SIGGRAPH 2021

HIGHLIGHT-AWARE TWO-STREAM NETWORK FOR SINGLE-IMAGE SVBRDF ACQUISITION

Jie Guo[1] **Shuichang Lai**[1] Chengzhi Tao[1]

Yuelong Cai[1] Lei Wang[2] Yanwen Guo[1] Ling-qi Yan[3]

[1]State Key Lab for Novel Software Technology, Nanjing University

[2]Guangdong OPPO Mobile Telecommunications Corp Ltd

[3]University of California, Santa Barbara

© 2021 SIGGRAPH. ALL RIGHTS RESERVED.

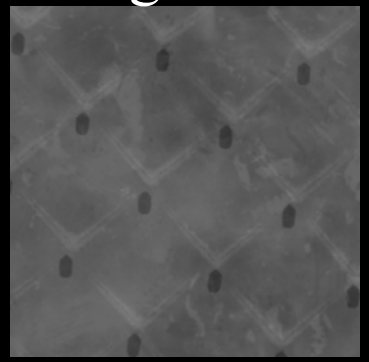
THE PREMIER **CONFERENCE** & **EXHIBITION** IN COMPUTER GRAPHICS & INTERACTIVE TECHNIQUES

SVBRDF Acquisition

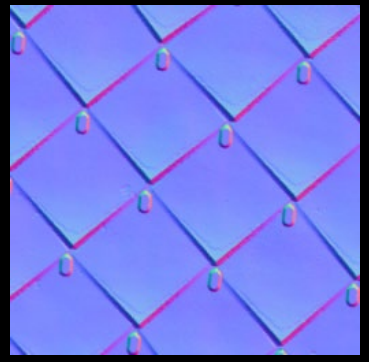
Input

Material

diffuse



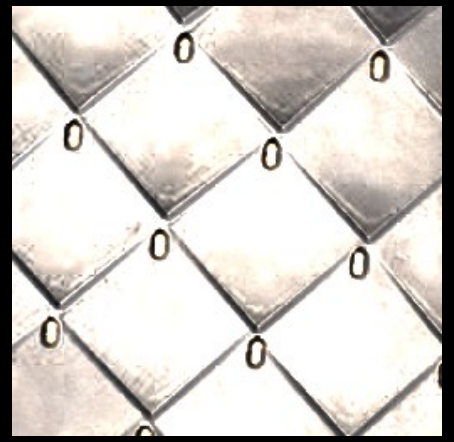
roughness



normal

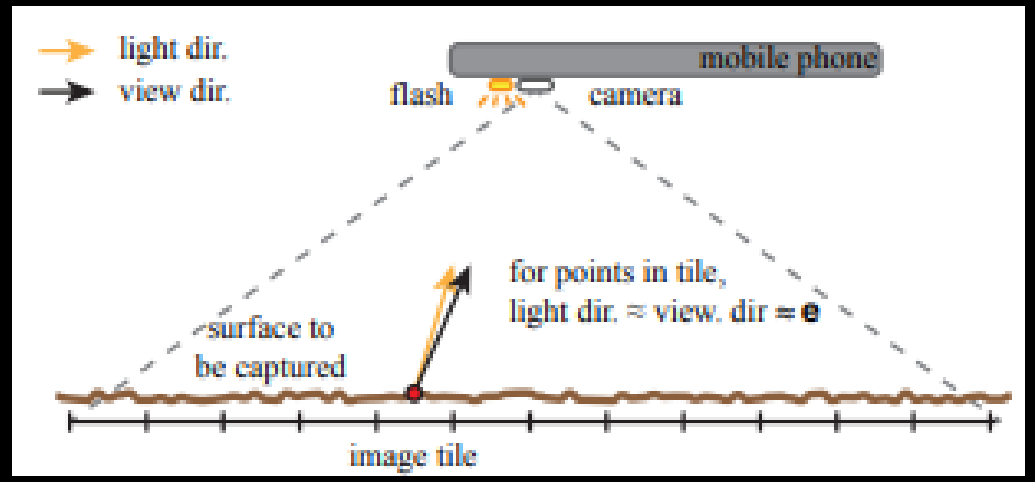
specular

Render



Related Work

Multiple input images



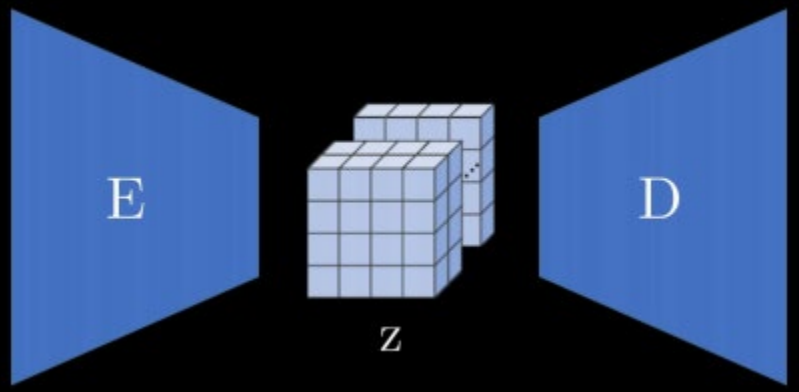
[Aittala et al. 2016]

[Asselin et al. 2020]

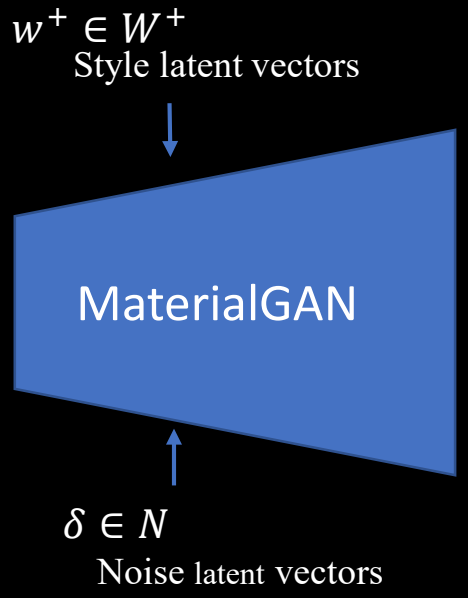
- High-quality
- Specialized hardware
- Complex operation

Related Work

Multiple input images



[Gao et al. 2019]



[Guo et al. 2020]

- Time-comsuming
- Cumbersome optimization
- Extra camera calibration

Related Work

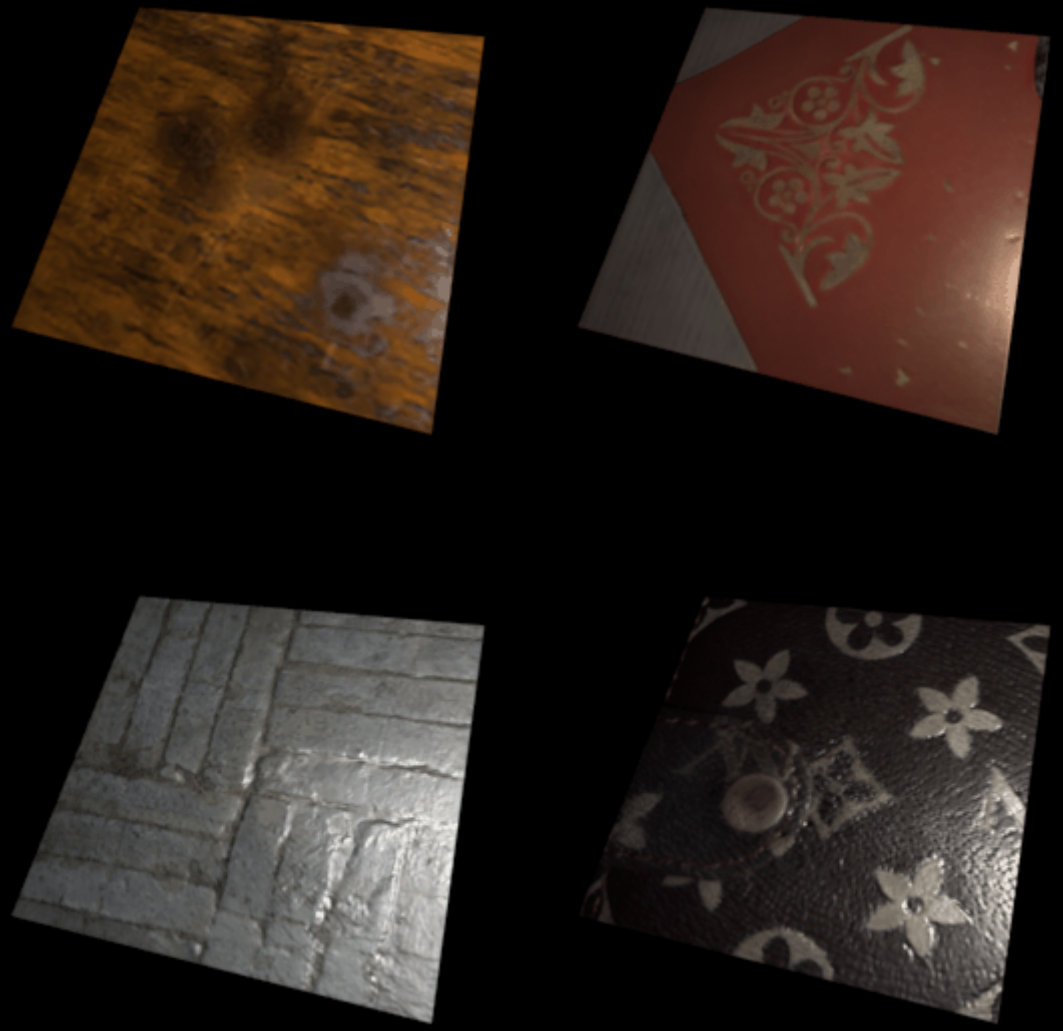
Single input image

[Deschaintre et al. 2018]

- Weak disentanglement
- Highlight ambiguity

Our Goal

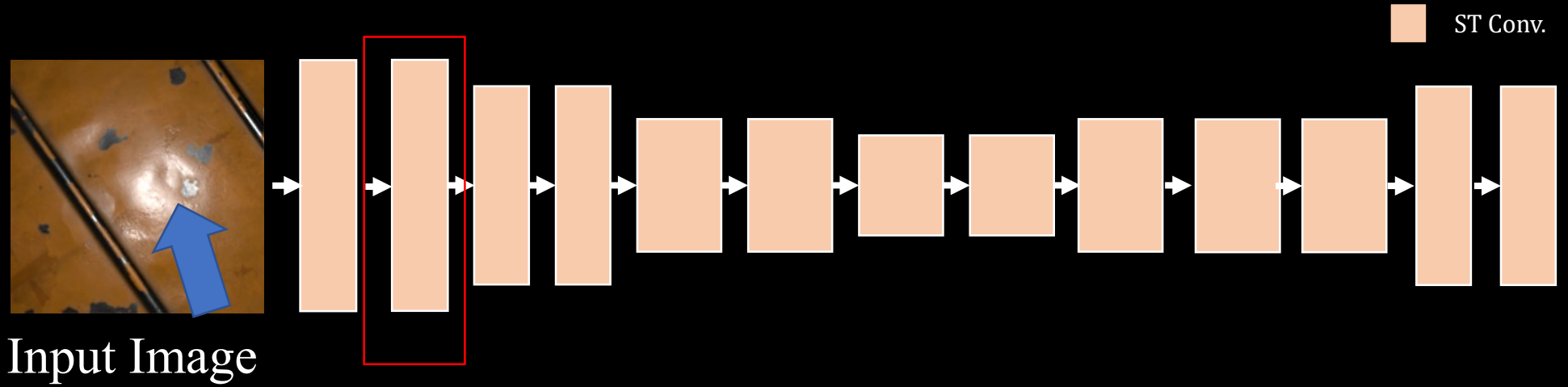
- Single, casually captured image as input (without camera calibration)
- Well disentanglement
- Highlight aware
- Fast evaluation



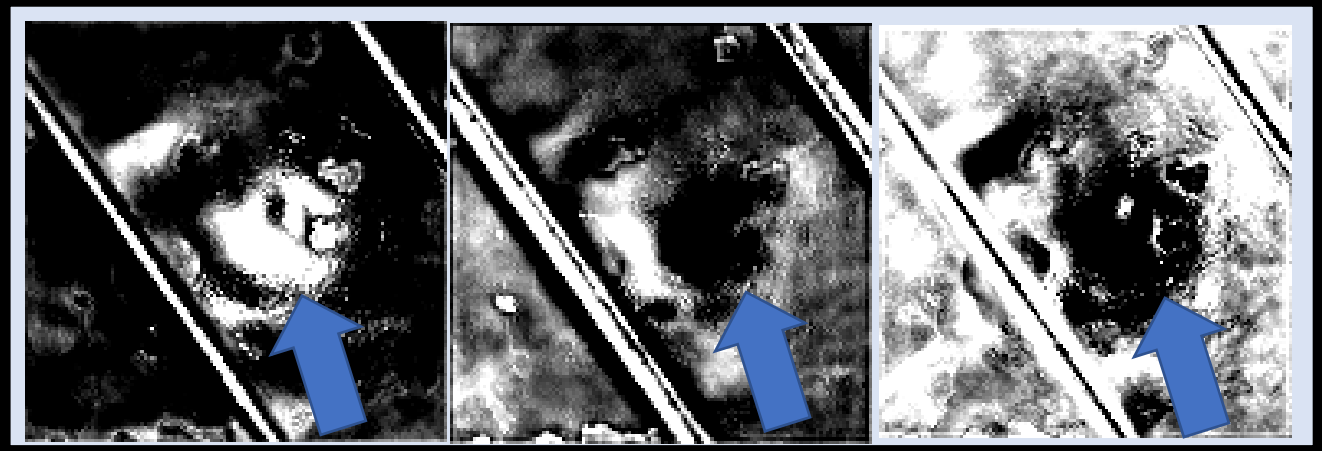
Key Challenge

Highlight pollution

Standard Unet Design



Input Image

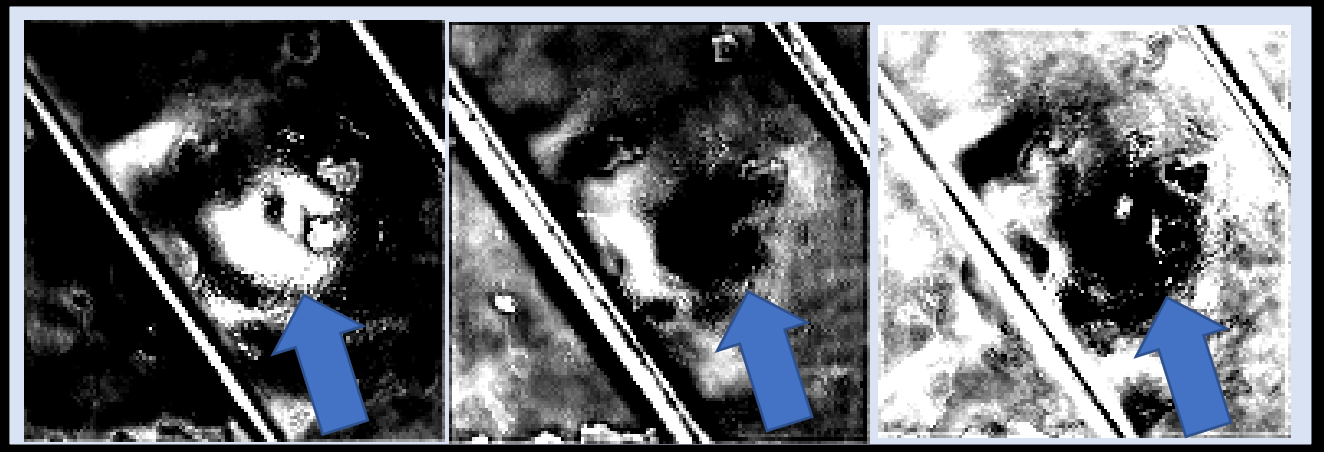


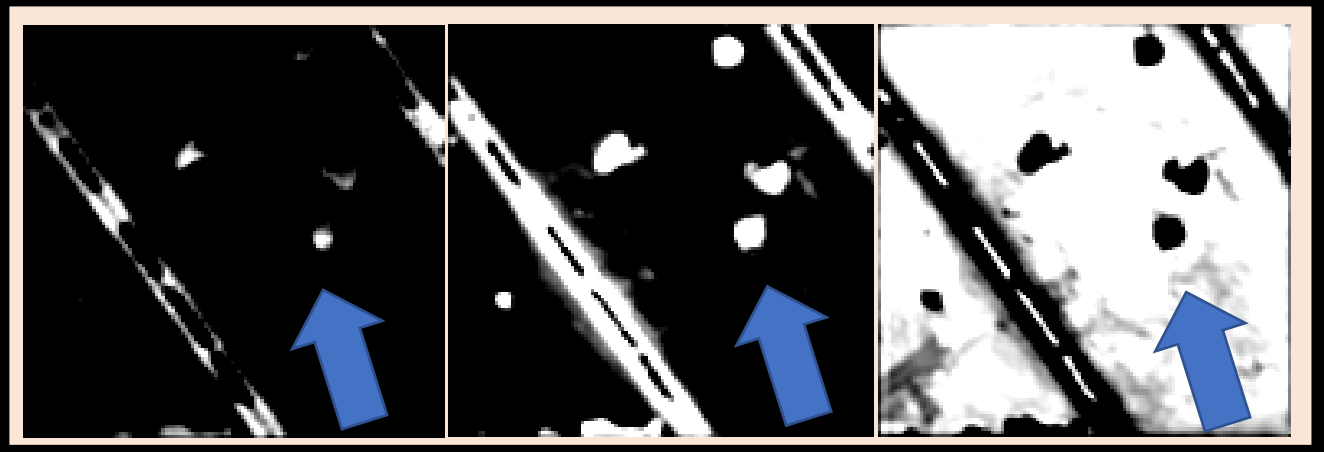
Standard features

Ambiguity

Highlight-Aware Convolution

Input Image

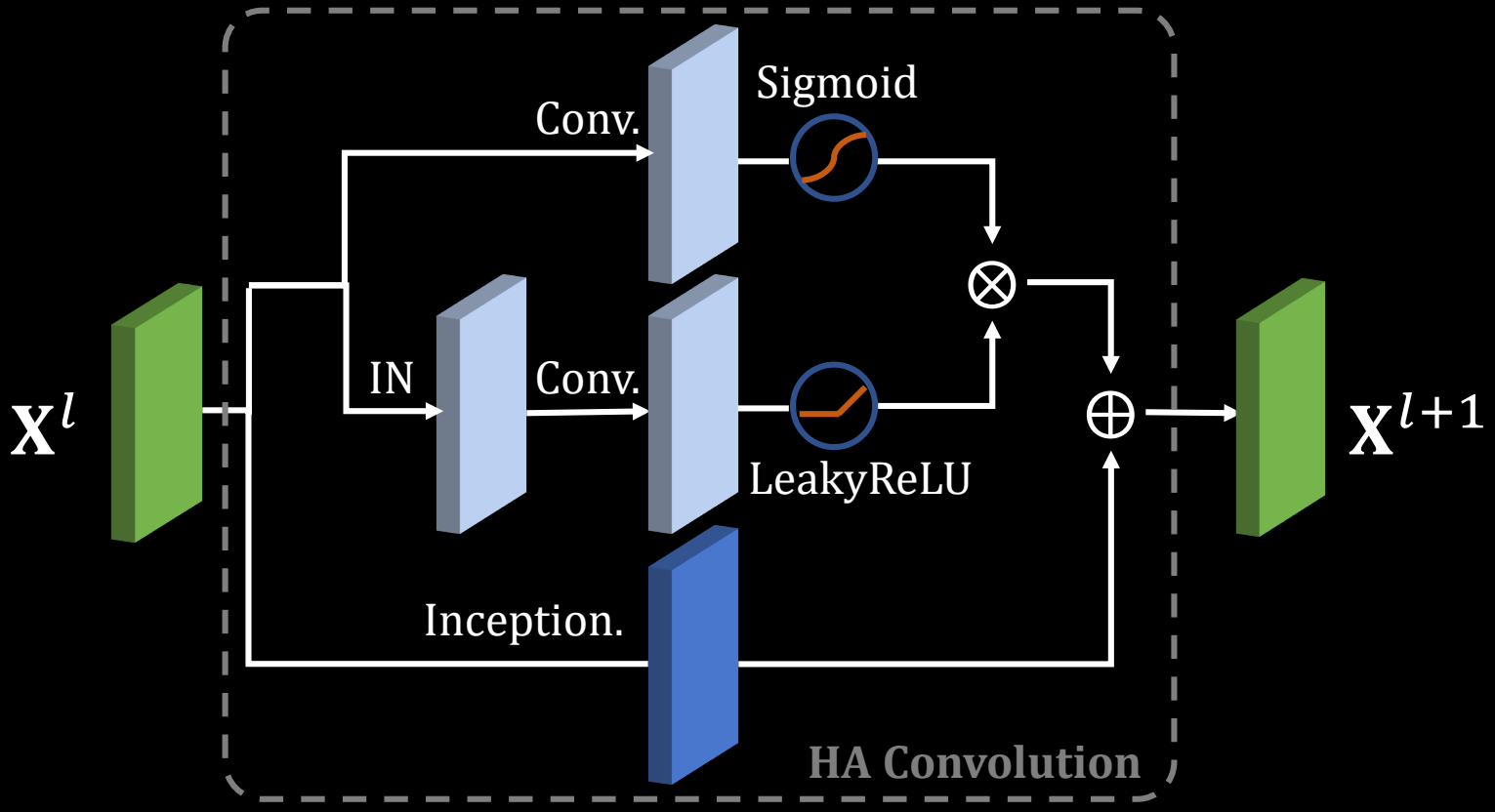


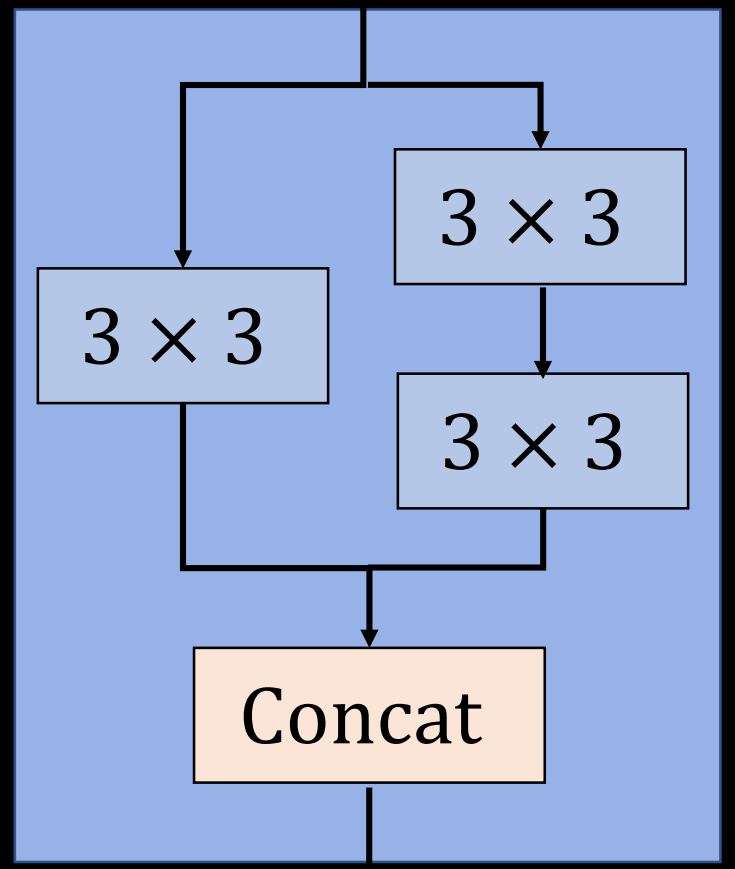


Highlight-Aware features

Recovered diffuse map

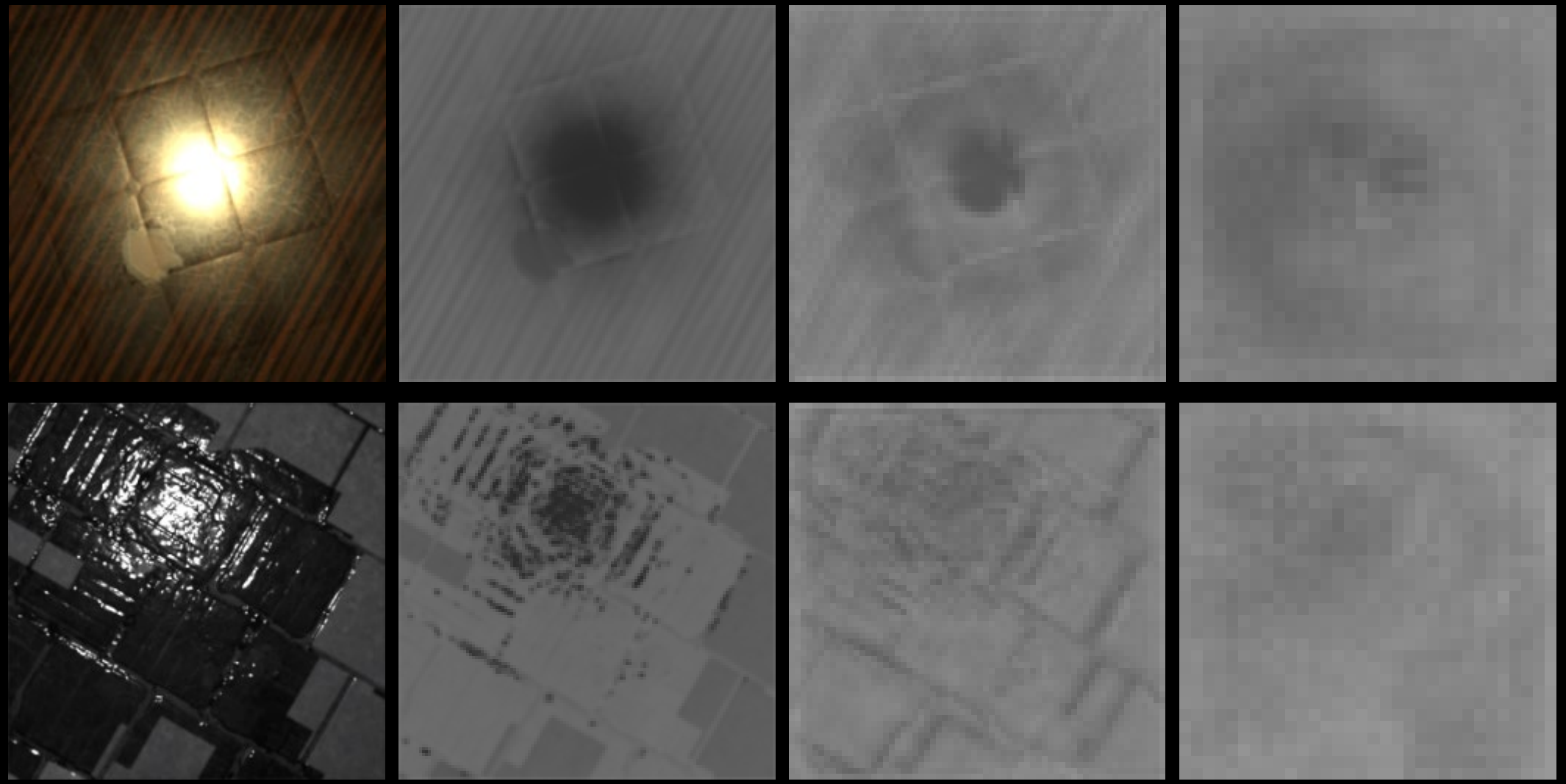
Highlight-Aware Convolution





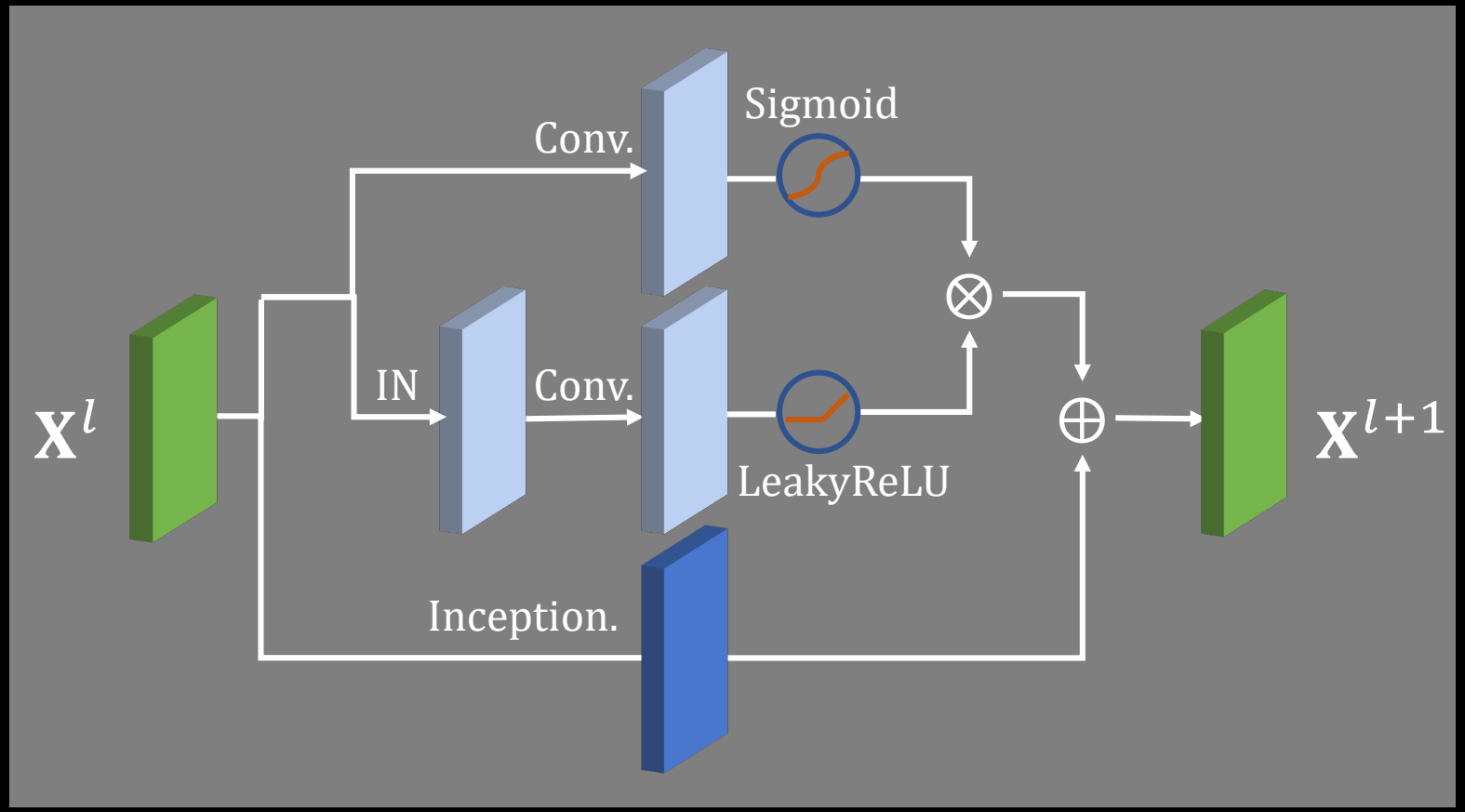
Inception

Highlight-Aware Convolution

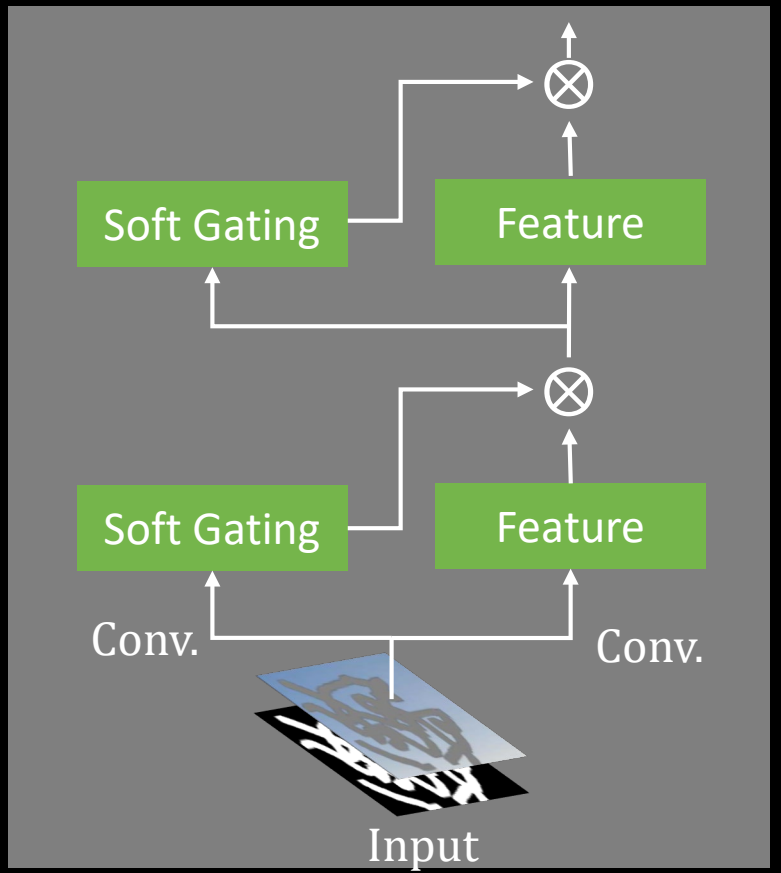


Input (256×256) Layer 1 (128×128) Layer 4 (64×64) Layer 7 (32×32)

Highlight-Aware Convolution vs Gated Conv.

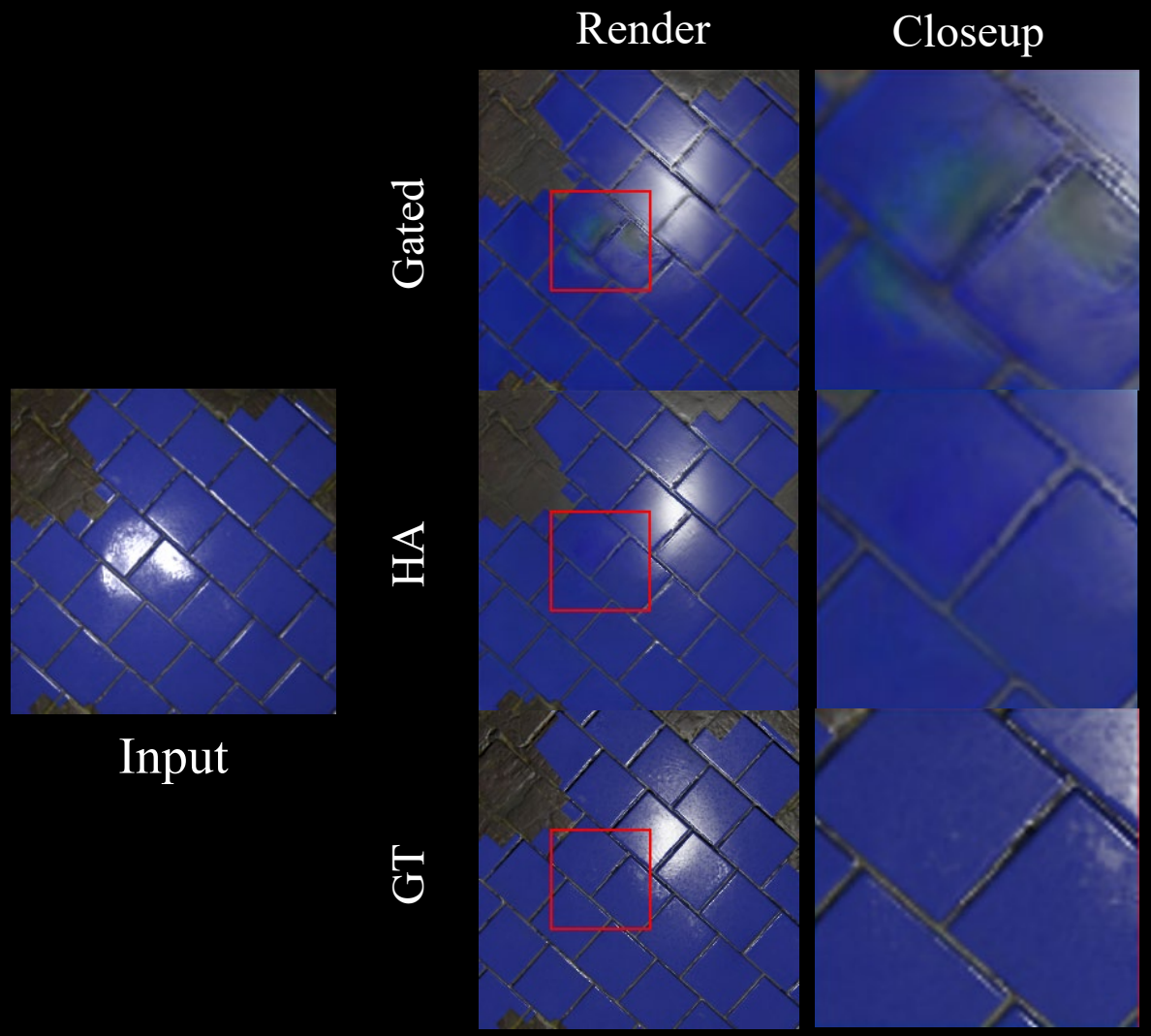


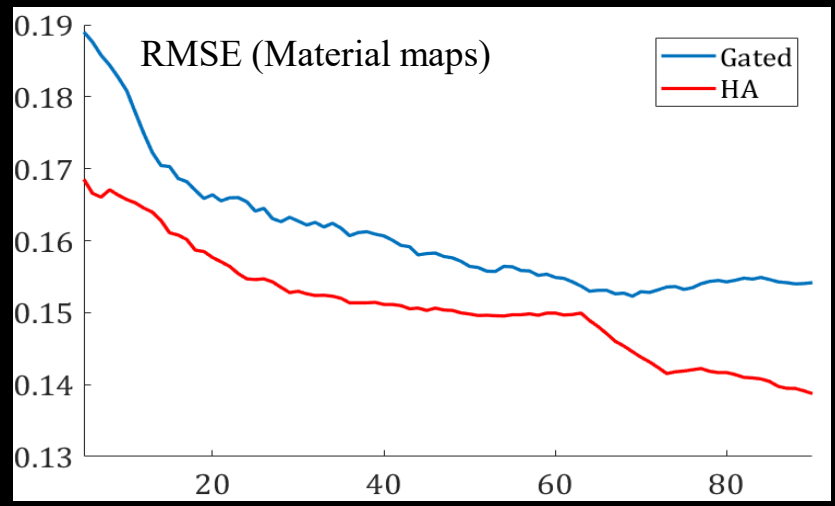
HA Convolution

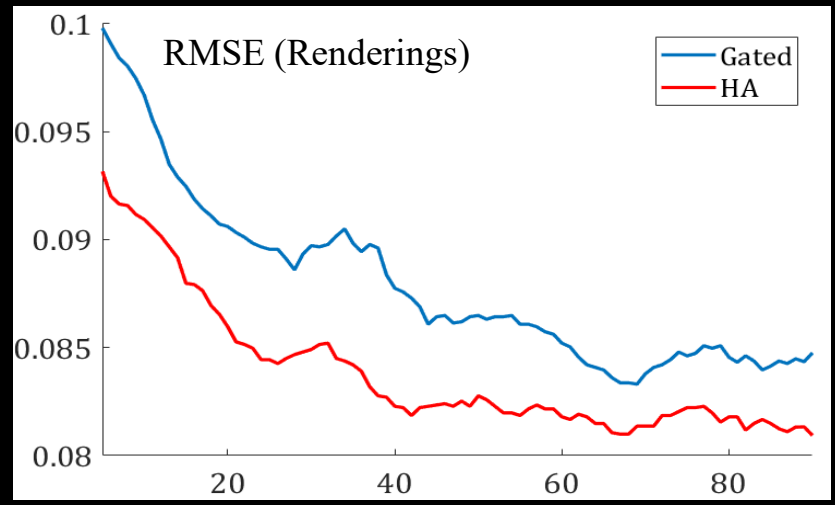


Gated Convolution
[Yu et al. 2019]

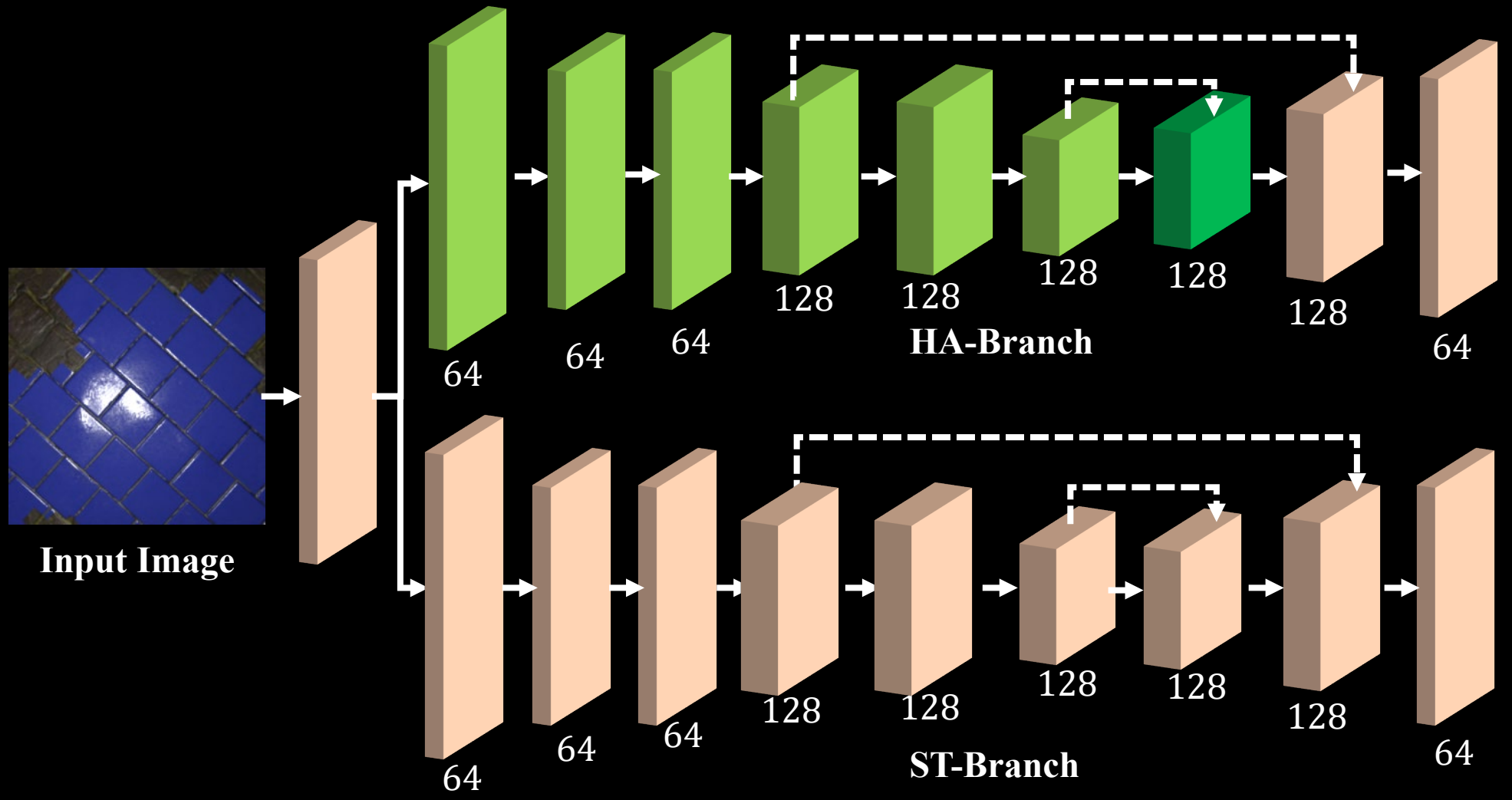
Highlight-Aware Convolution vs Gated Conv.



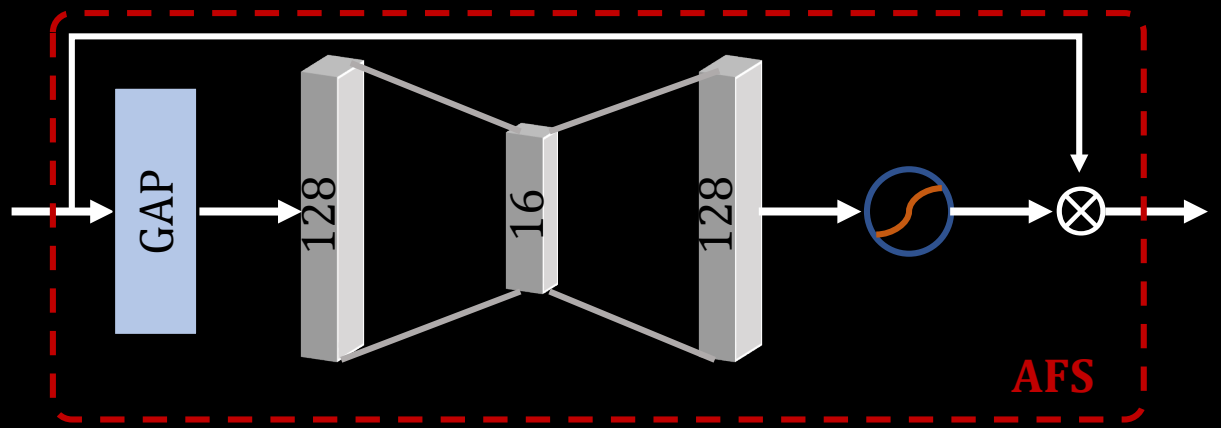


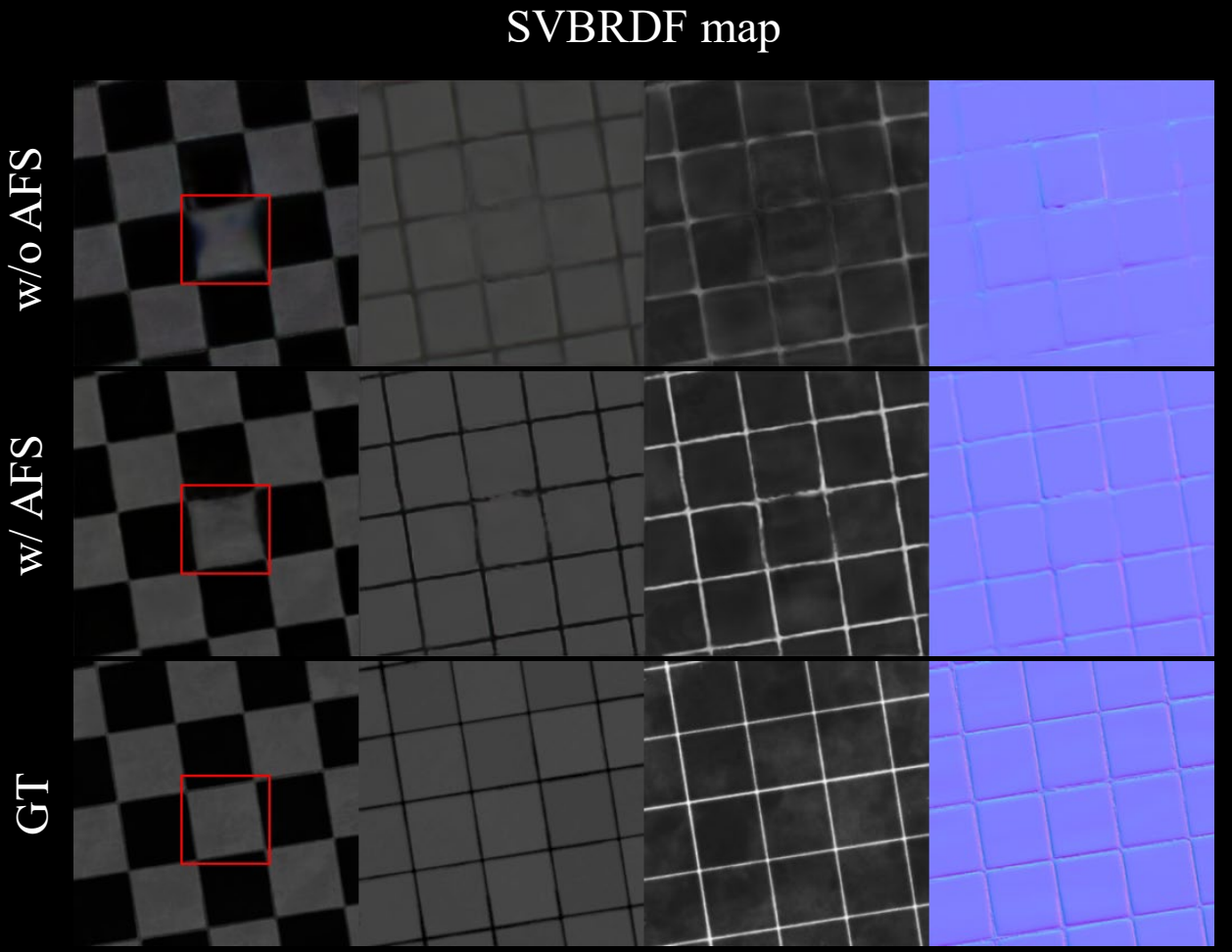


Feature Extraction with Two-Stream

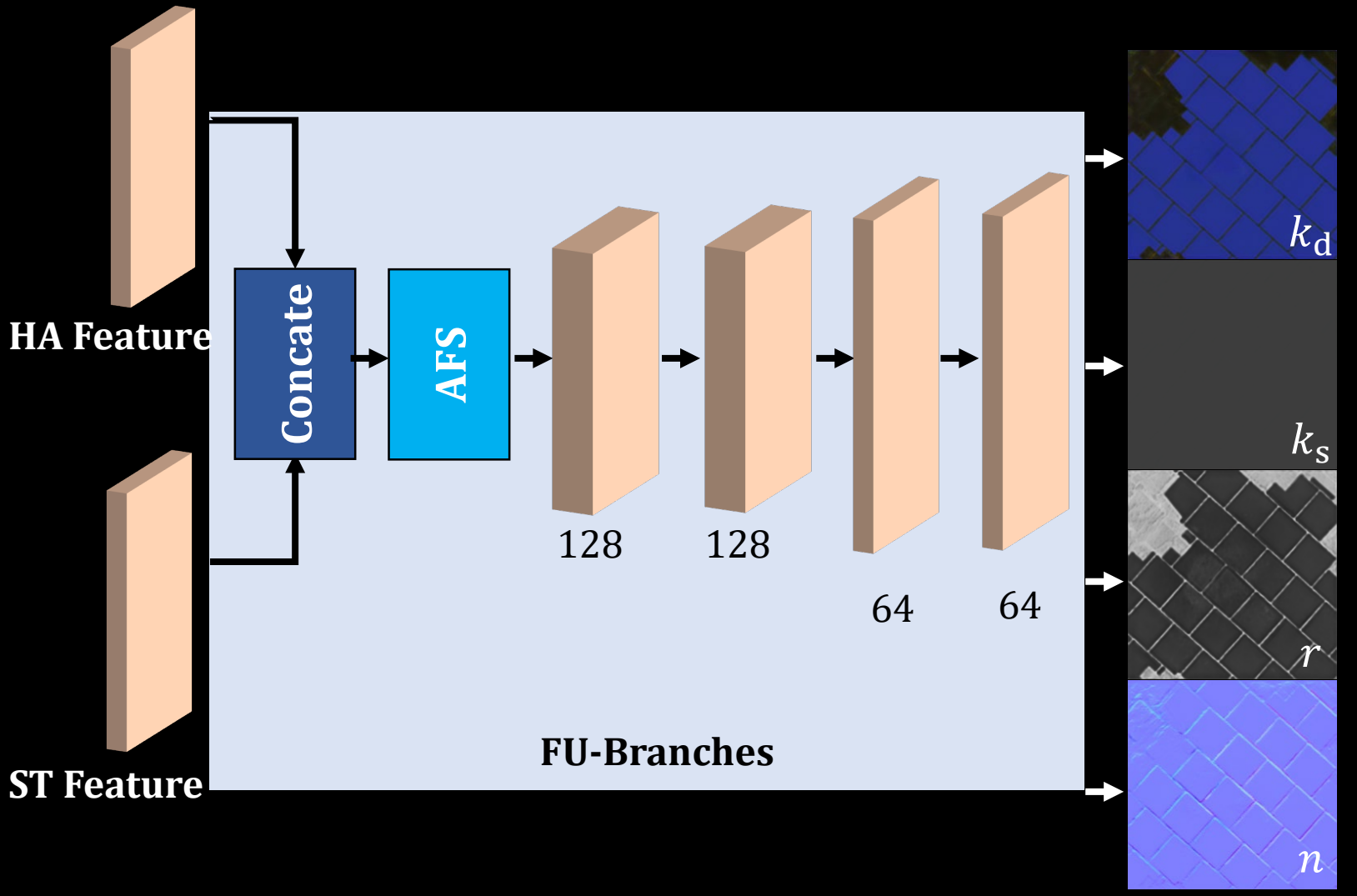


Attention-based Feature Selection Module

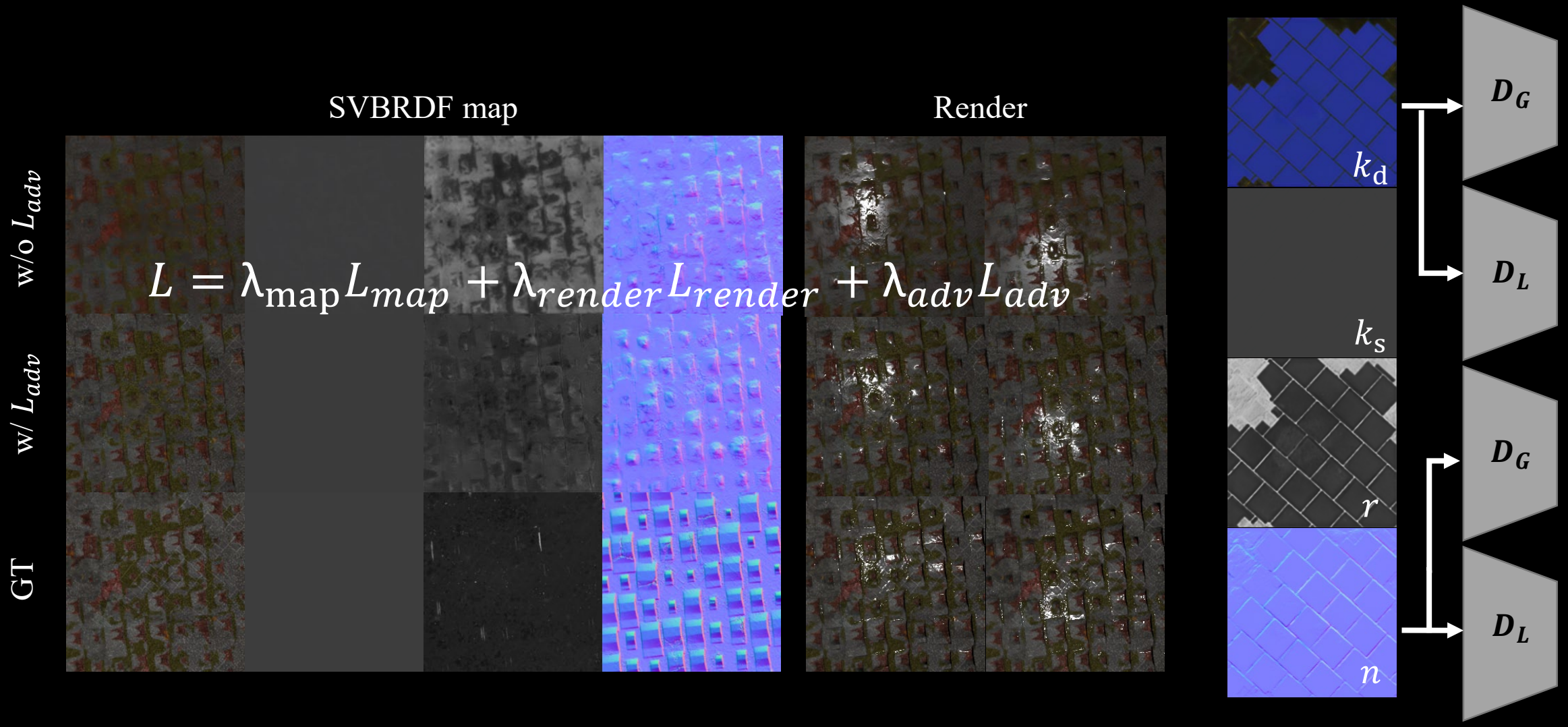




Feature Fusion and Selection with FU-Branch

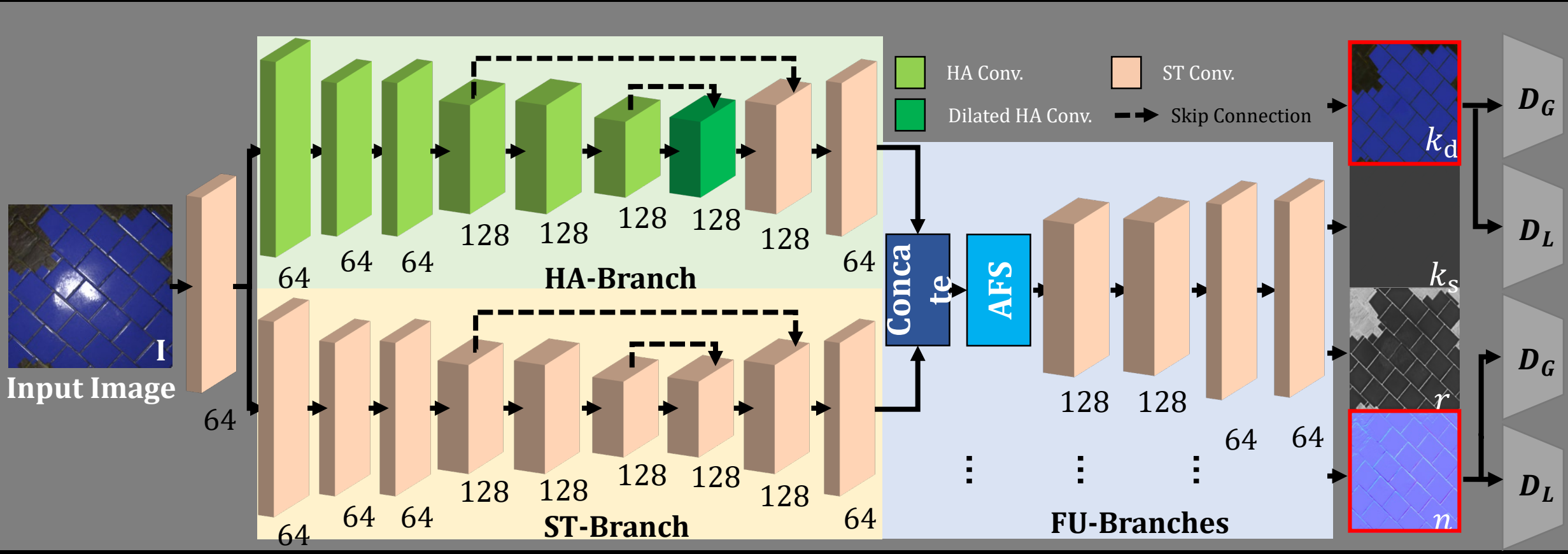


Loss


SVBRDF map
Render
w/o L_{adv}
w/ L_{adv}
GT
$L = \lambda_{\text{map}} L_{map} + \lambda_{render} L_{render} + \lambda_{adv} L_{adv}$
k_d
k_s
r
n
D_G
D_L
D_G
D_L

Pipeline

Highlight-Aware Two-Stream Network

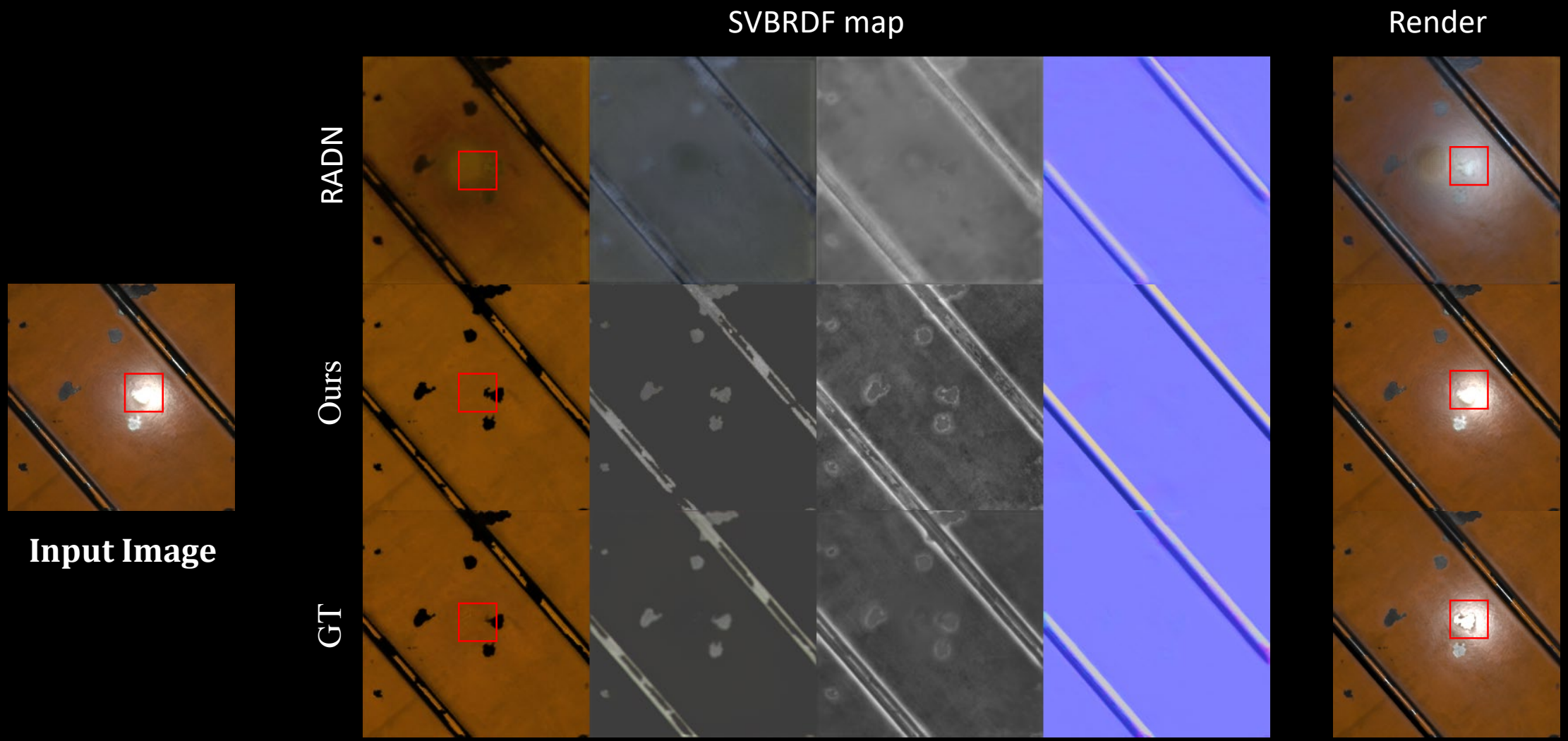


Results

We compare our method to

- Rendering-Aware Deep Network **(RADN)** [Deschaintre et al. 2018]: a **single image** SVBRDF acquisition method using classical U-Net architecture.
- Deep Inverse Rendering **(DIR)** [Gao et al. 2019]: a unified **deep inverse rendering** framework for estimating SVBRDF from an **arbitrary number of input** photographs.
- **MaterialGAN** [Guo et al. 2020]: a deep generative convolutional network based on StyleGAN2, trained to synthesize realistic SVBRDF parameter maps, needing **a number of images** as input.

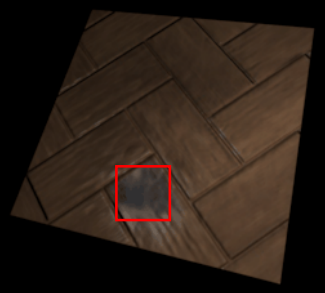
Comparison on Synthetic Data



Comparison on Synthetic Data

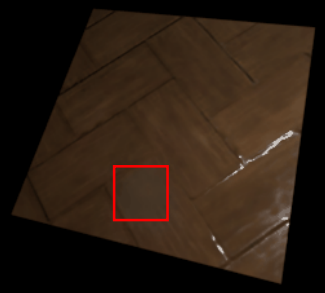
Input Image

RADN

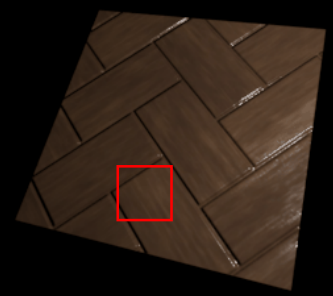


DIR

MaterialGAN



Ours



GT

Comparison on Real Data

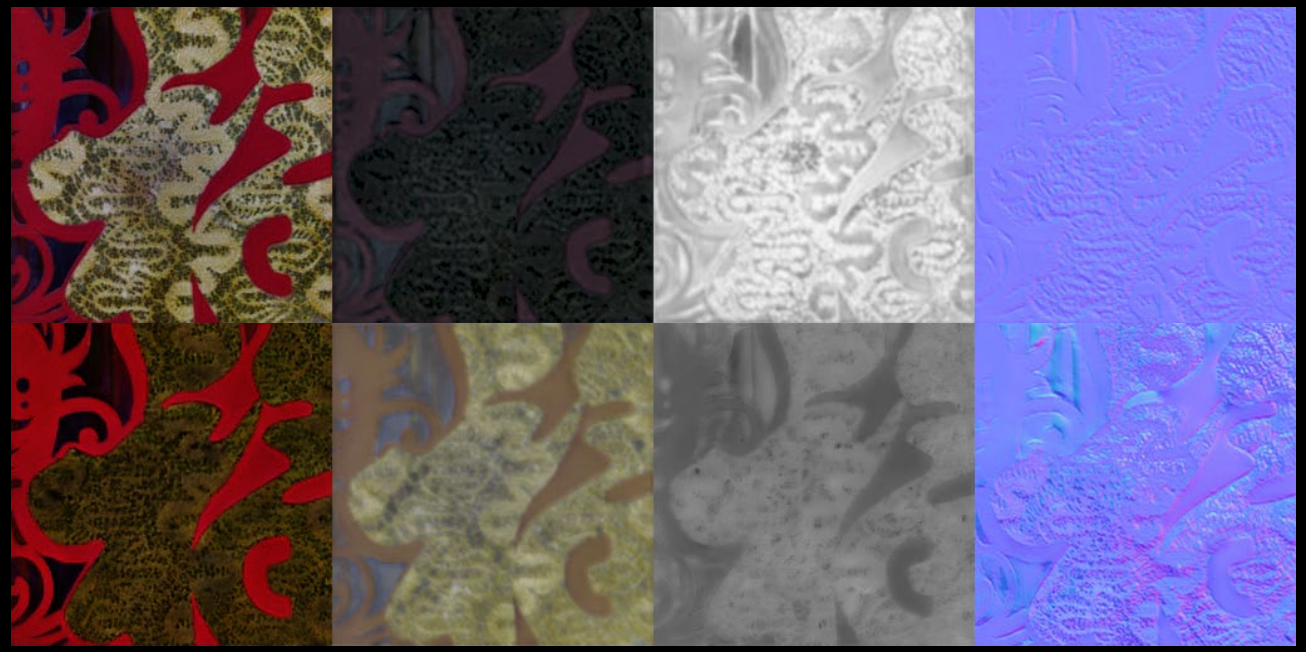
Input Image

SVBRDF map

Render

RADN

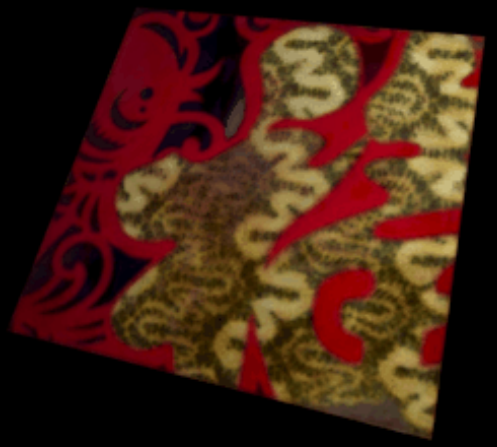
Ours



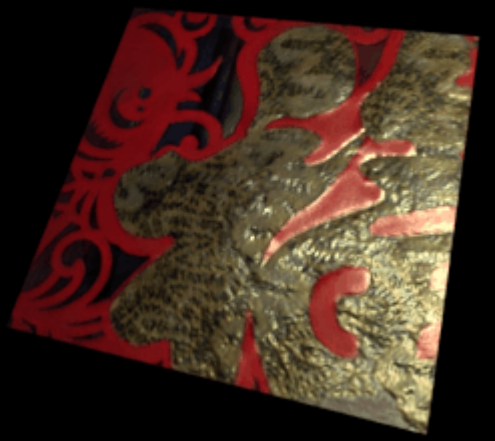
Comparison on Real Data

Input Image

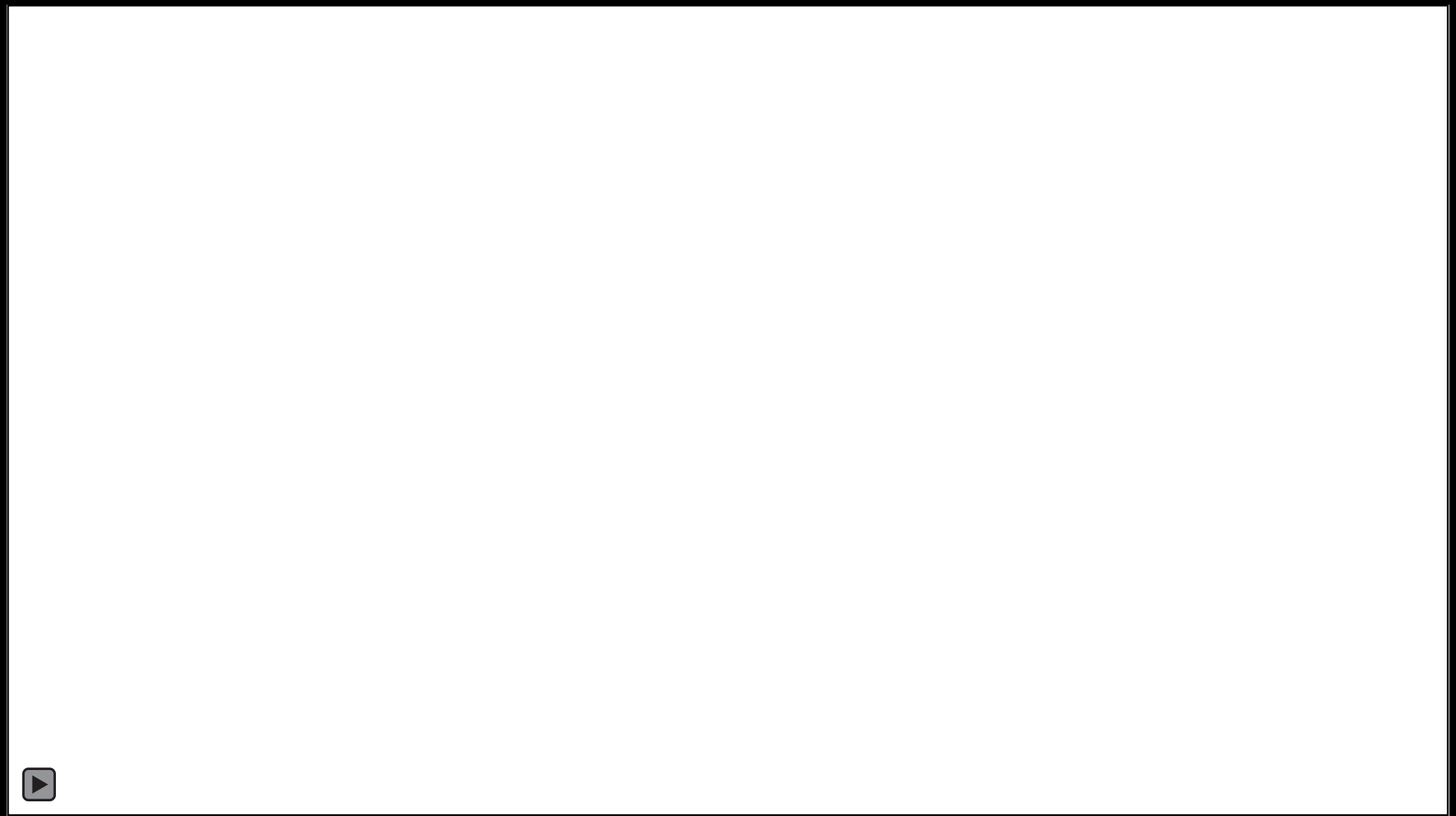
RADN



Ours

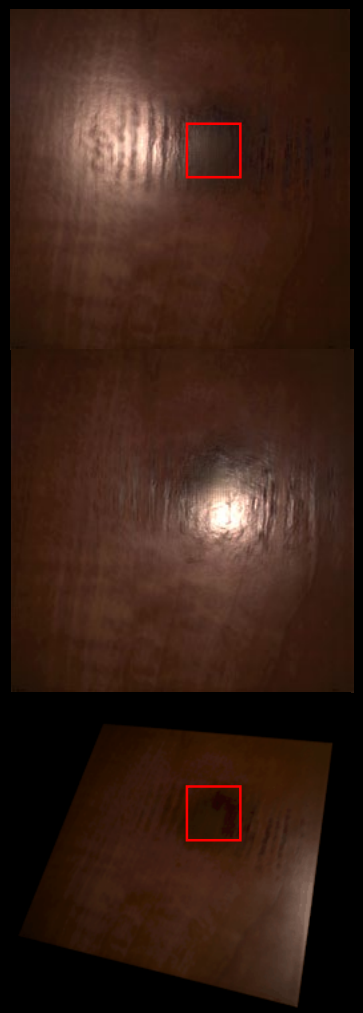


Comparison on Real Data

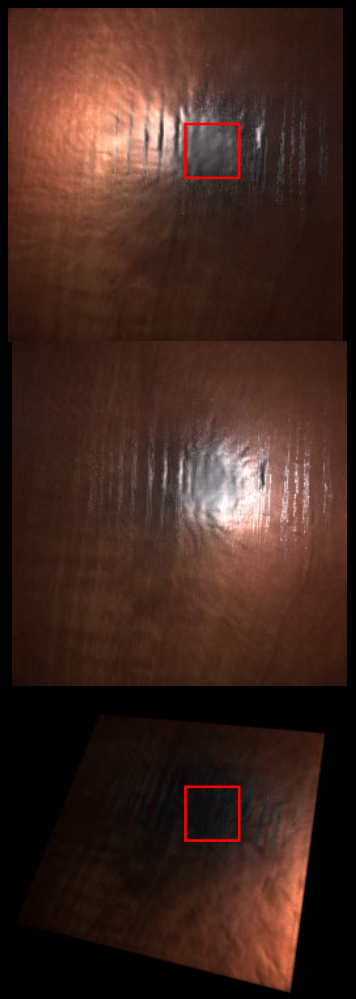


Comparison on Real Data

Input Image

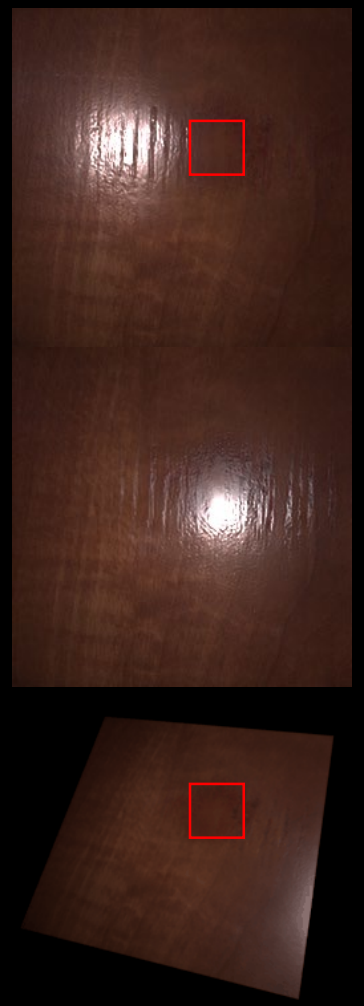


[Deschaintre et al. 2018]



[Gao et al. 2019]

[Guo et al. 2020]



Ours

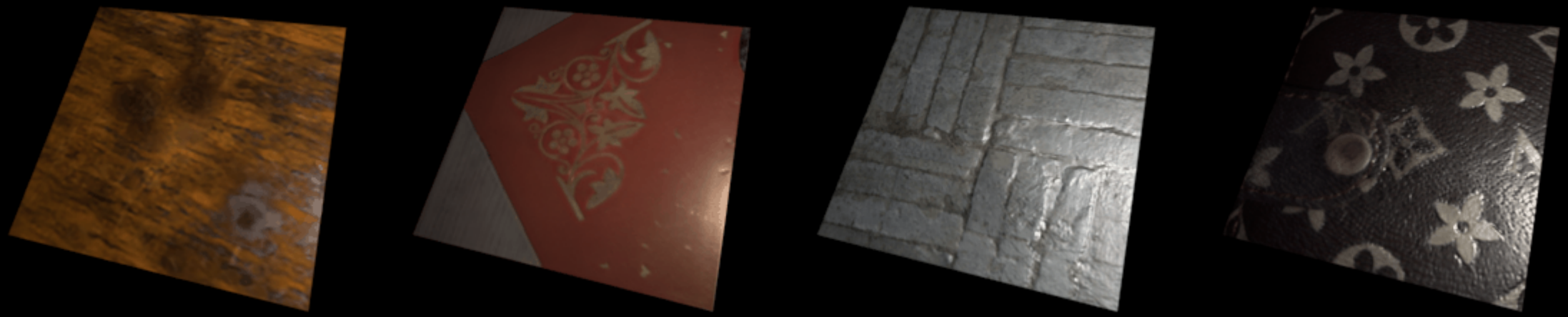
High-resolution (1024 × 1024) Results

Input Image

Ours render result

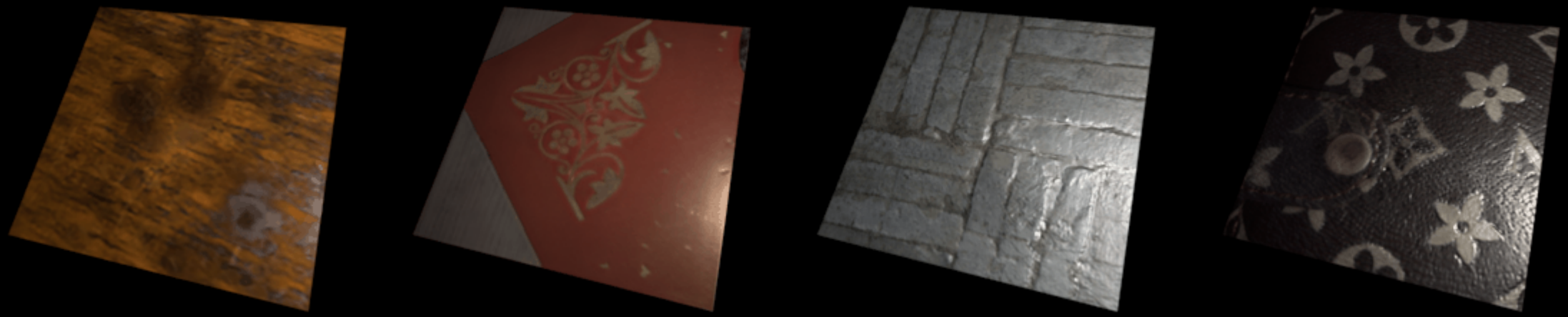
Conclusion

- A well-designed HA convolution layer for hallucinating missing contents in overexposed imageregions.
- A well-designed two-stream network to fully exploit useful features from any casually captured image, facilitating disentangled learning of material properties.



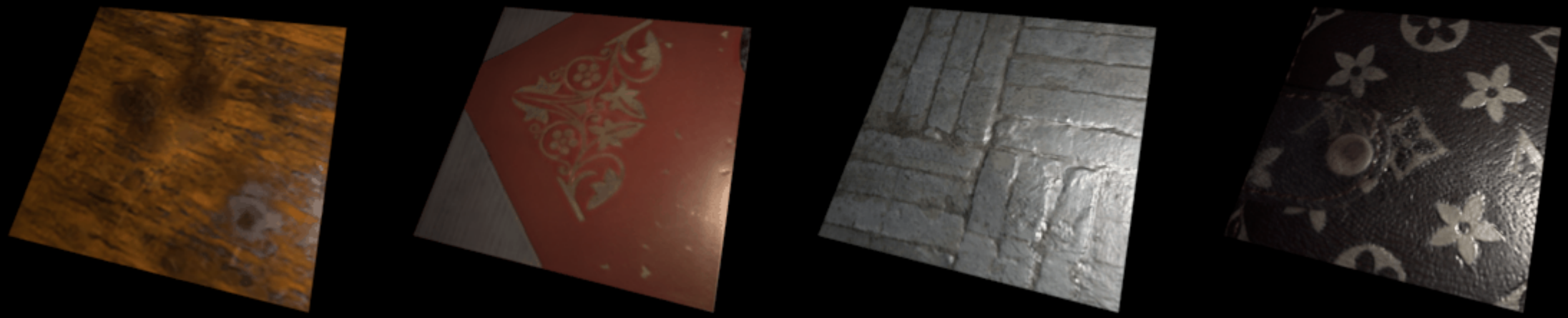
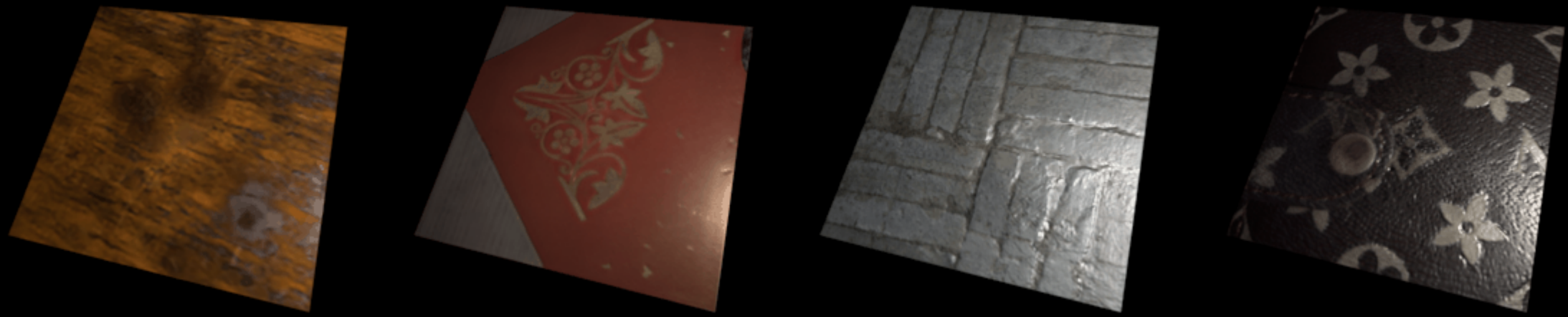
Future Work

- Use a more complex network which could infer the geometry with the help of large-scale datasets containing both geometry and material.
- Extend our method to support more than one input image.



Acknowledgements

- Anonymous reviewers
- Deschaintre et al. [2018], Gao et al. [2019] and Guo et al. [2020] for their help
- National Natural Science Foundation of China (62032011 and 61972194)



Thanks
Q&A