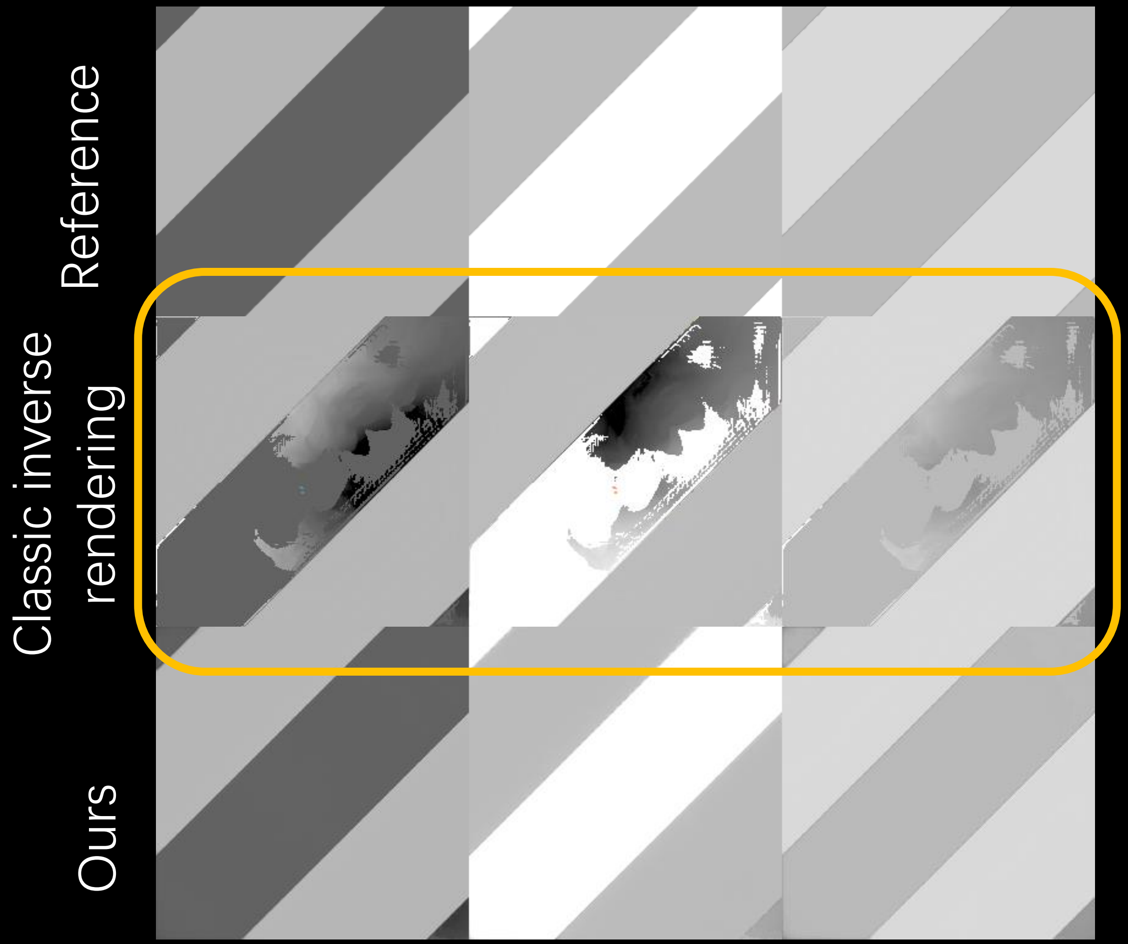
Classic inverse rendering ours





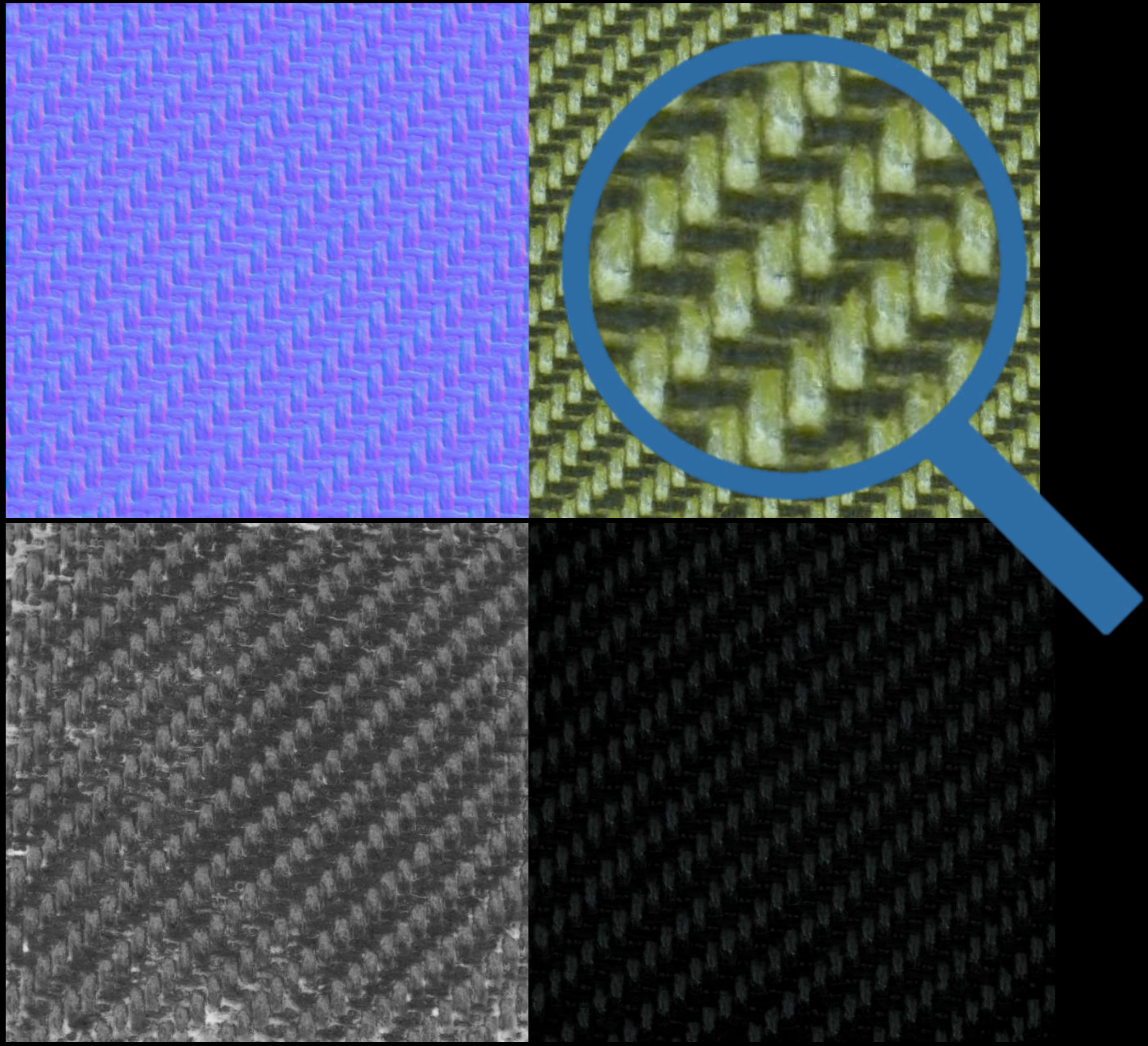
COMPARISON WITH CLASSIC INVERSE RENDERING

Classic inverse rendering   ours

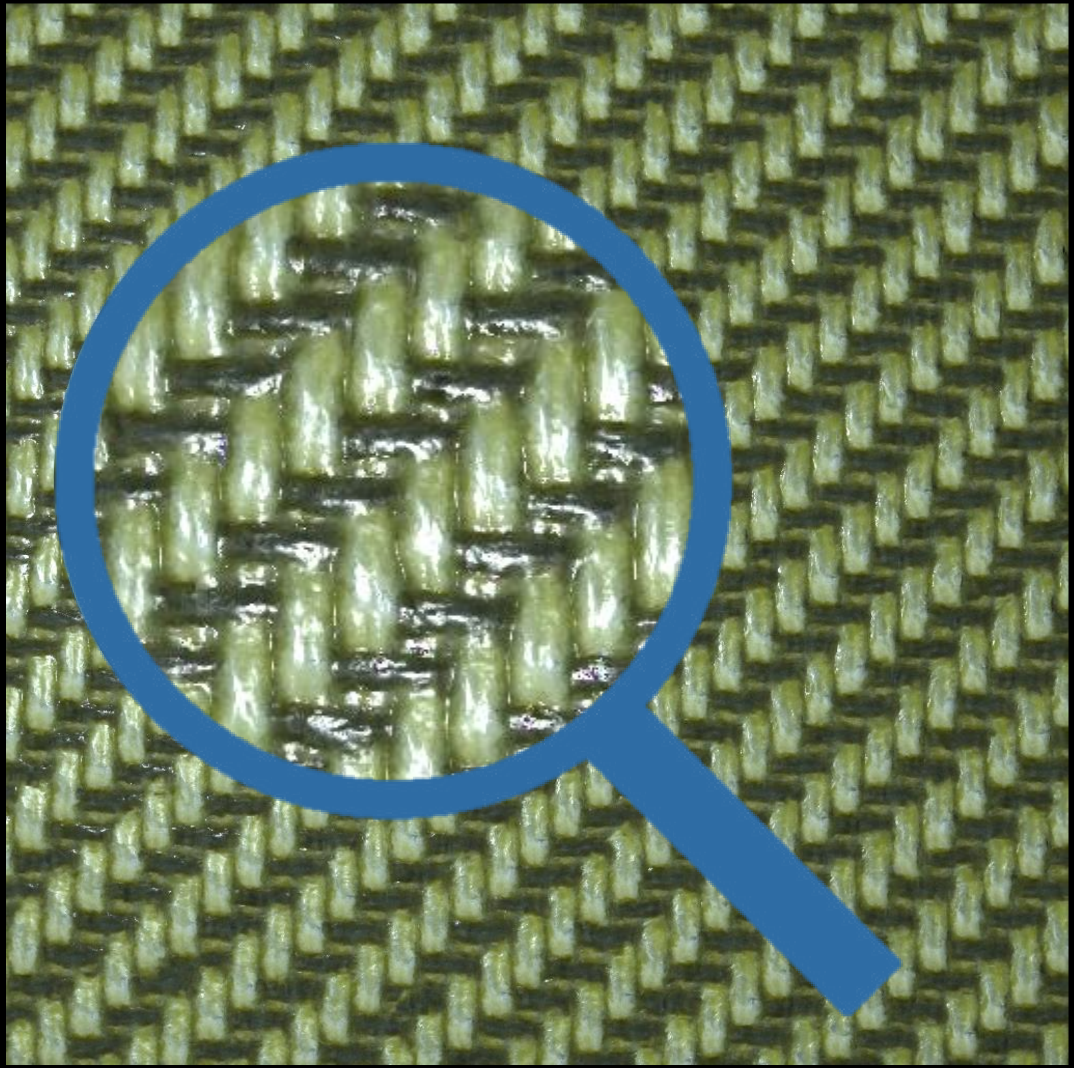


HIGH RESOLUTION RESULTS

Support arbitrary resolution!



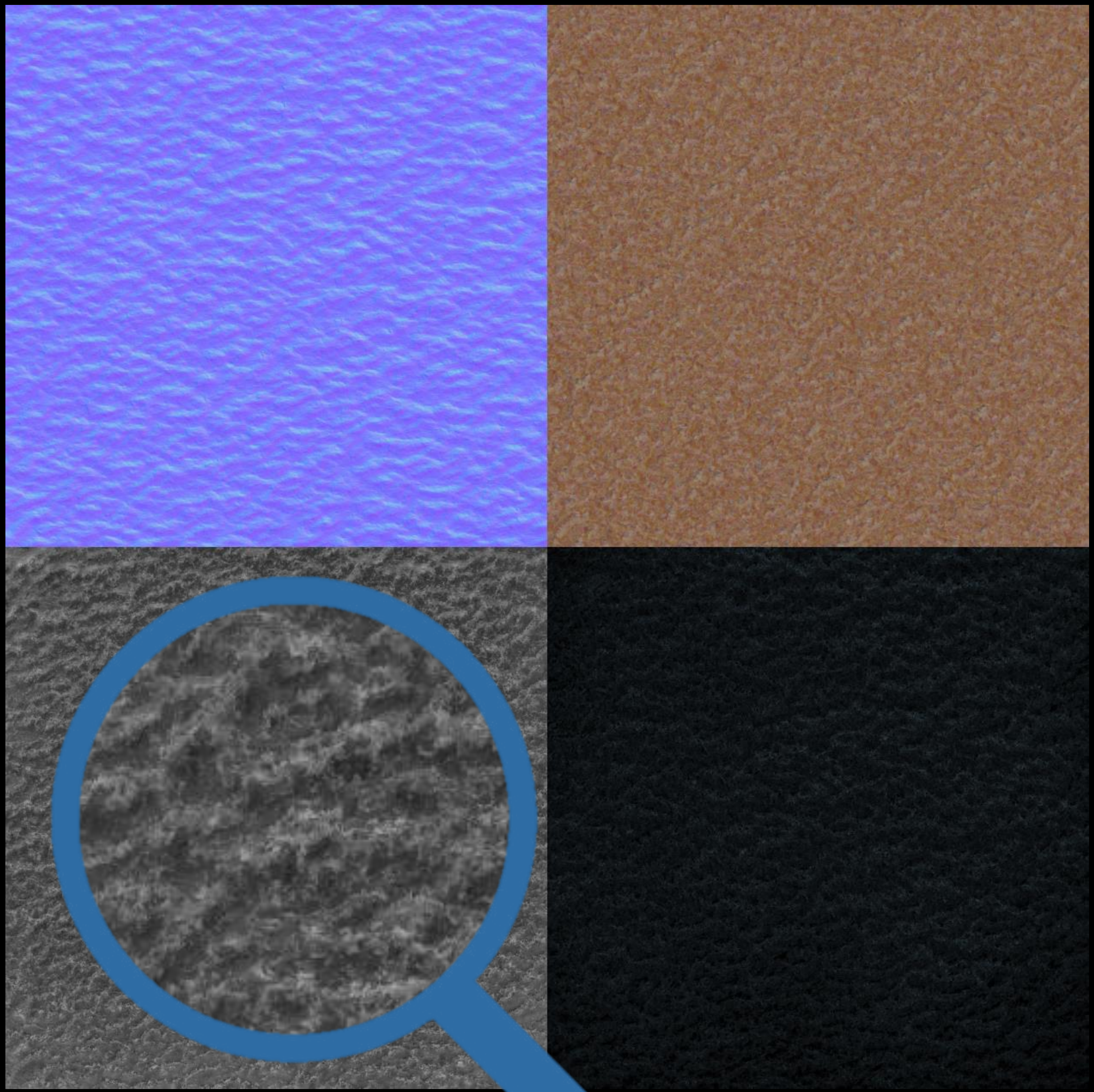
Estimated SVBRDF with 20 input photos



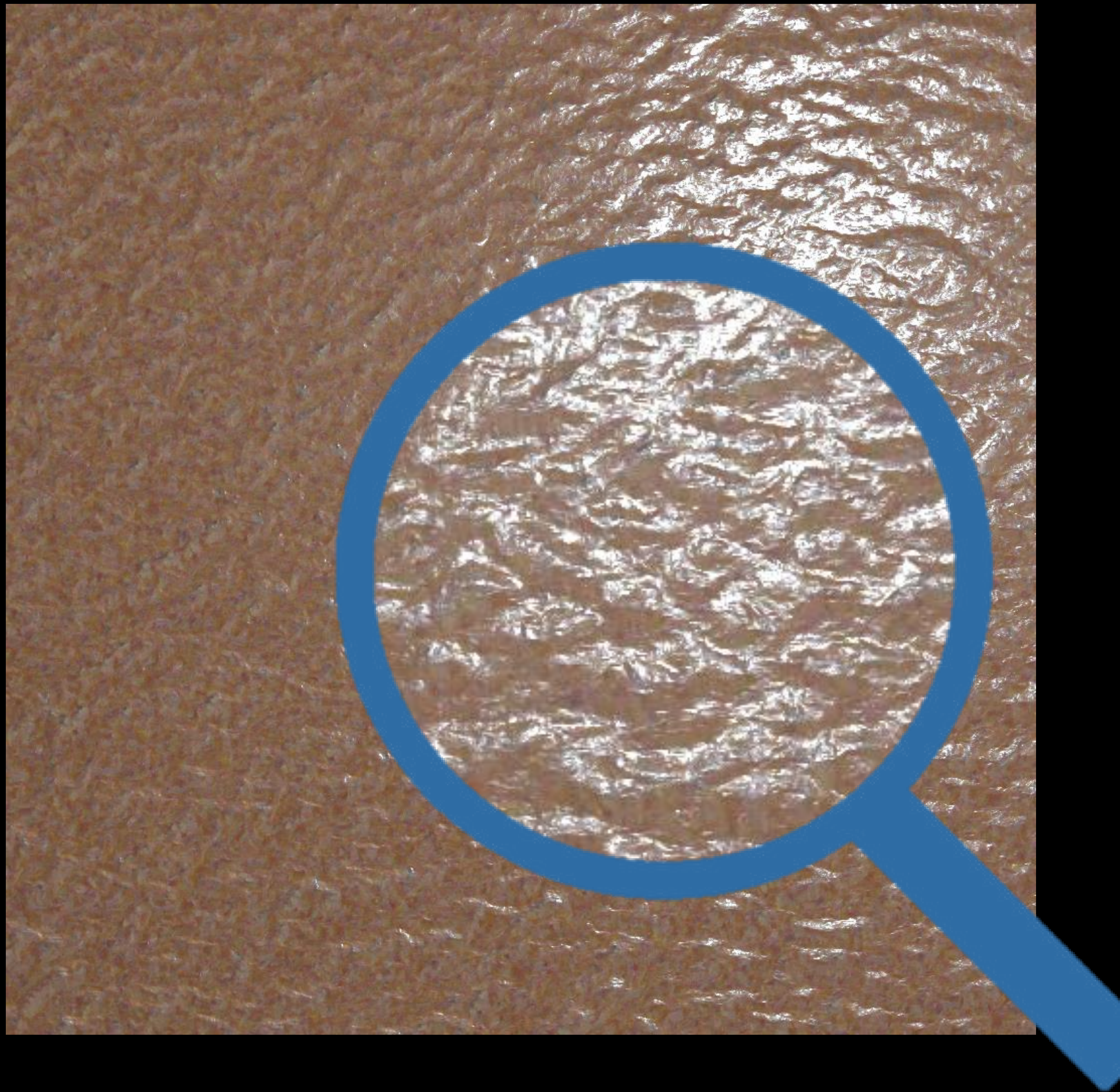
Novel view rendering

HIGH RESOLUTION RESULTS

Support arbitrary resolution!



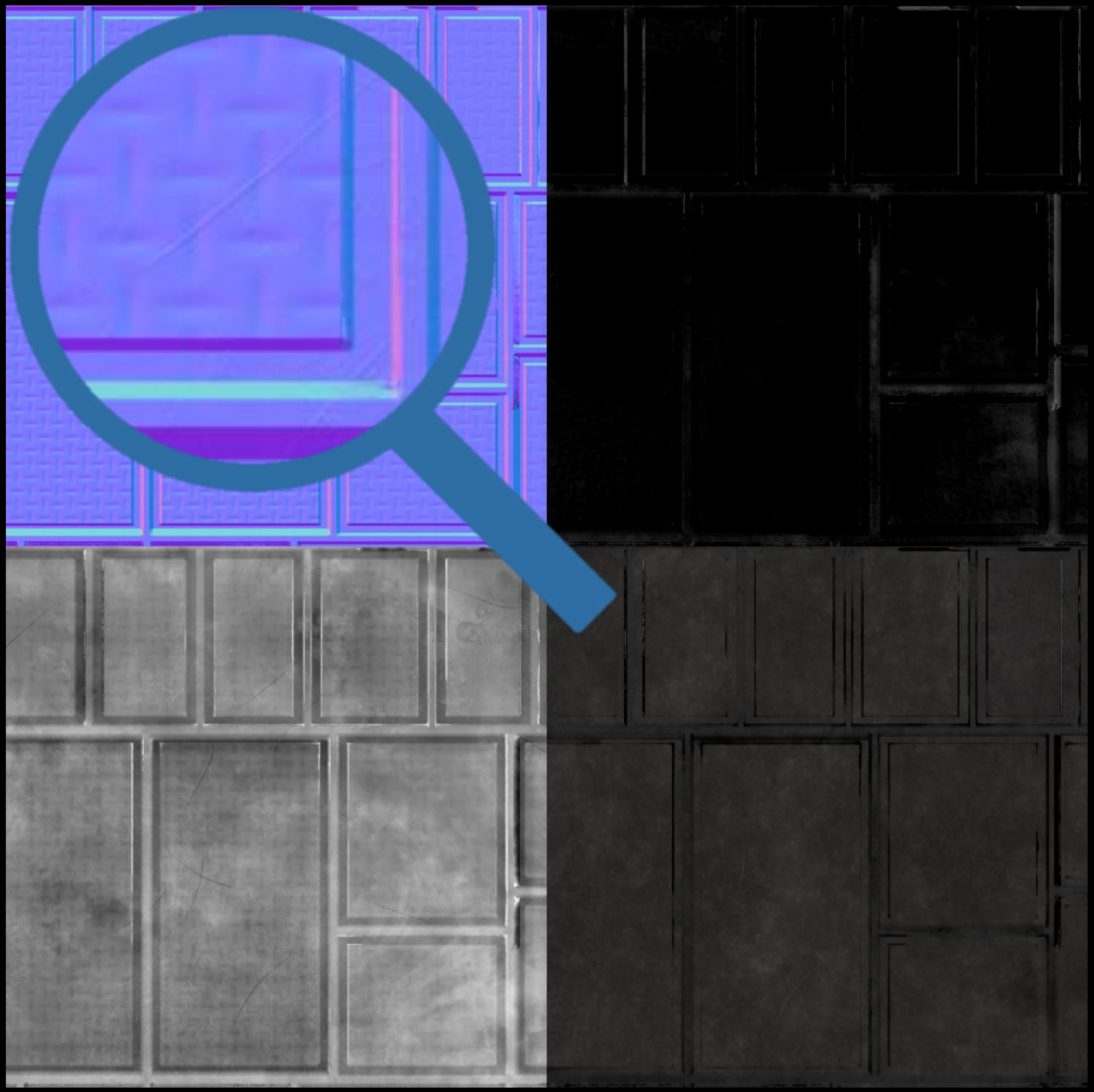
Estimated SVBRDF with 20 input photos



Novel view rendering

HIGH RESOLUTION RESULTS

Support arbitrary resolution!

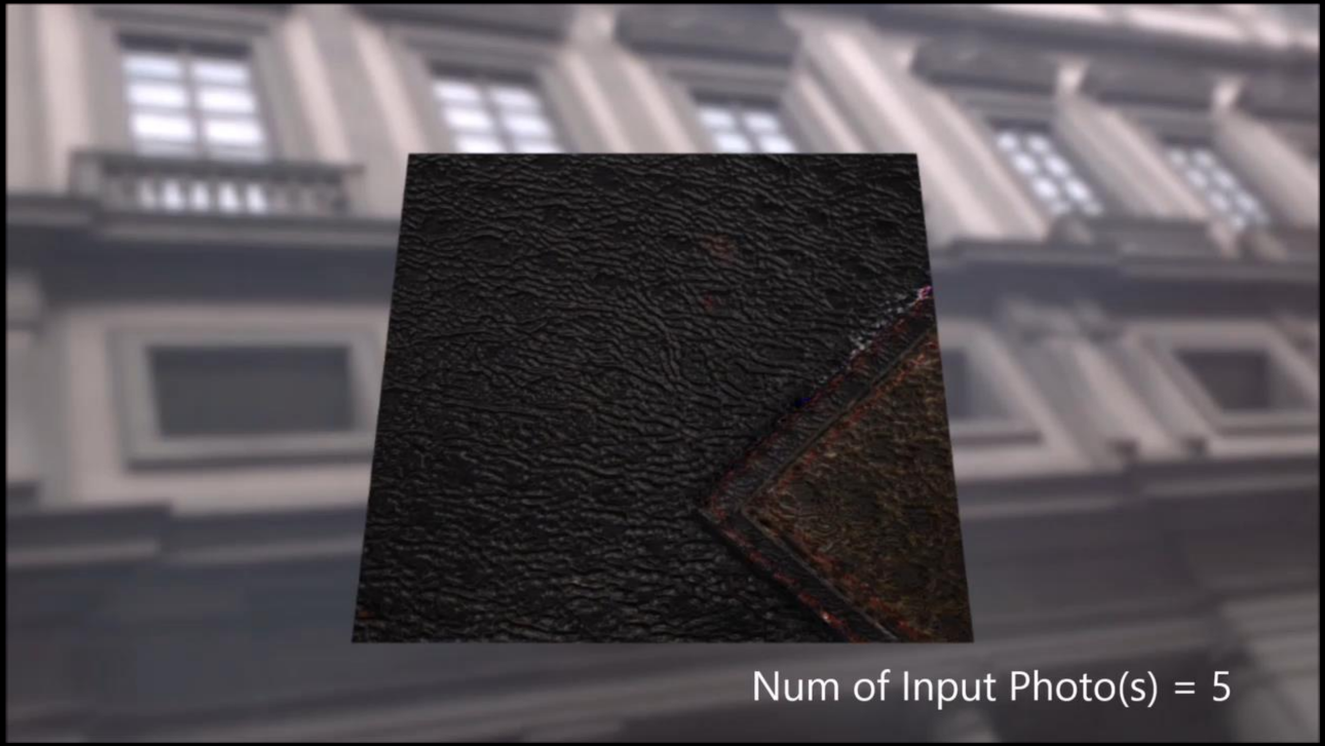


Estimated SVBRDF with 20 input photos

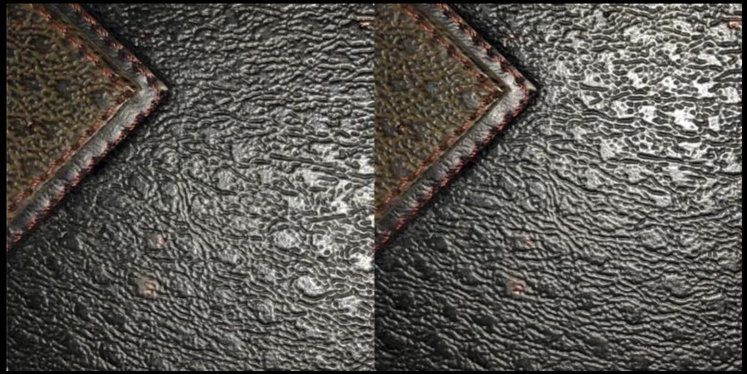
Novel view rendering

REAL CAPTURED RESULTS

Leather, 1k resolution, 5 input images

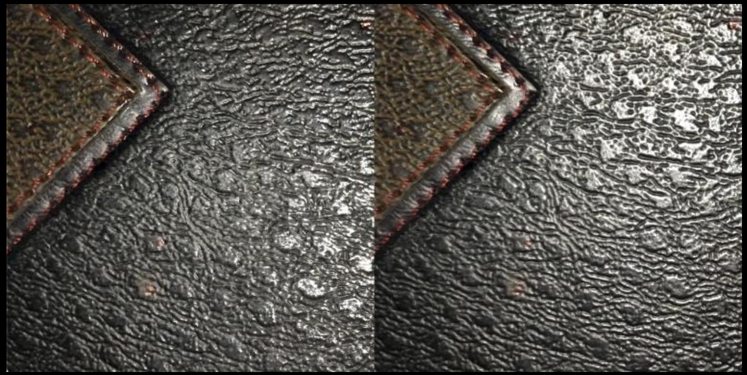


GT



N=1

N=5



Novel view

REAL CAPTURED RESULTS

Card, 1k resolution, 20 input images

GT

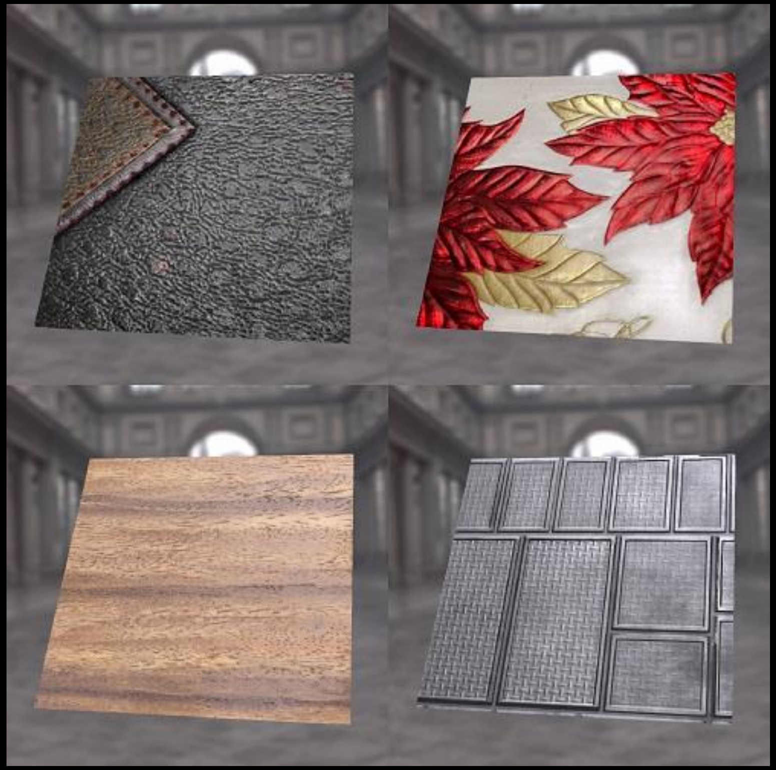
N=1

N=20

Novel view

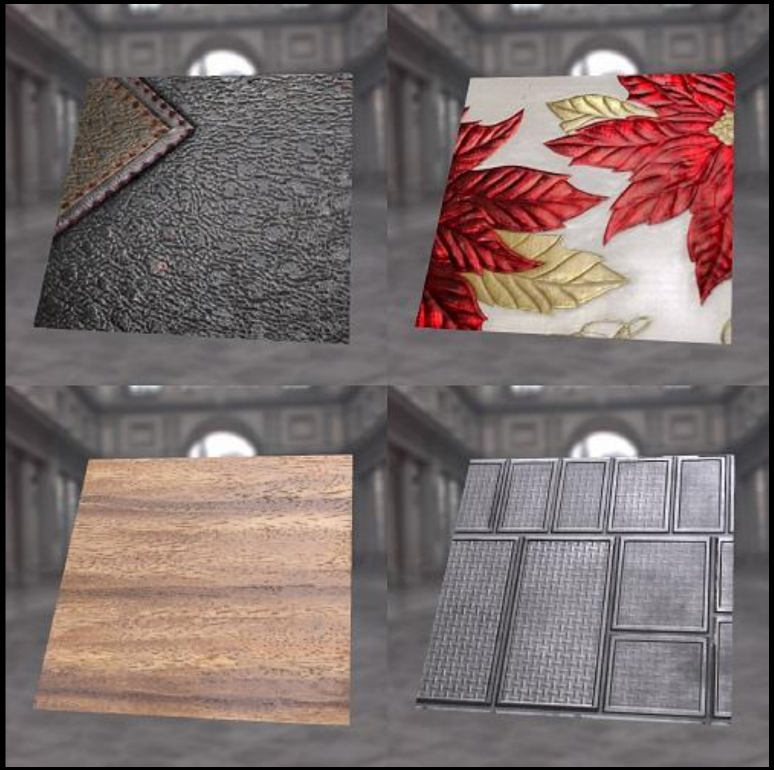
CONCLUSION & FUTURE WORK

- A unified deep inverse rendering framework
  - Performs optimization in SVBRDF latent space
  - Handles arbitrary number of inputs
- Future Work
  - Leverage better initialization strategy
  - Geometry + appearance estimation



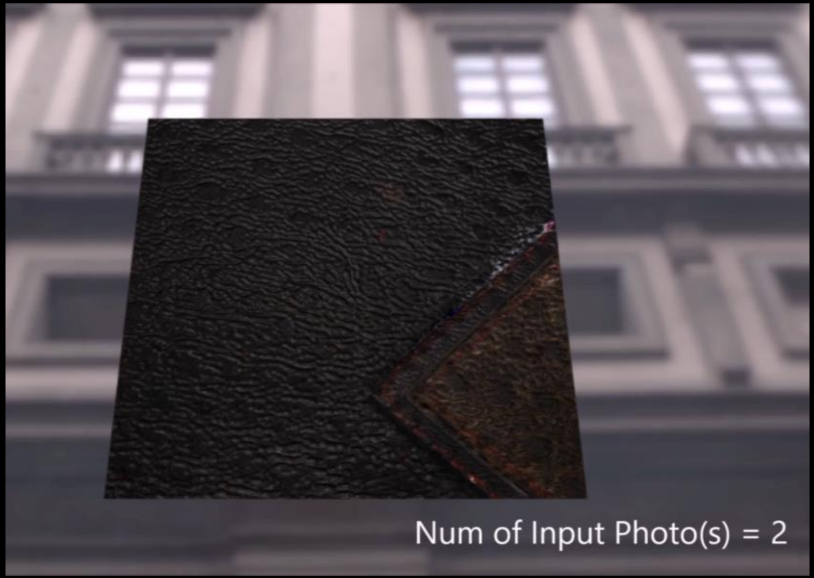
CONCLUSION & FUTURE WORK

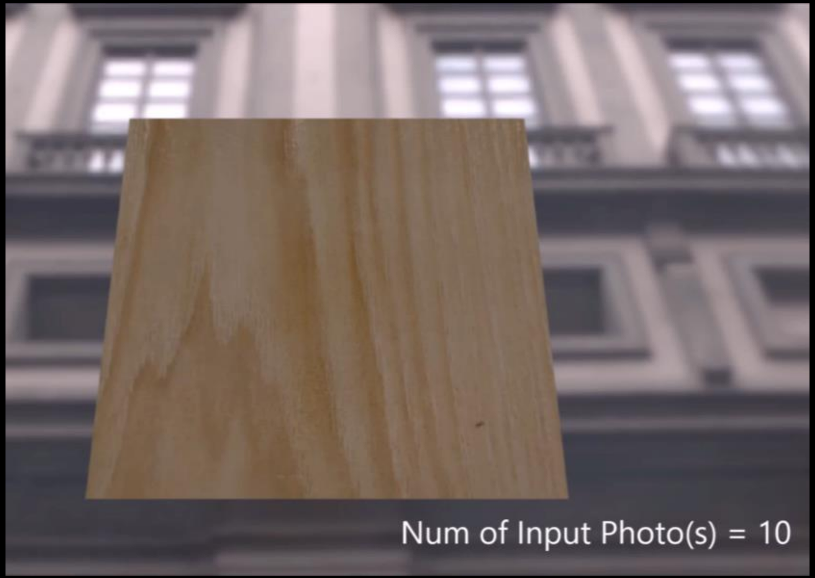
- A unified deep inverse rendering framework
  - Performs optimization in SVBRDF latent space
  - Handles arbitrary number of inputs
- Future Work
  - Leverage better initialization strategy
  - Geometry + appearance estimation



ACKNOWLEDGEMENTS

- Anonymous Reviewers
- Deep Materials dataset and model [Deschaintre et al. 2018]
- NSF grant IIS - 1350323
- National Natural Science Foundation of China


Num of Input Photo(s) = 2


Num of Input Photo(s) = 10


Num of Input Photo(s) = 20

Thanks