



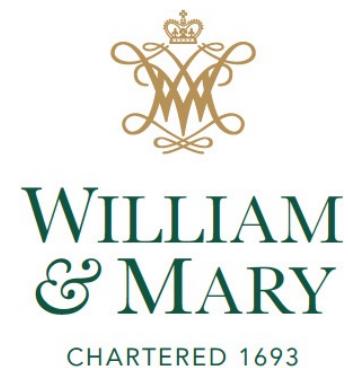
**SIGGRAPH 2025**  
Vancouver+ 10-14 August

# **RenderFormer: Transformer-based Neural Rendering of Triangle Meshes with Global Illumination**

**Chong Zeng, Yue Dong, Pieter Peers, Hongzhi Wu, Xin Tong**

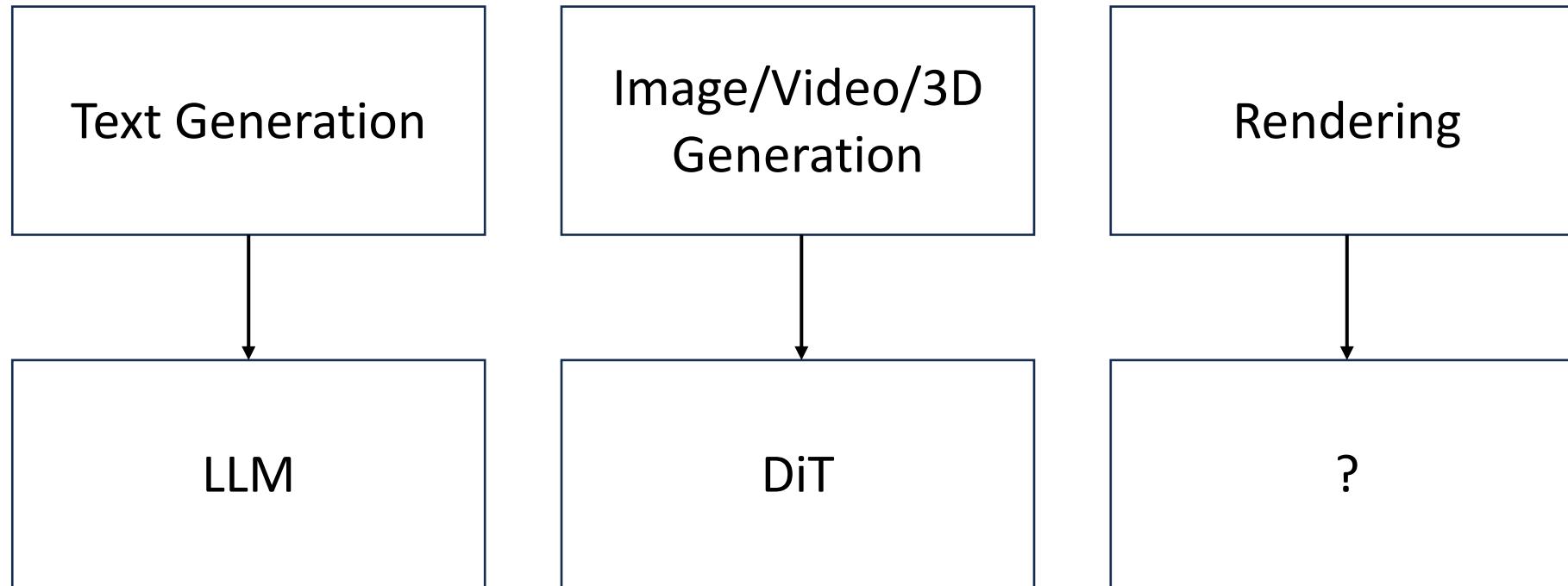


Microsoft  
**Research**  
微软亚洲研究院

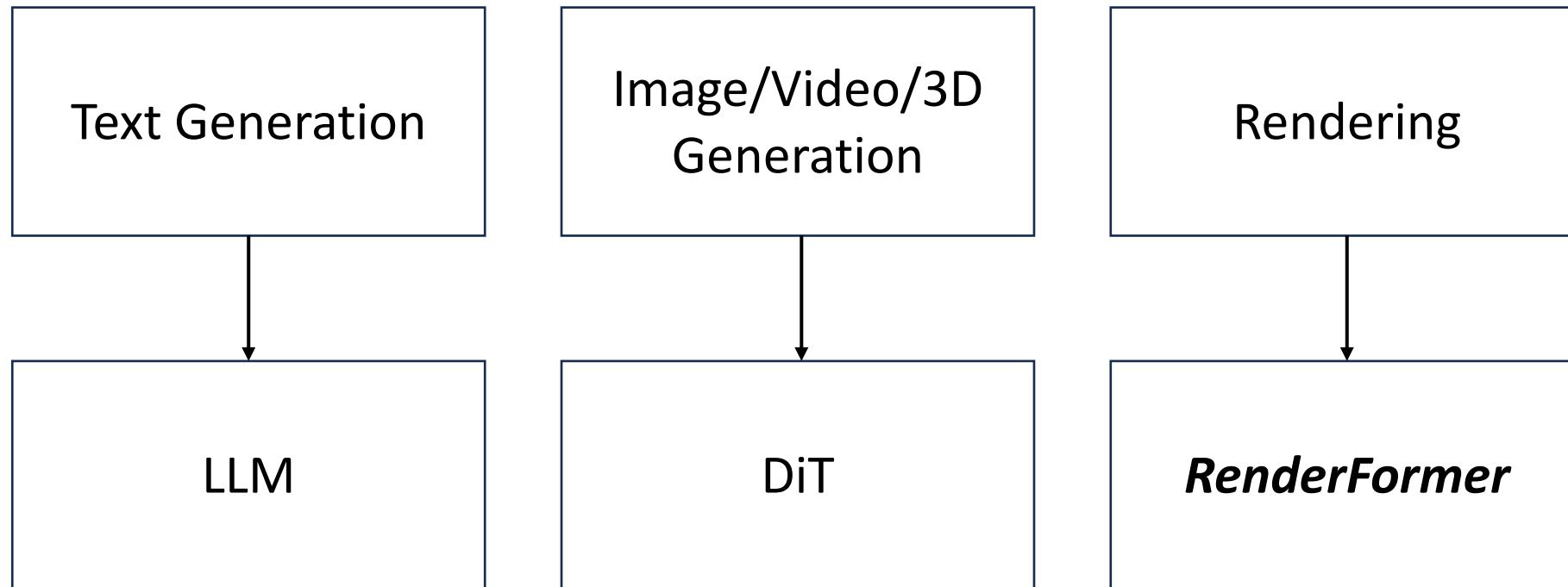


WILLIAM  
& MARY  
CHARTERED 1693

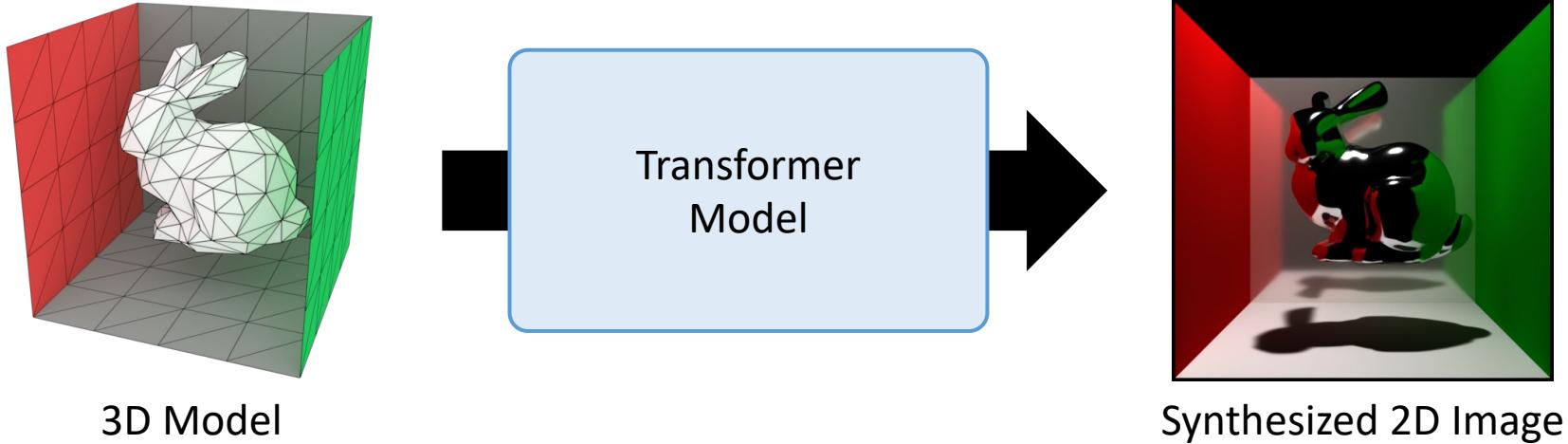
# Attention Is All You Need!



# Attention Is All You Need for Rendering



# RenderFormer: A *Fully Transformer* Rendering Pipeline



End-to-end with Raw 3D Input

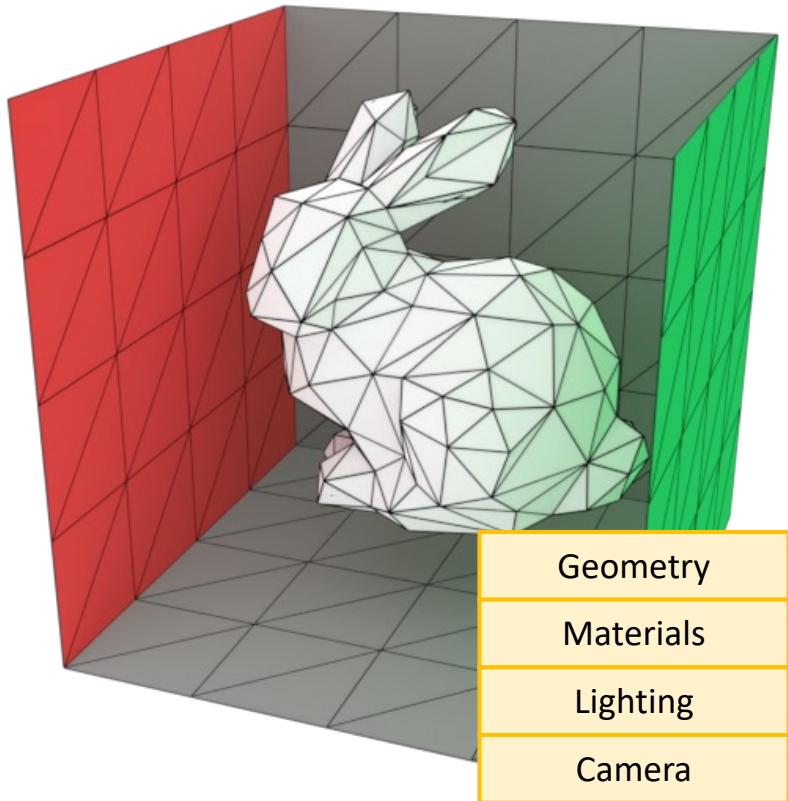
No Per-scene Training

Full Global Illumination Effects

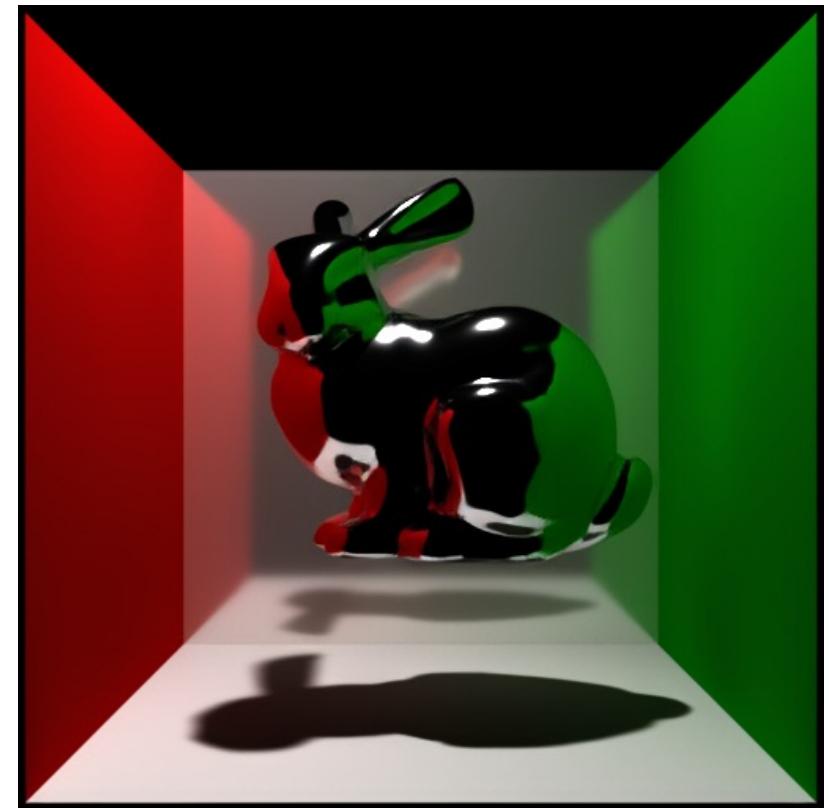
Minimal Prior Constraints

# Definition of 3D Rendering

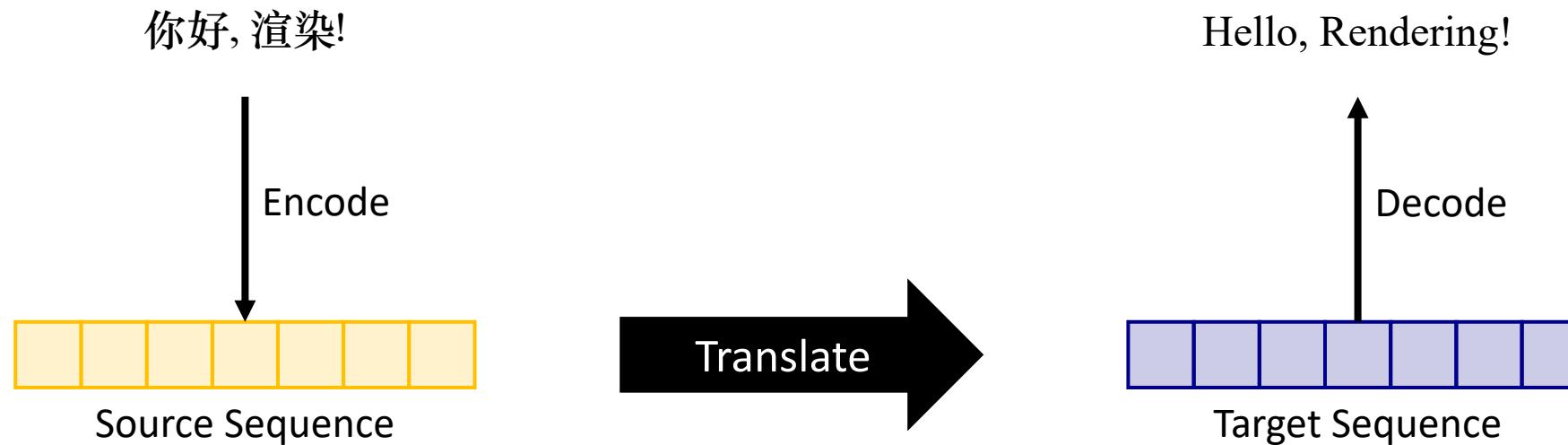
3D Model



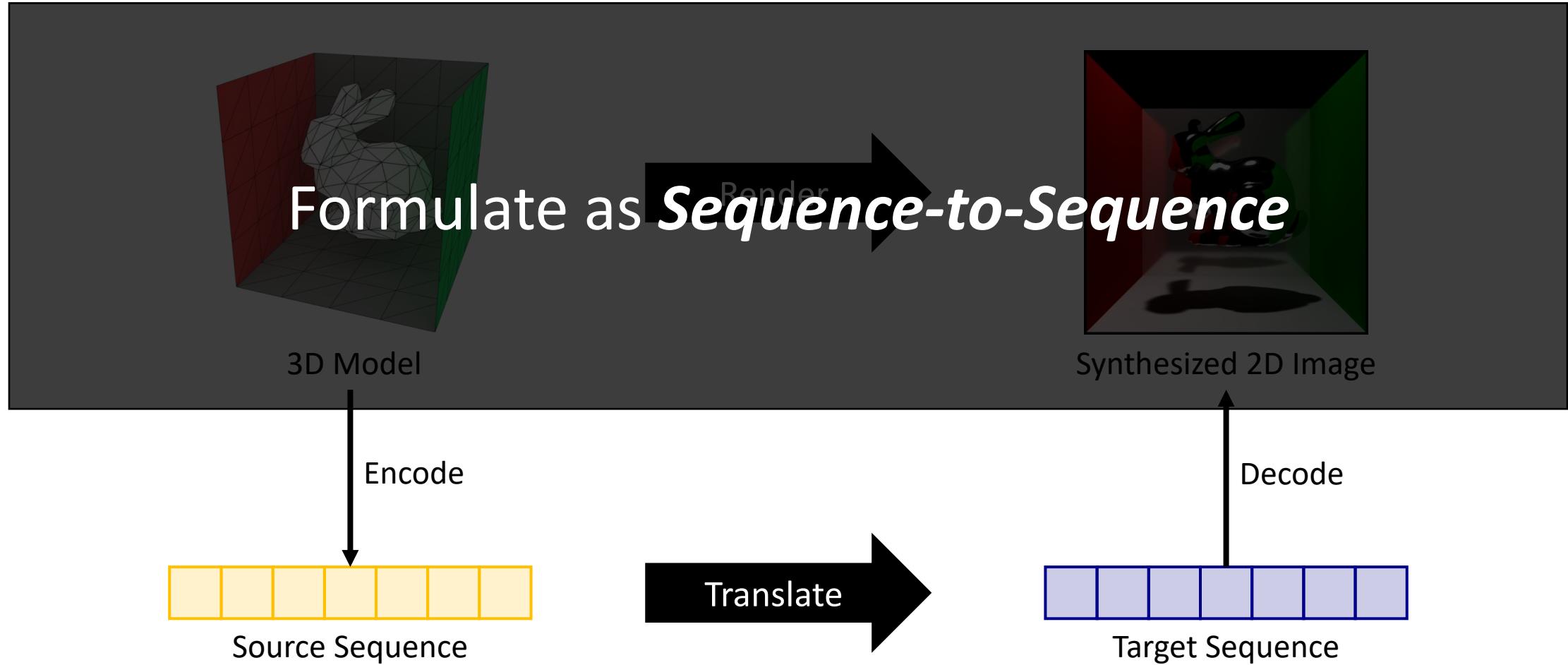
Synthesized 2D Image



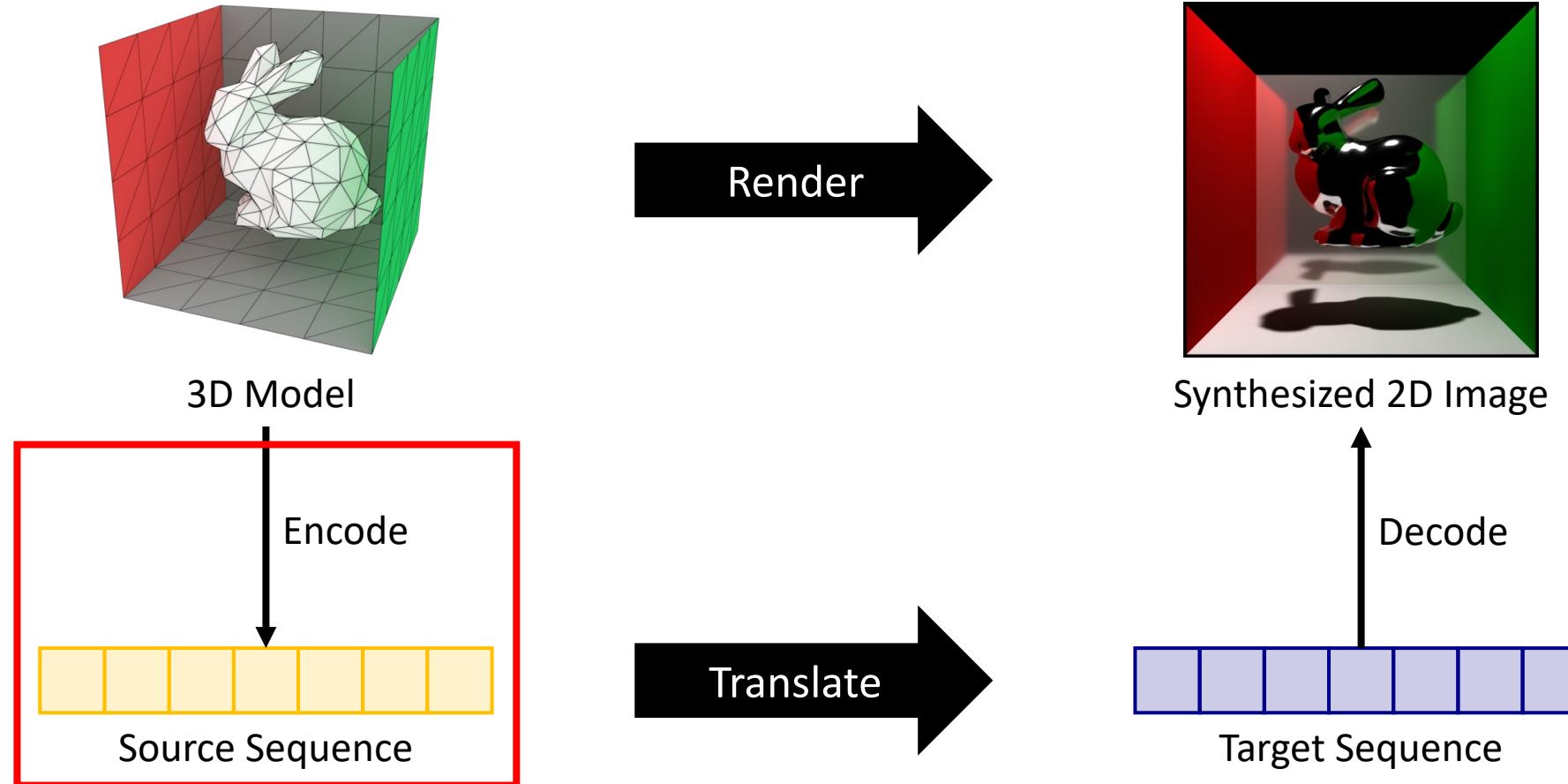
# Sequence-to-Sequence Machine Translation



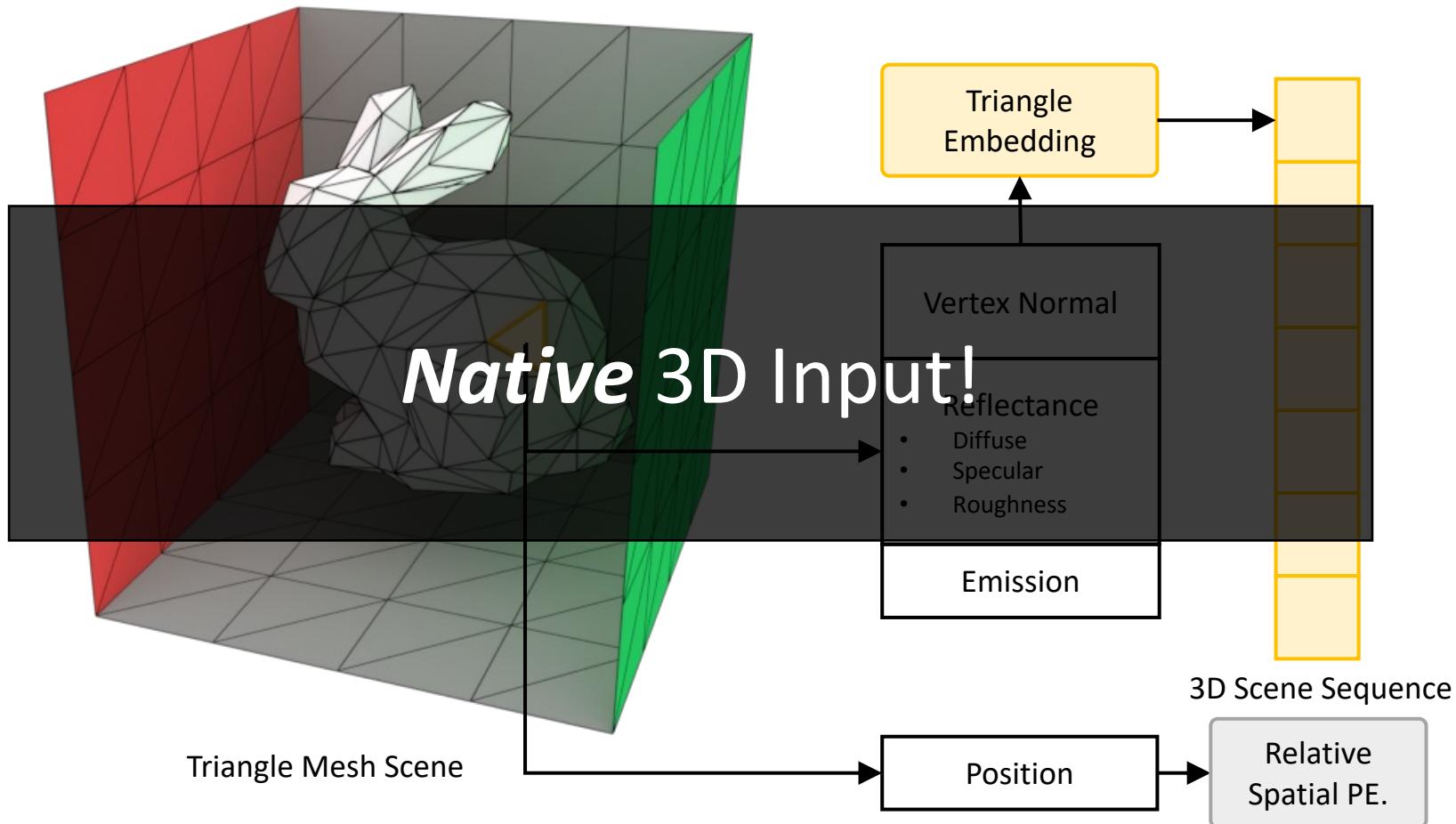
Idea: 3D Rendering = Translating 3D to 2D



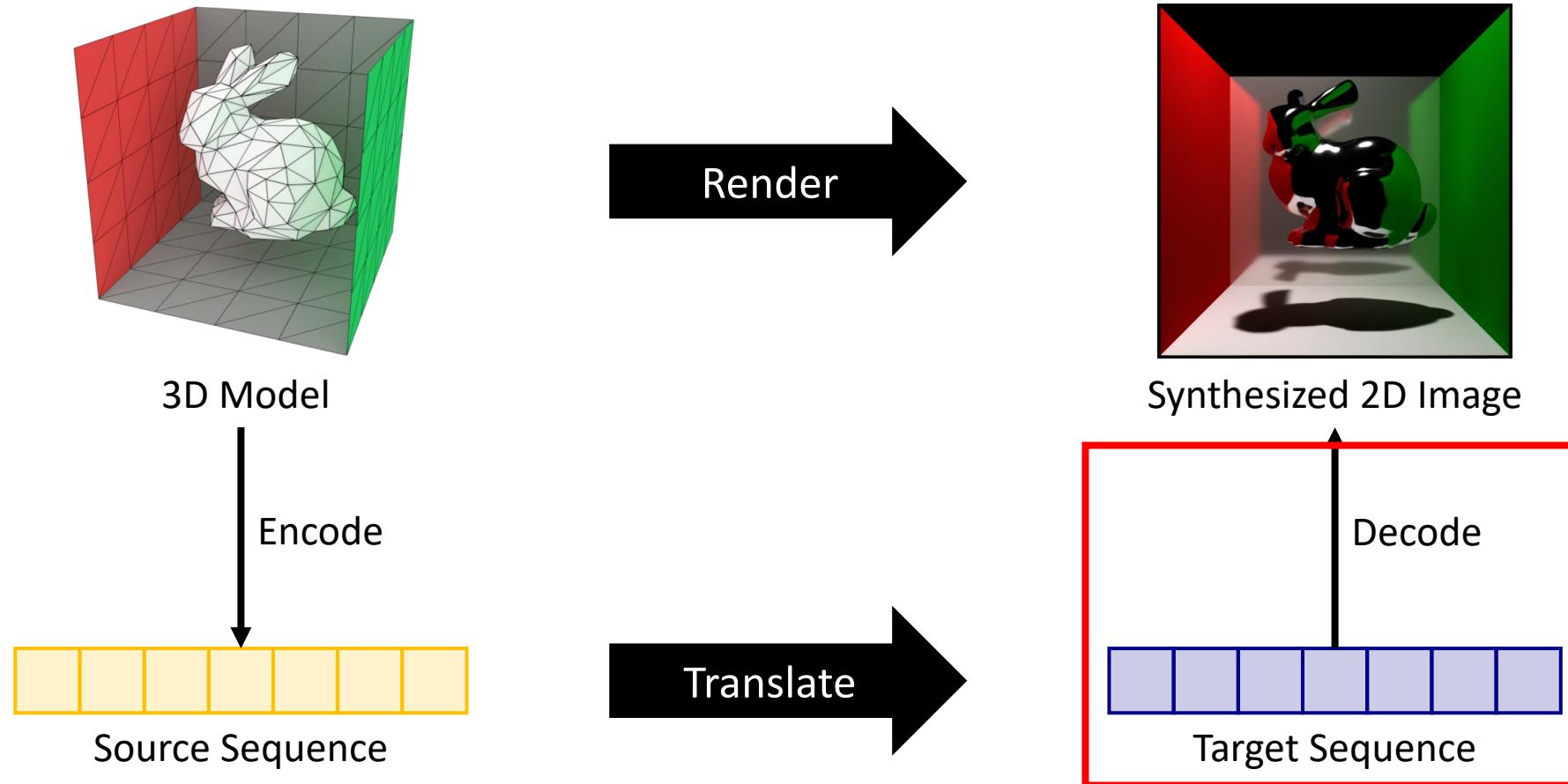
# Idea: 3D Rendering = Translating 3D to 2D



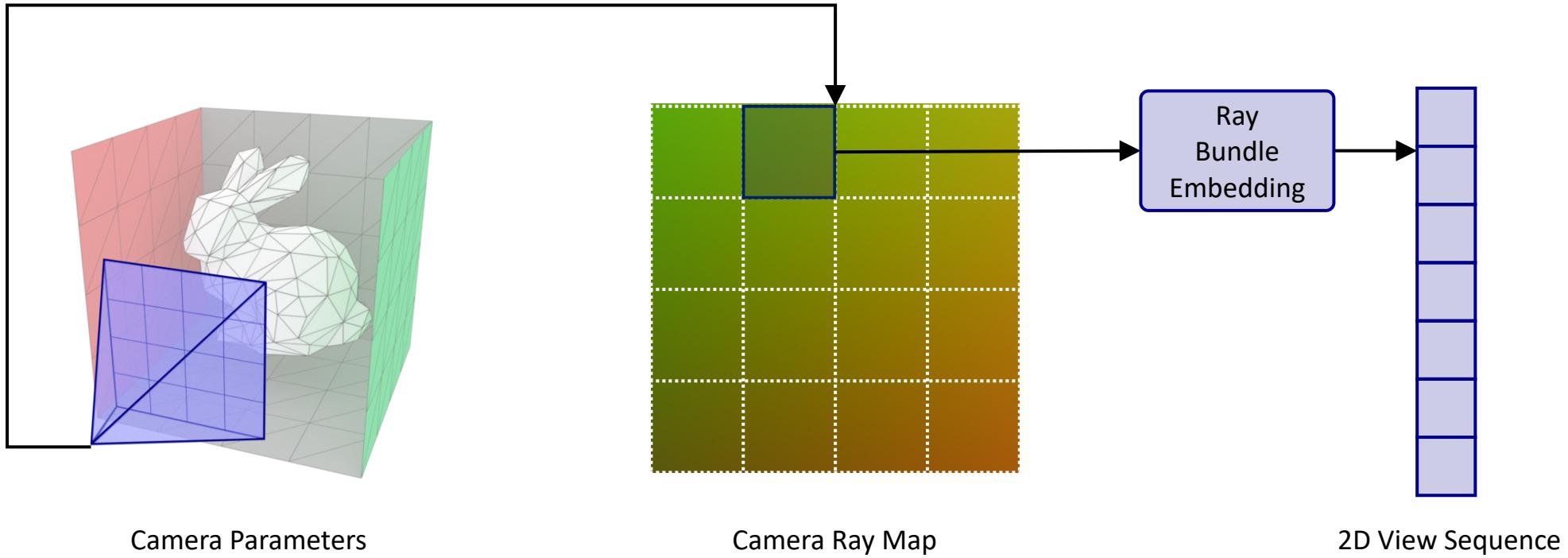
# 3D Sequence: Tokenize Mesh Scene by Triangles



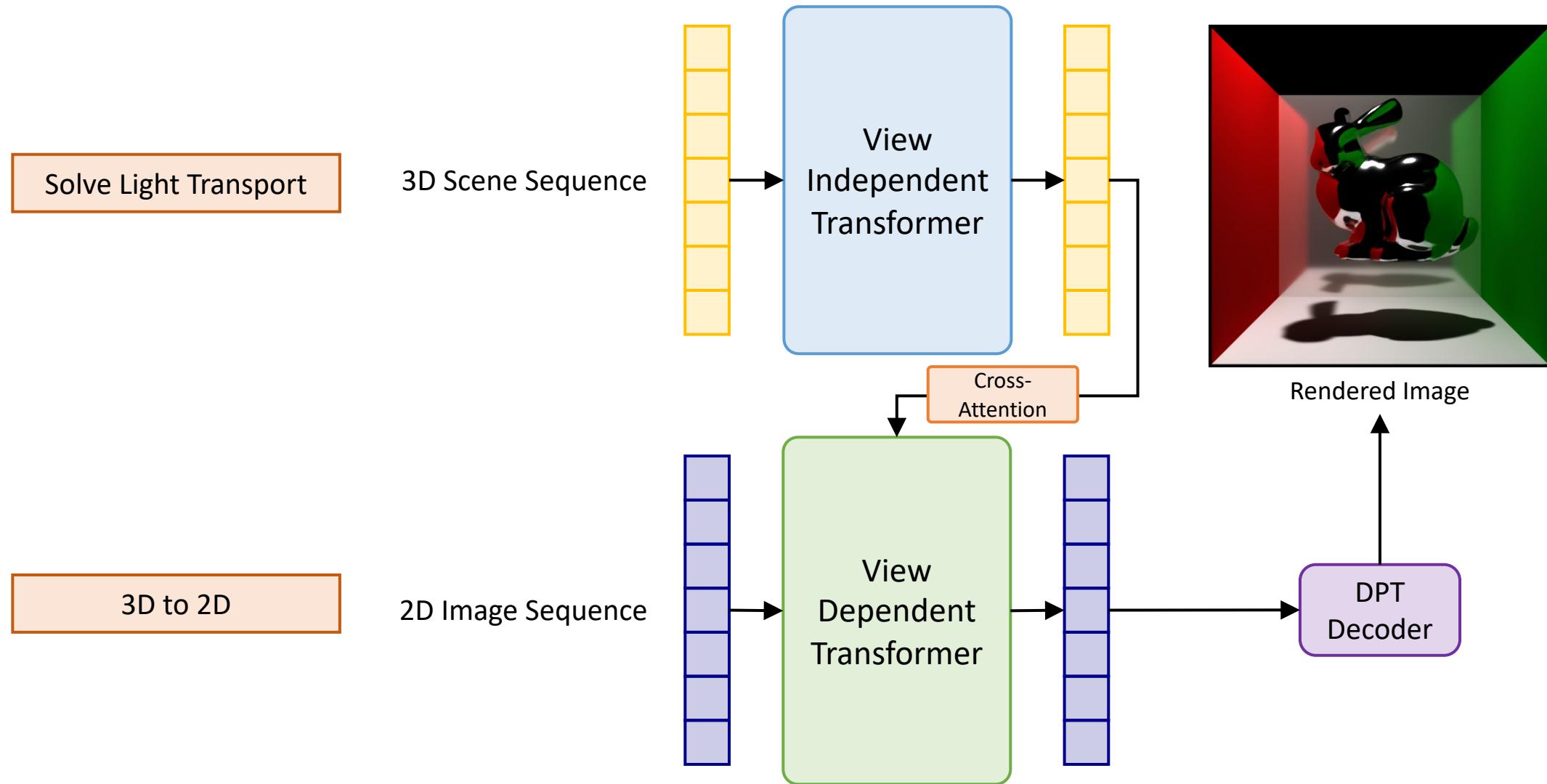
# Idea: 3D Rendering = Translating 3D to 2D



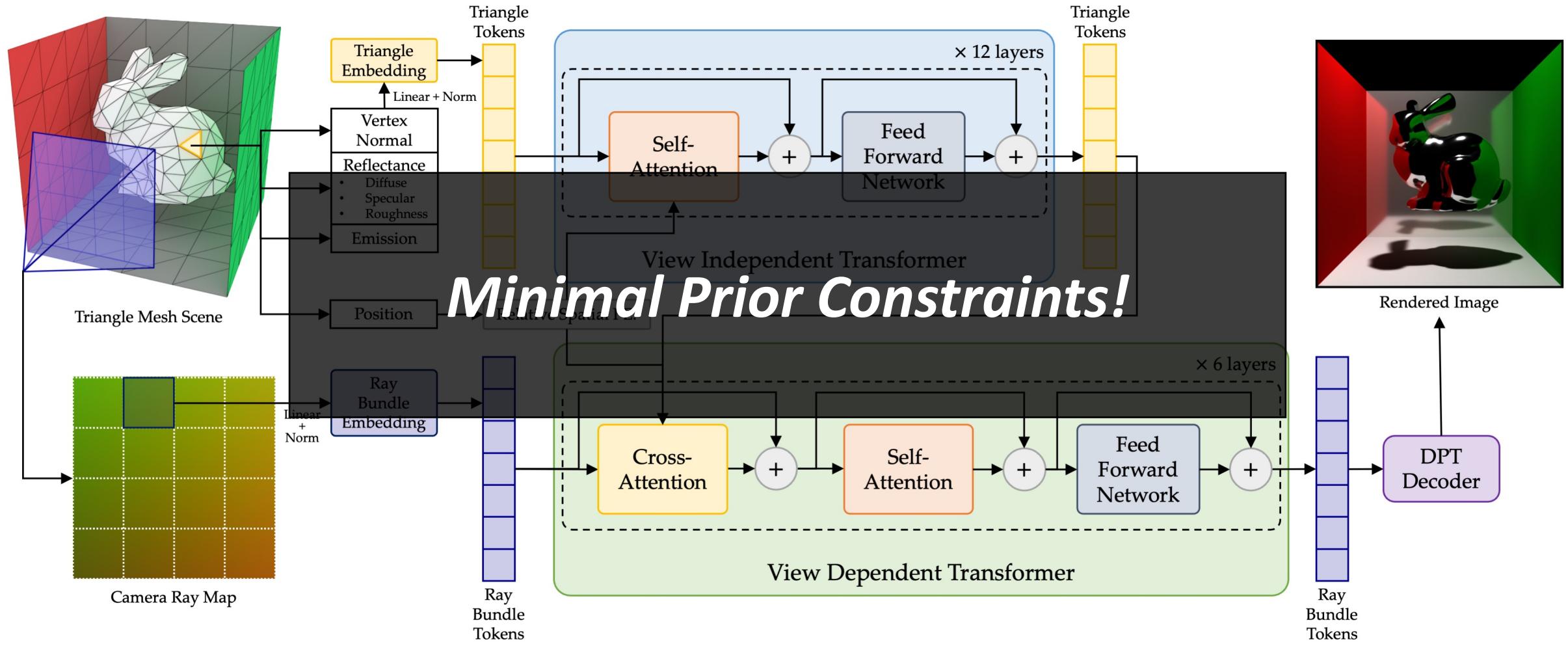
# 2D Sequence: Tokenize View using Ray Bundles



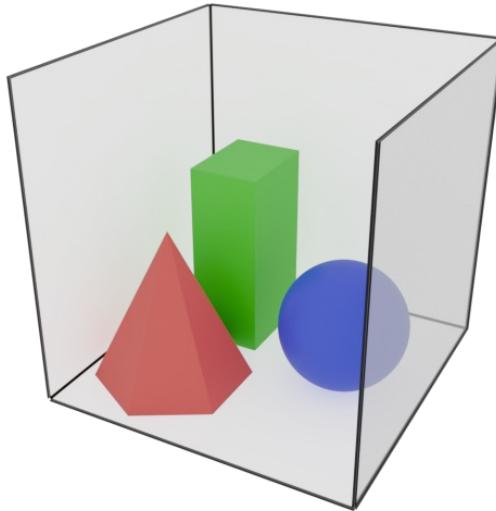
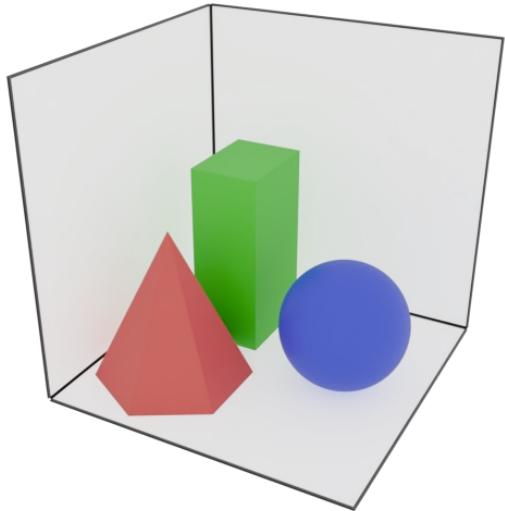
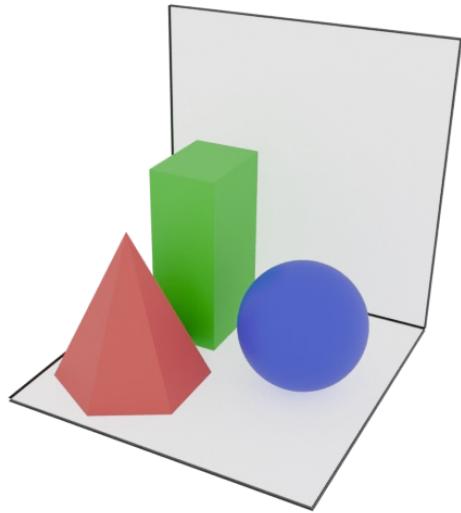
# Rendering with Transformer Architecture



# Full Pipeline



# Training Data – Template-based Scene Generation



**4,096**

Max #Triangles

**0.01-1.0**

Roughness Range

**8**

Max #Lights

**2.1-2.7**

Light Distance

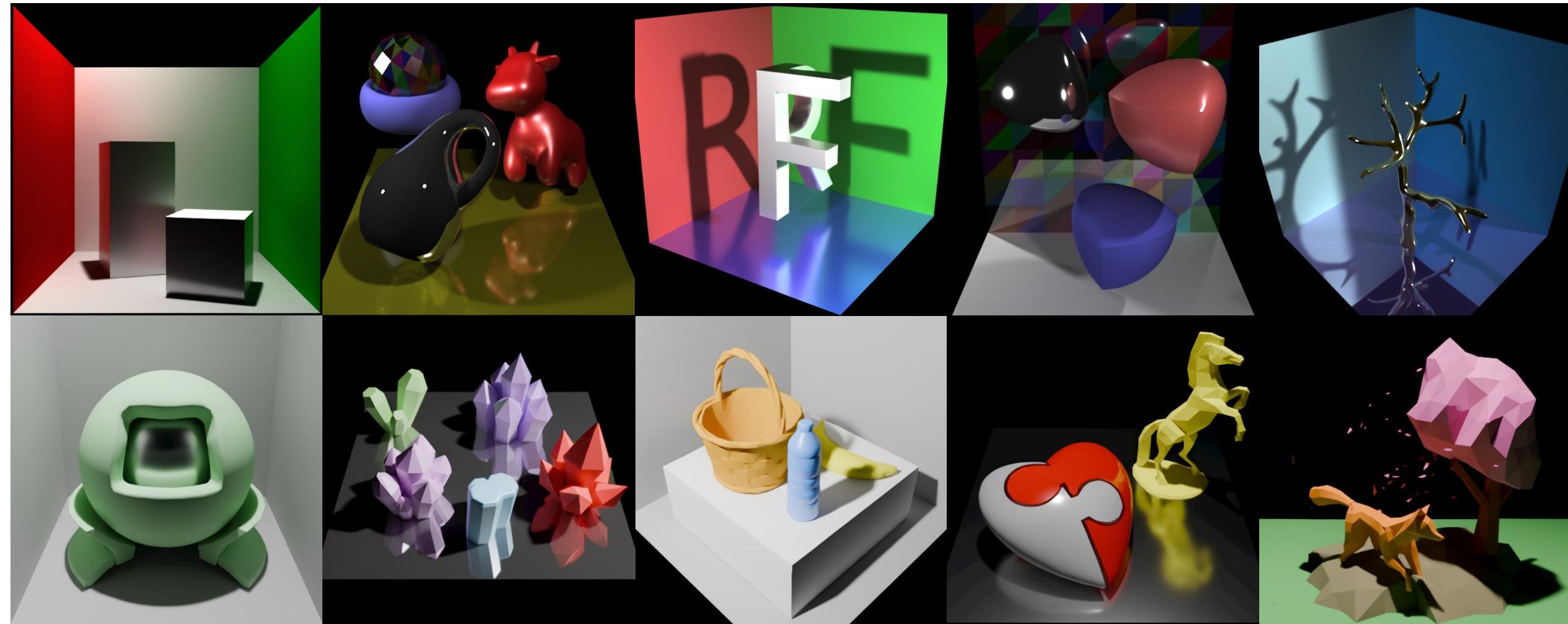
**30-60**

Camera FOV

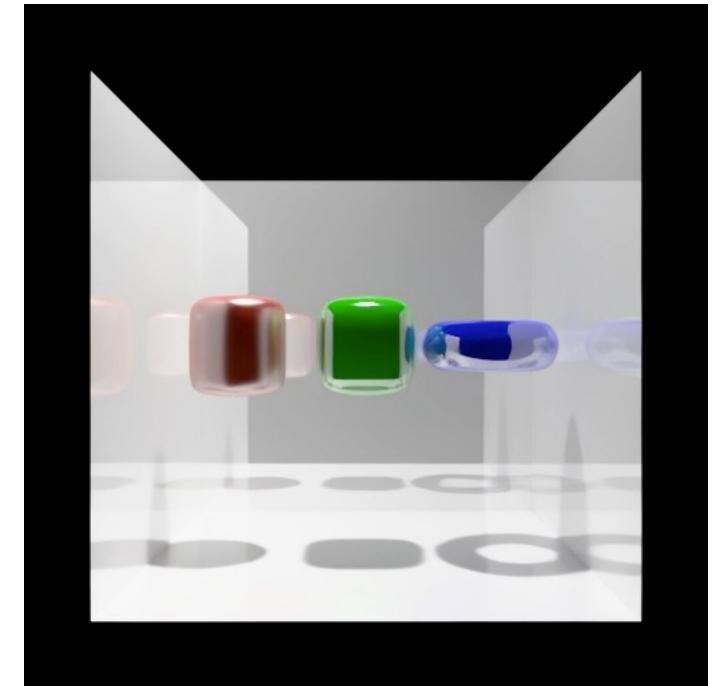
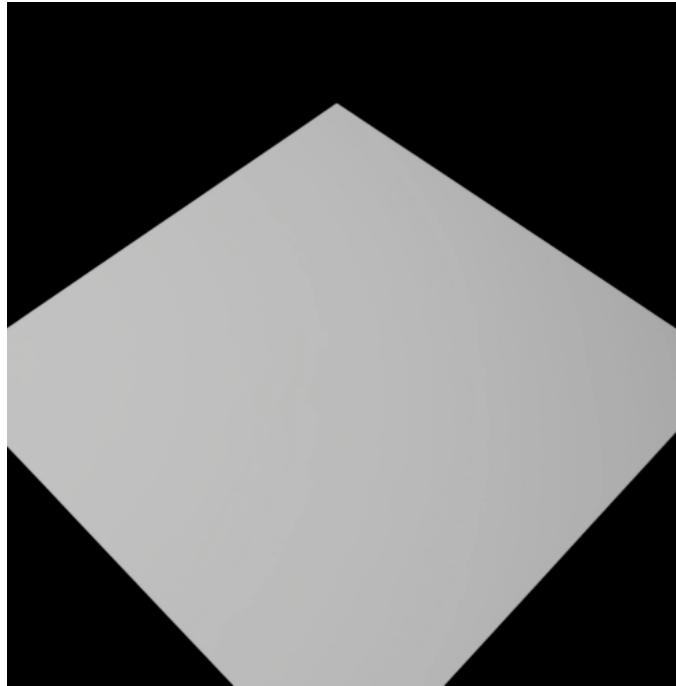
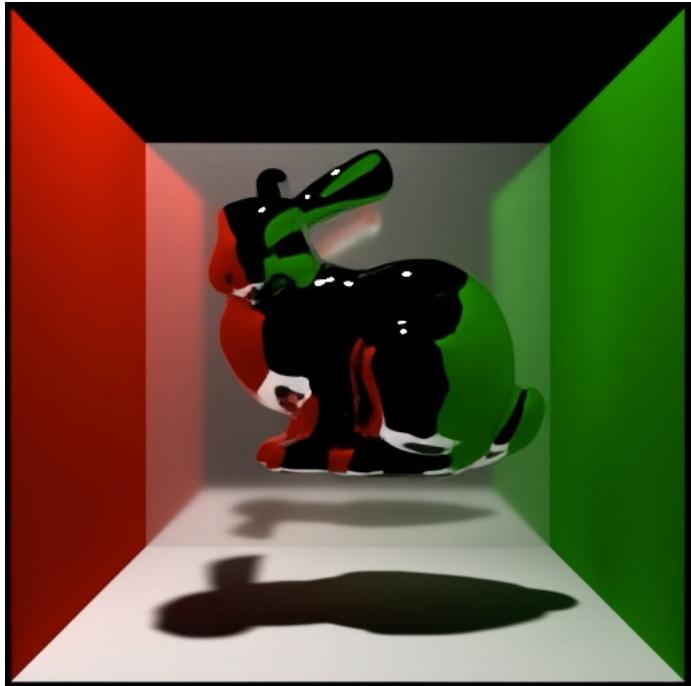
**1.5-2.0**

Camera Distance

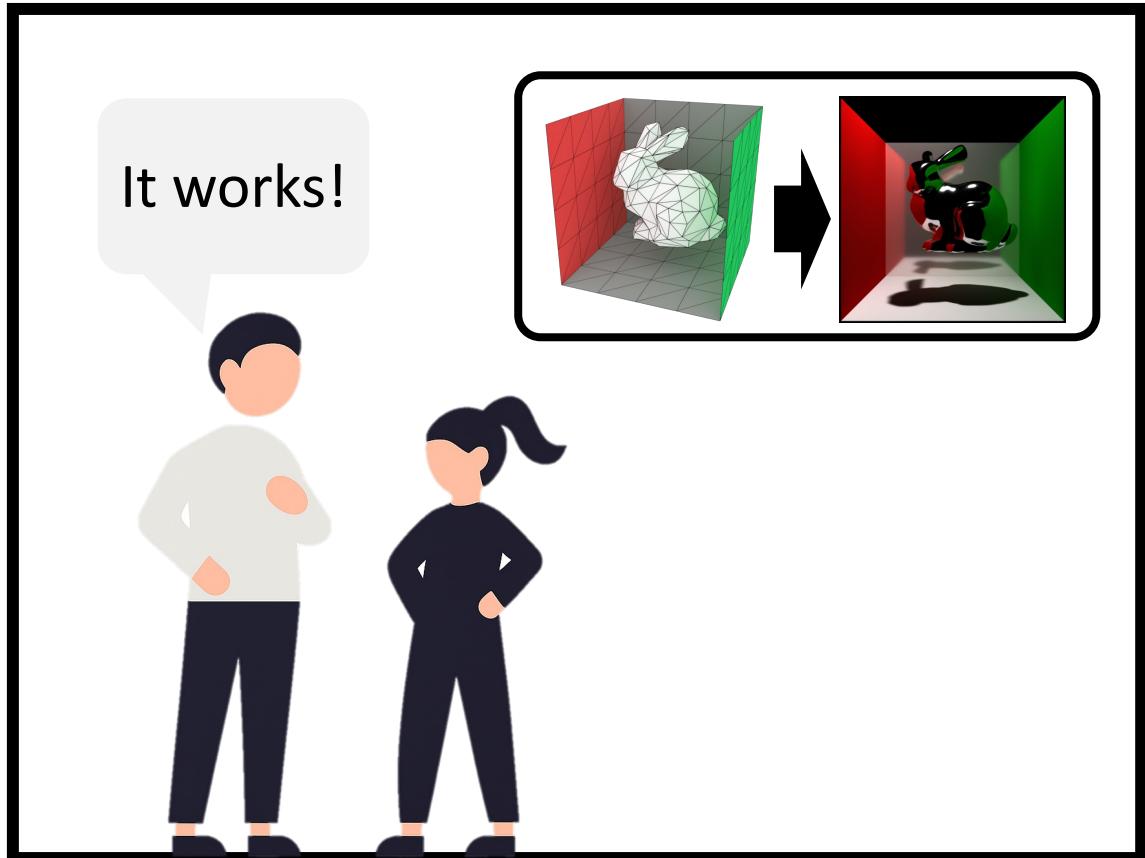
# Results – Static Scene Rendering



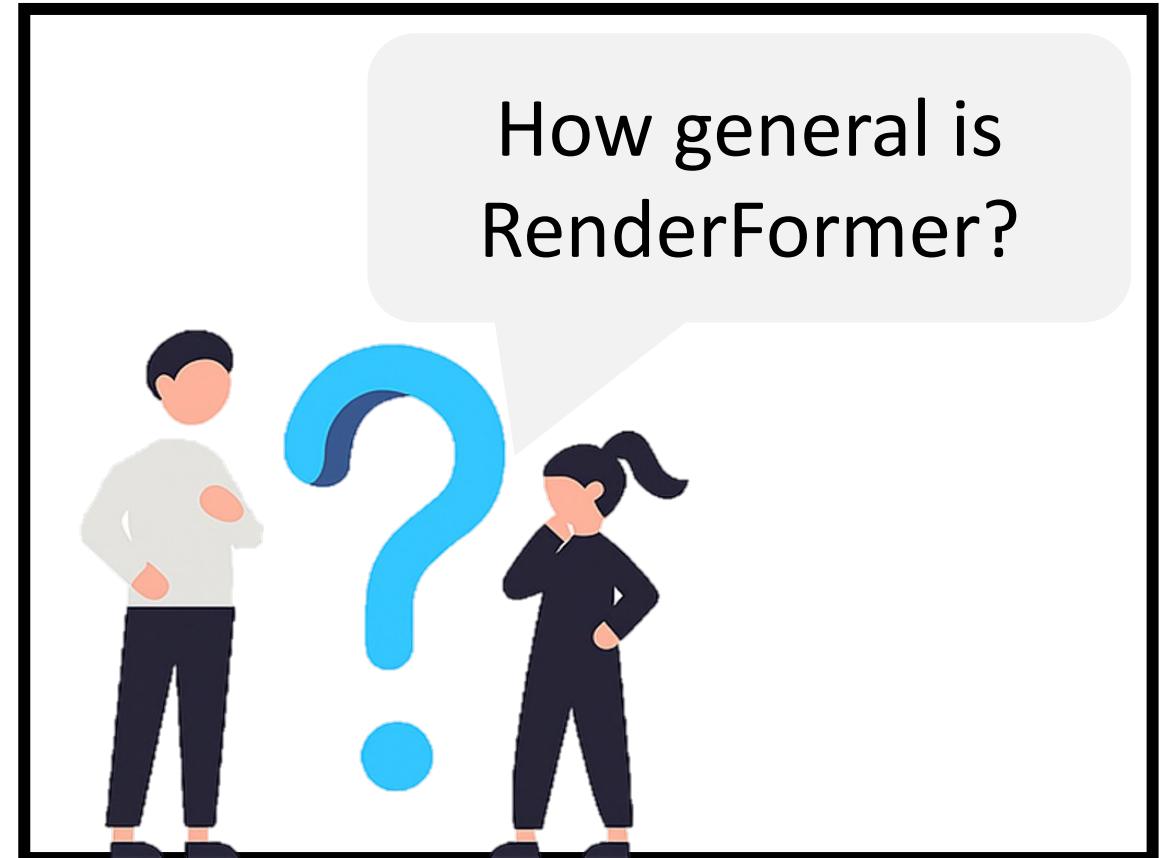
# Results – Dynamic Scene Rendering



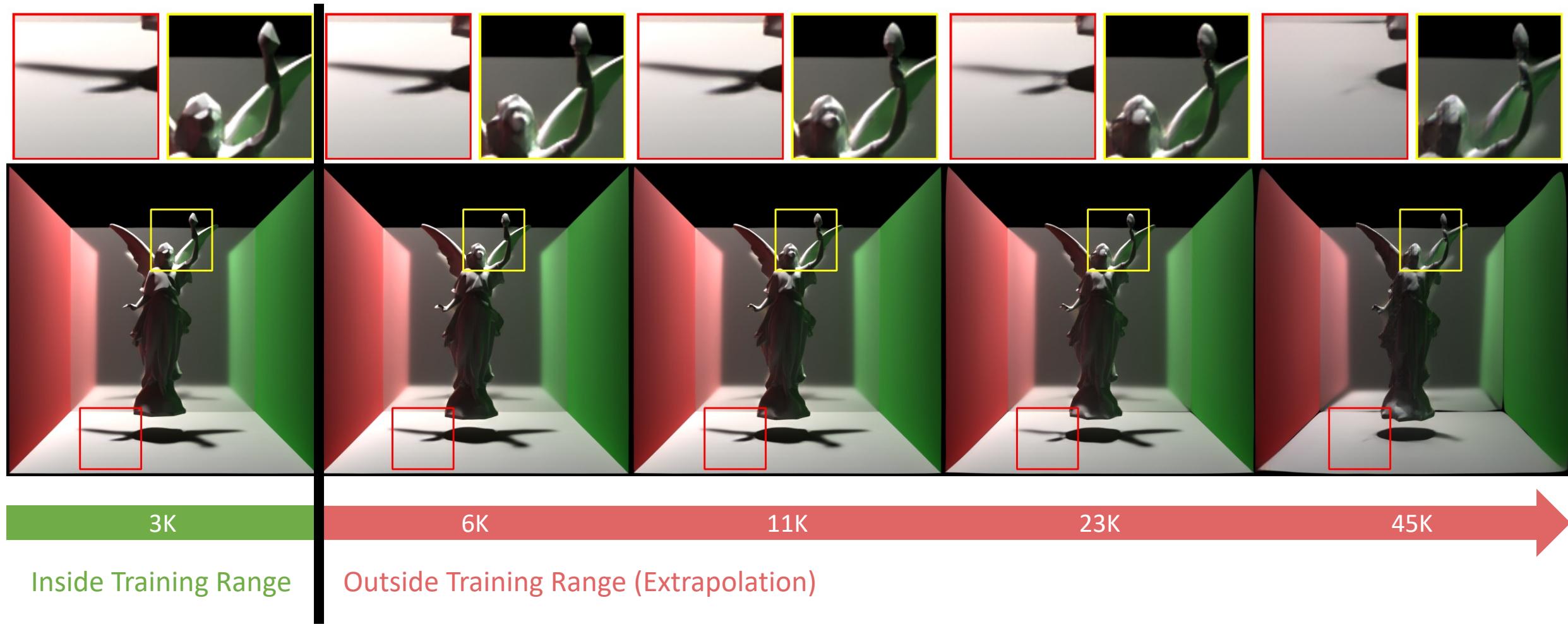
# YES,



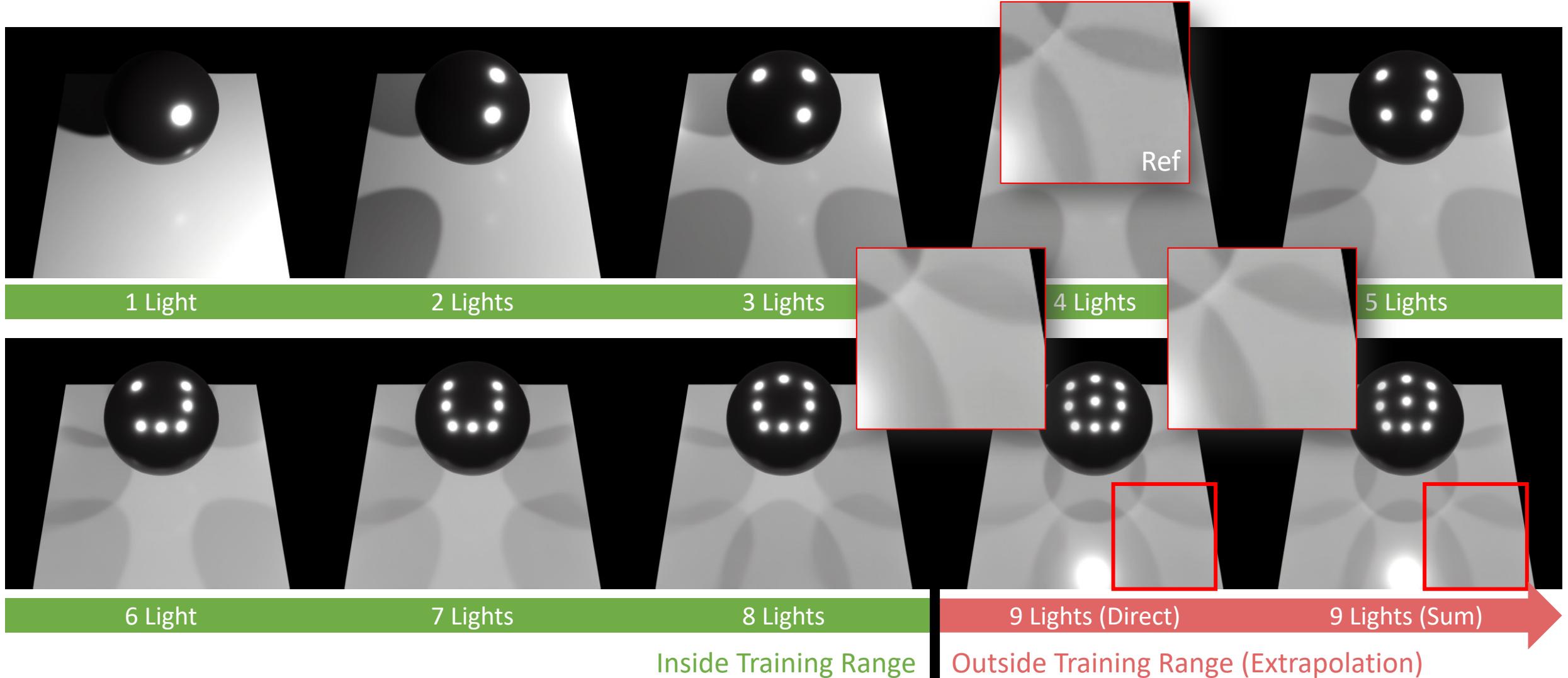
# BUT



# Generalization – Number of Triangles

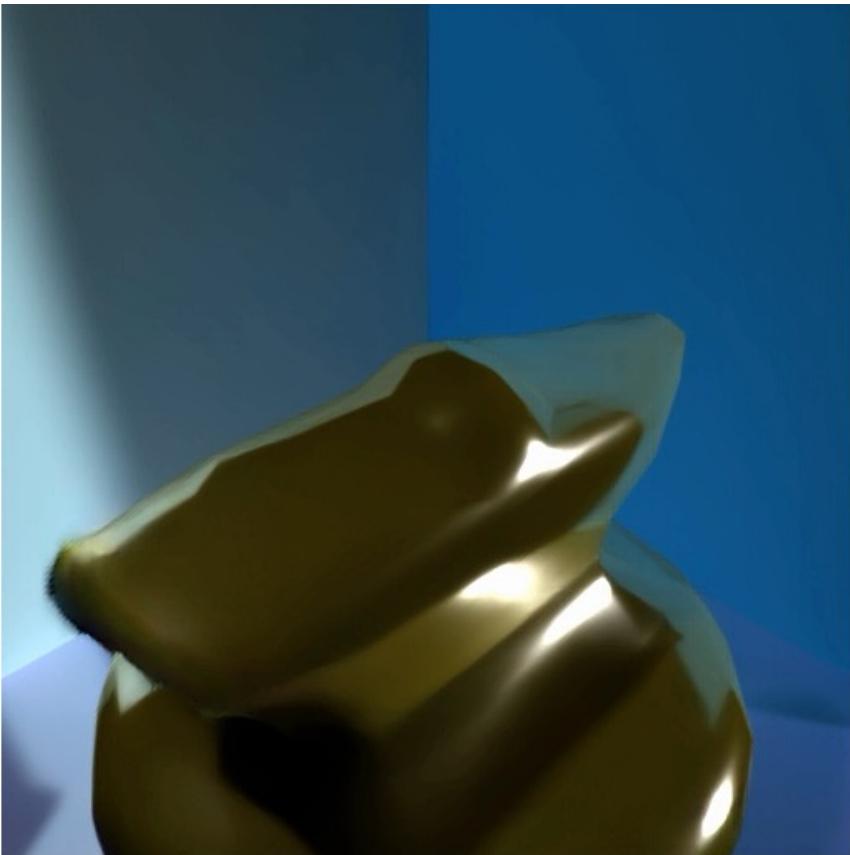


# Generalization – Lighting Number

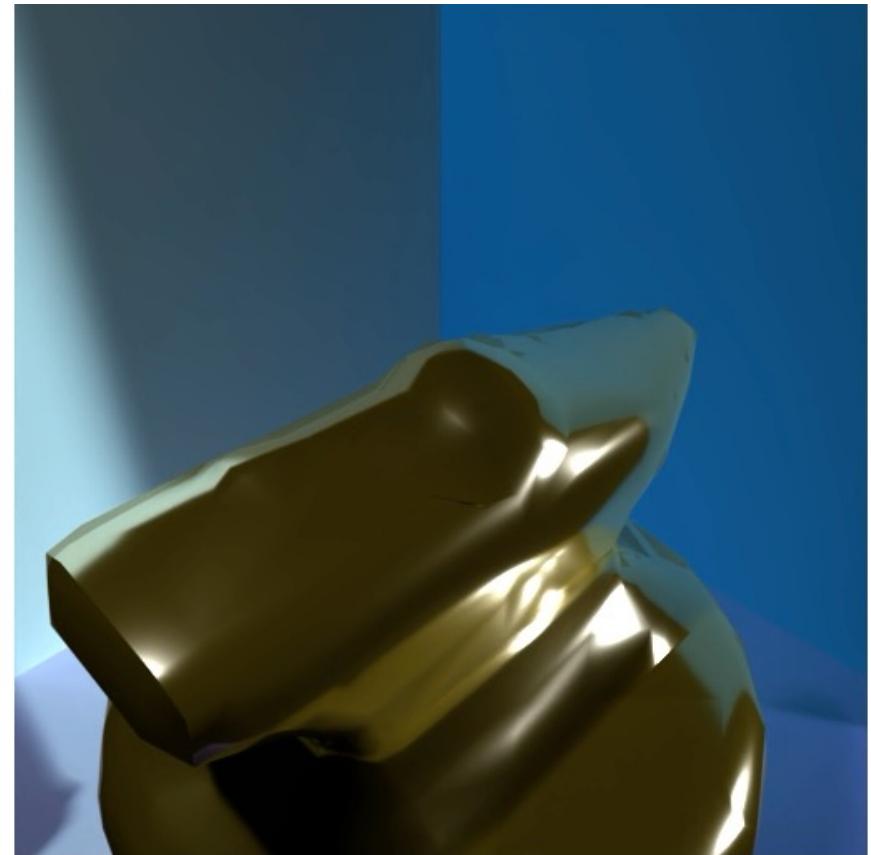


# Generalization – Camera Distance

RenderFormer

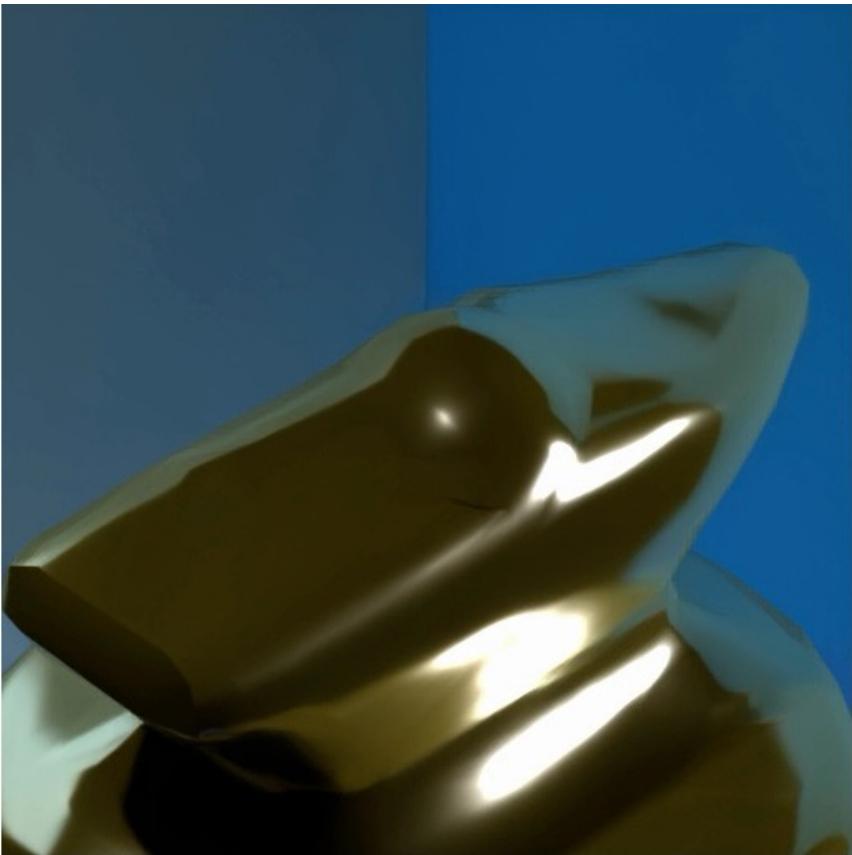


Reference

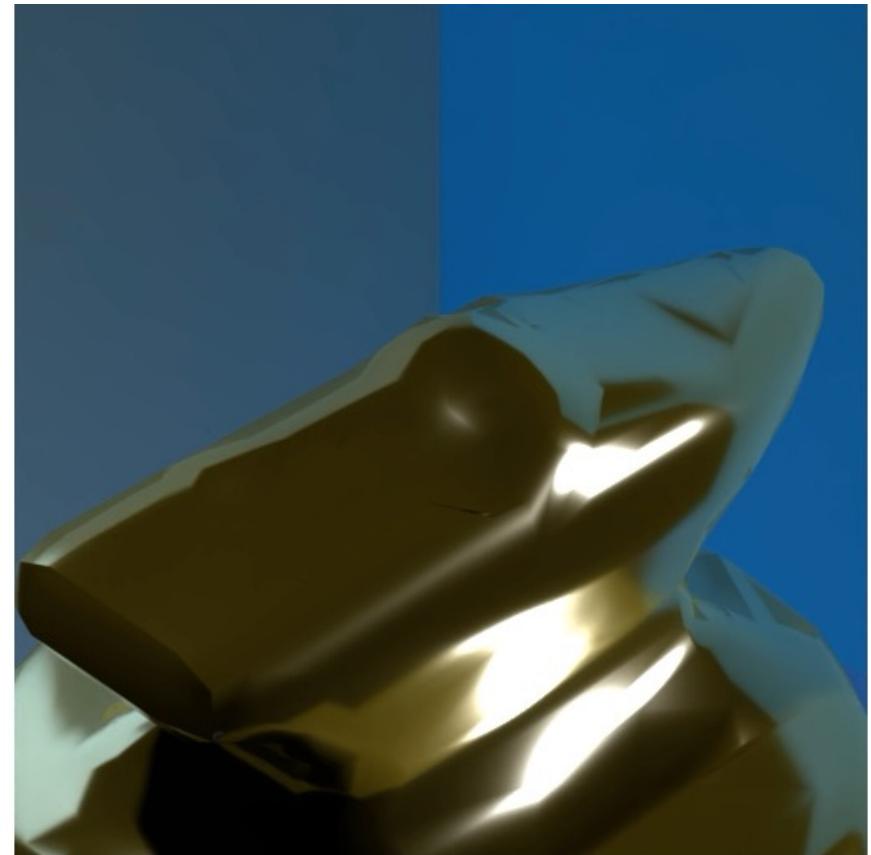


# Generalization – Camera FOV

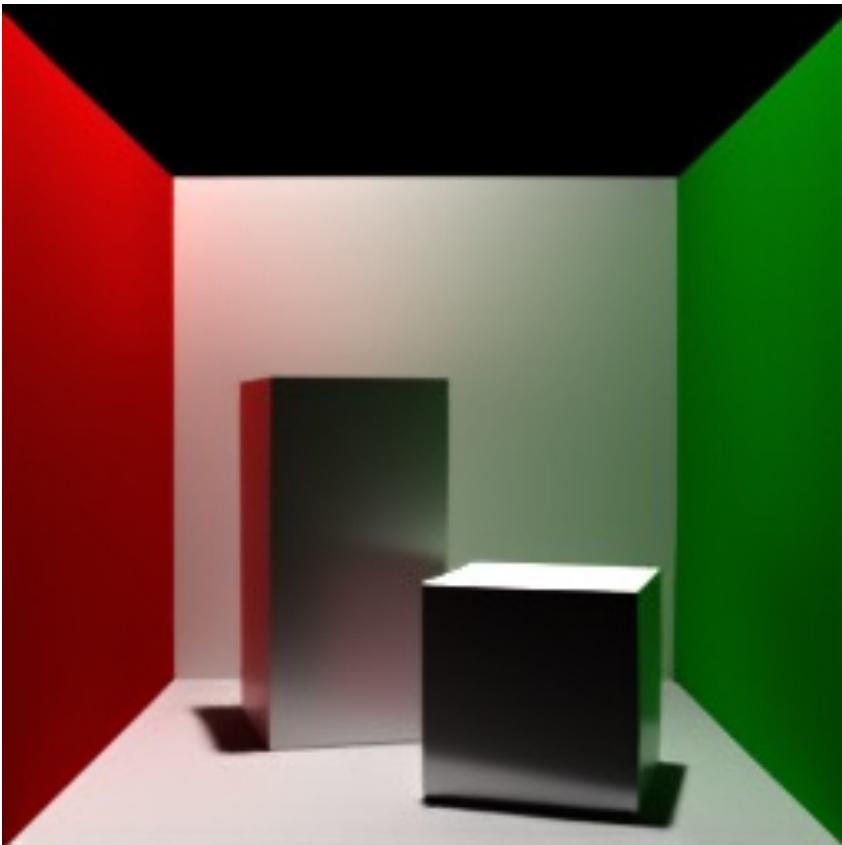
RenderFormer



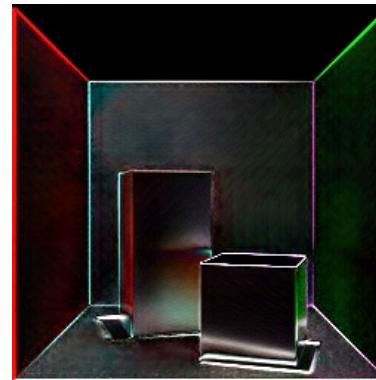
Reference



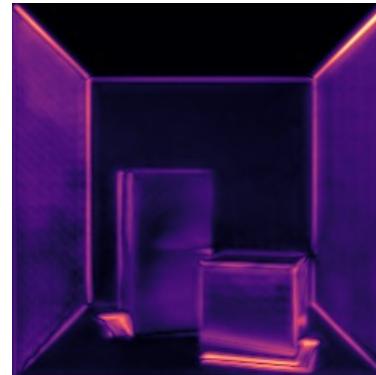
# Generalization – Resolution



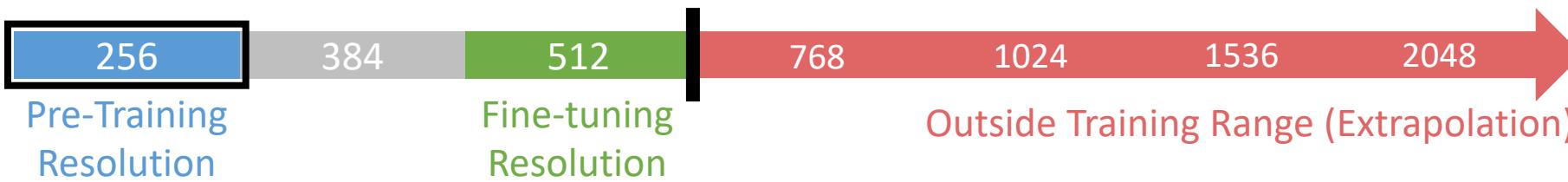
Rendered Image



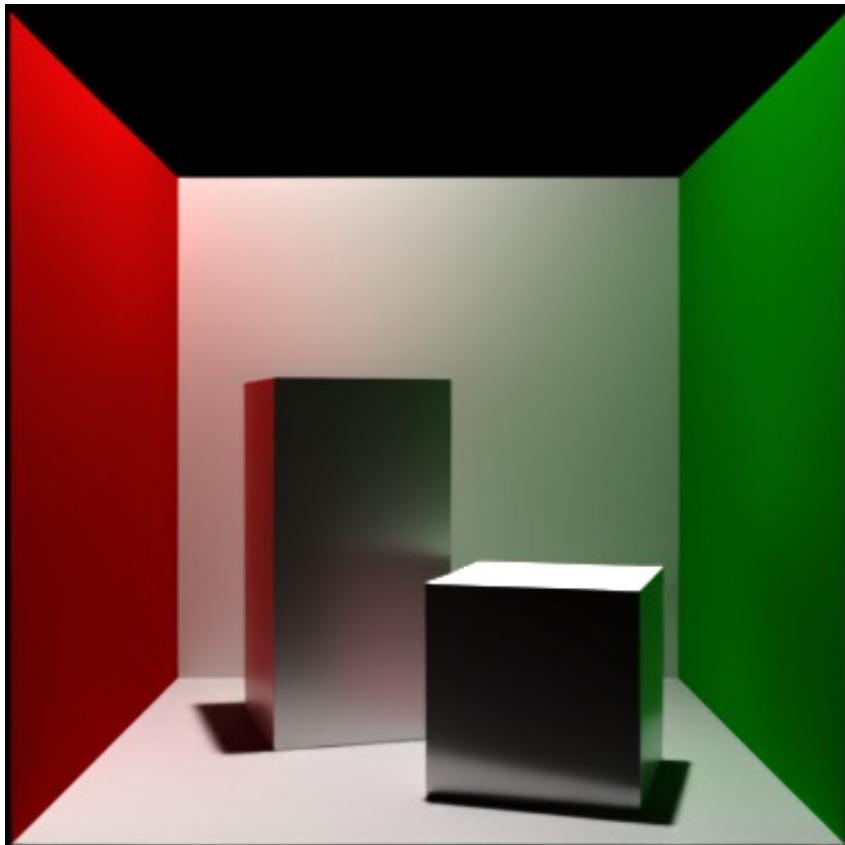
Error Image (x5)



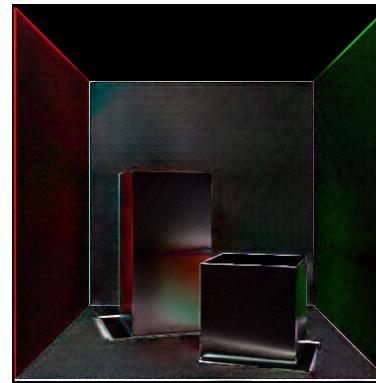
FLIP Error



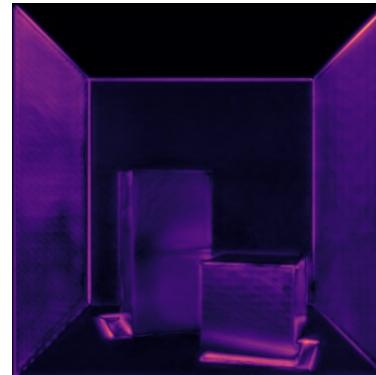
# Generalization – Resolution



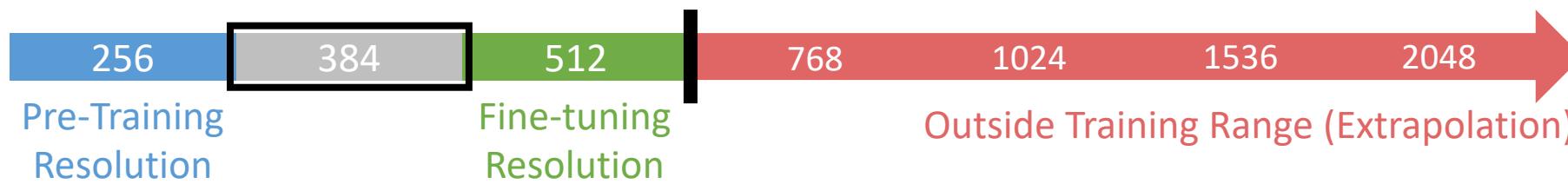
Rendered Image



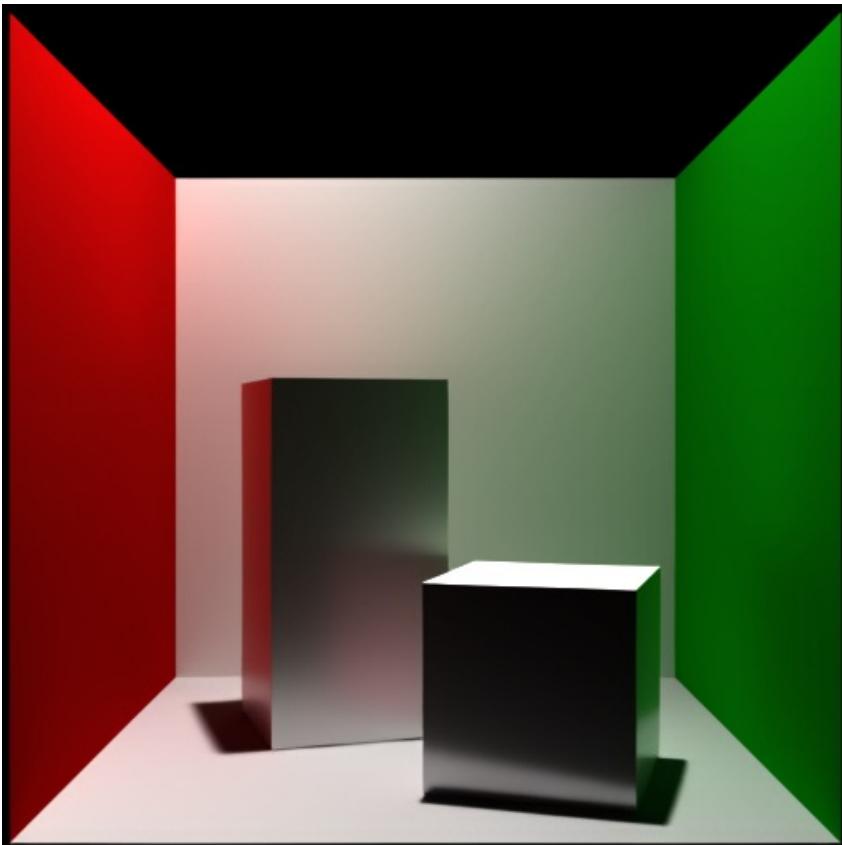
Error Image (x5)



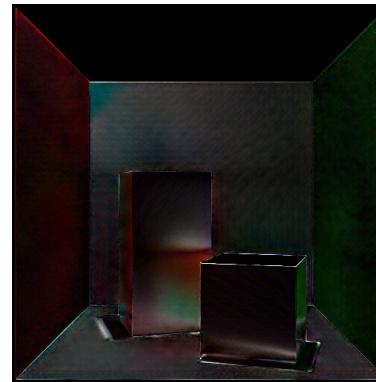
FLIP Error



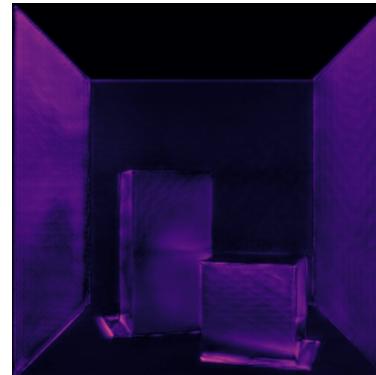
# Generalization – Resolution



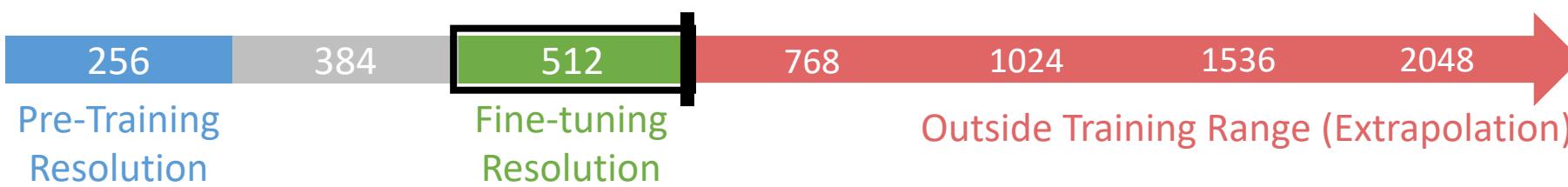
Rendered Image



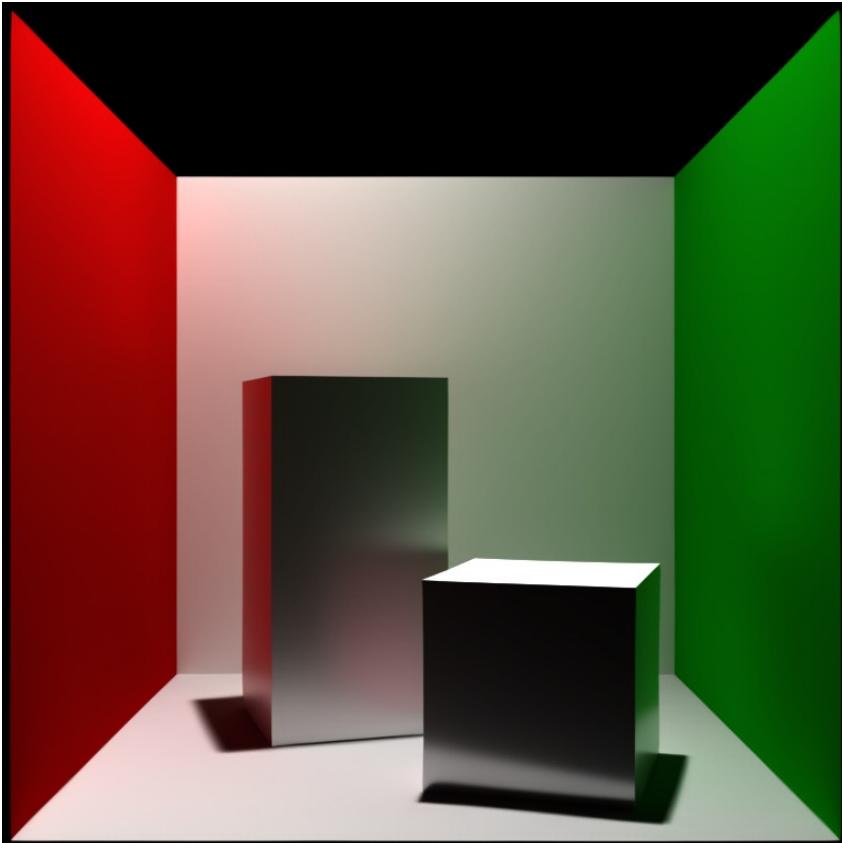
Error Image (x5)



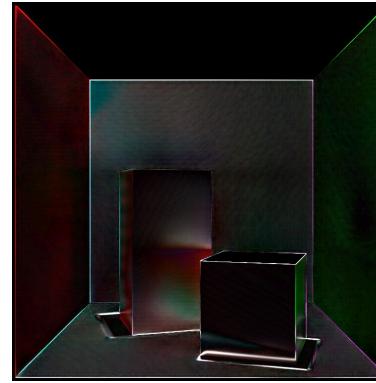
FLIP Error



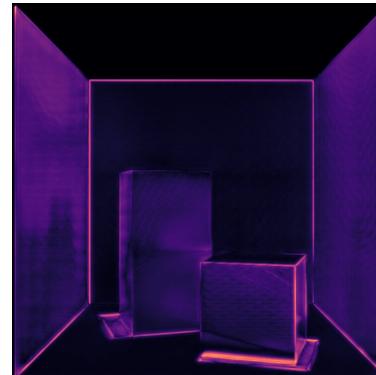
# Generalization – Resolution



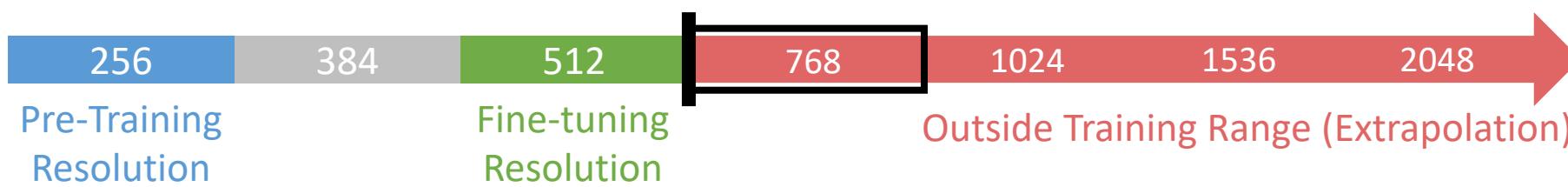
Rendered Image



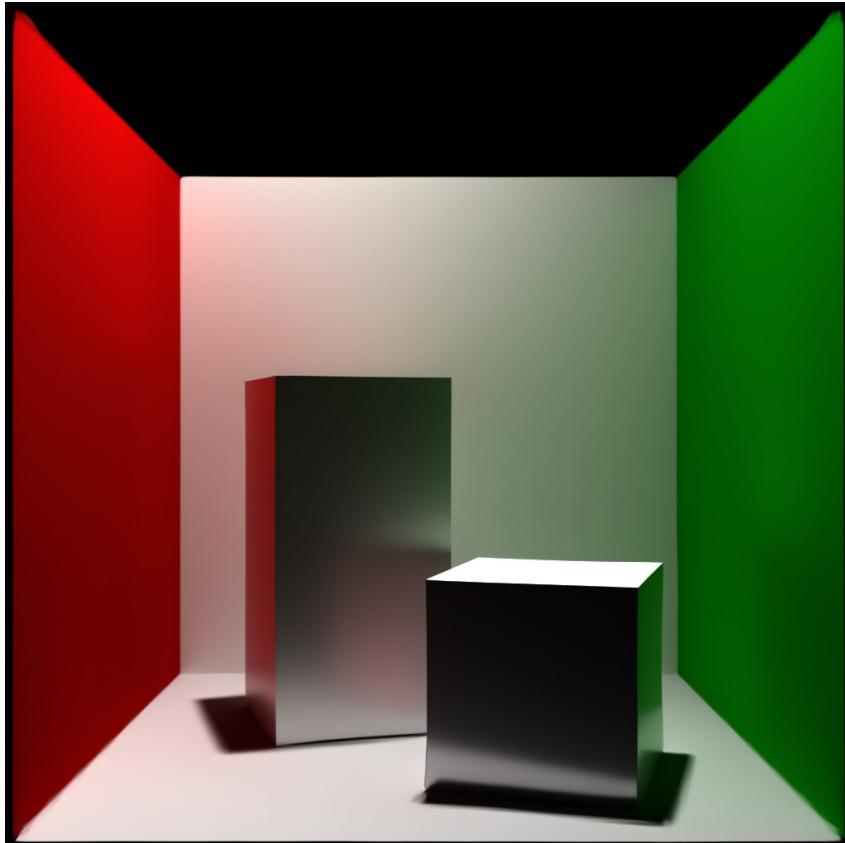
Error Image (x5)



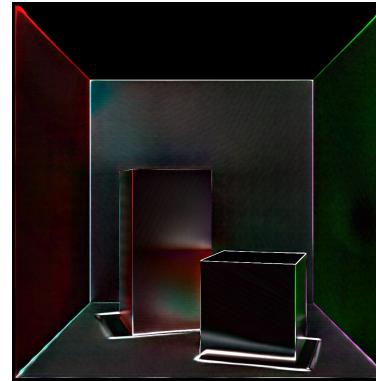
FLIP Error



# Generalization – Resolution



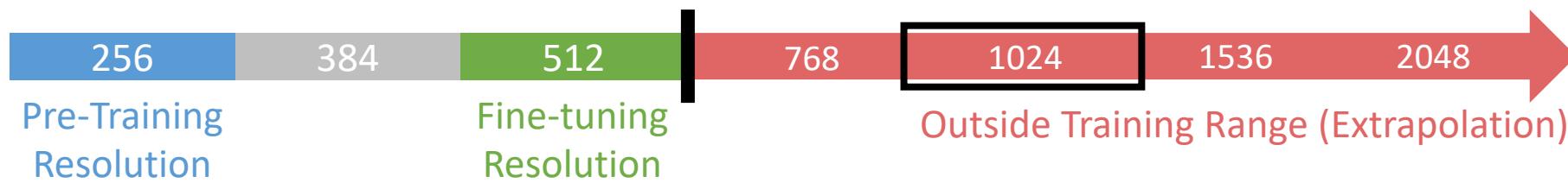
Rendered Image



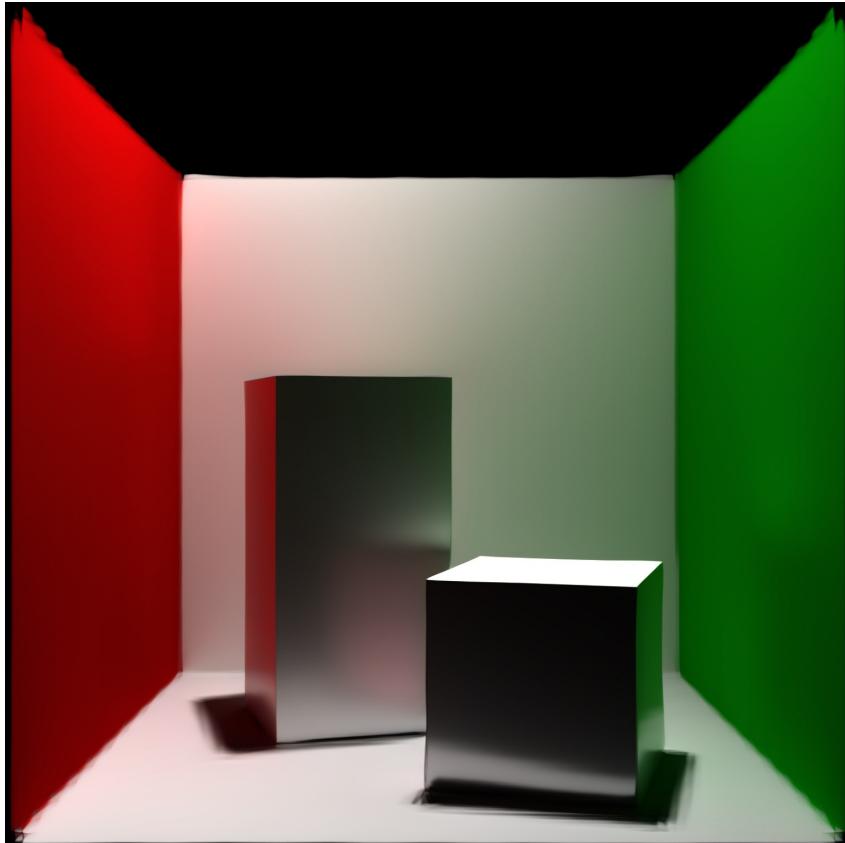
Error Image (x5)



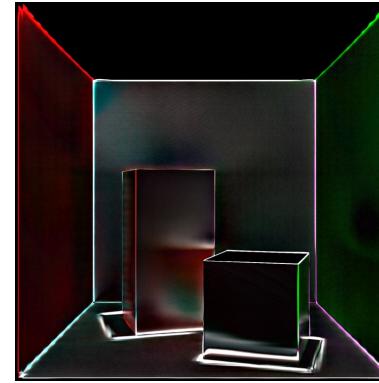
FLIP Error



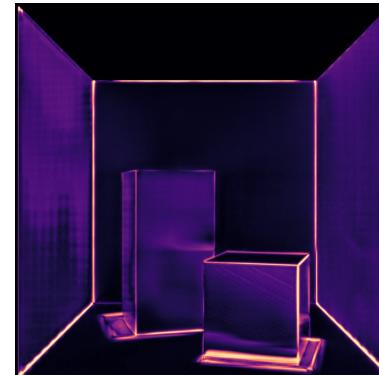
# Generalization – Resolution



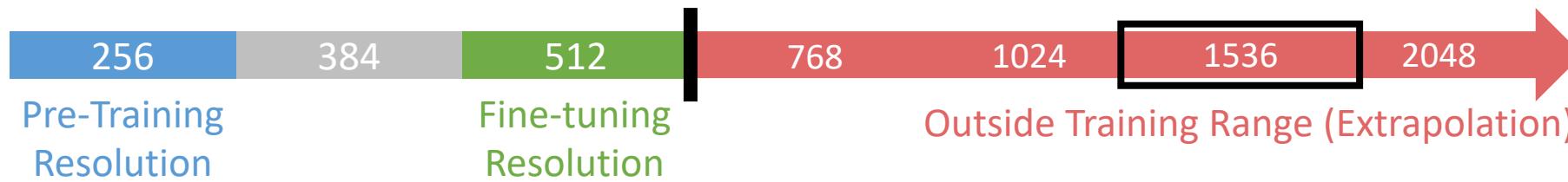
Rendered Image



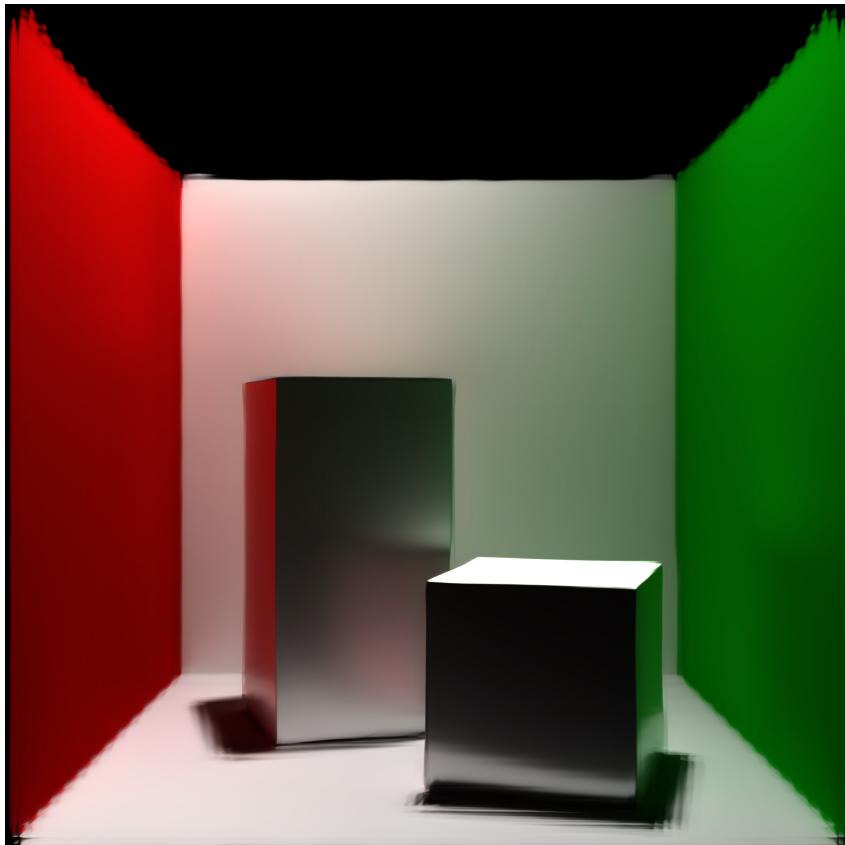
Error Image (x5)



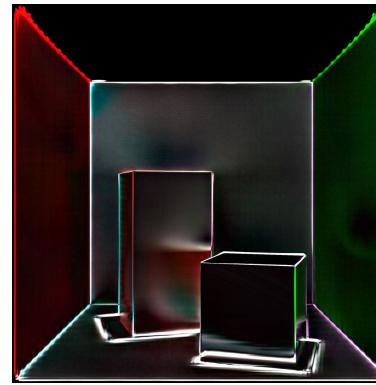
FLIP Error



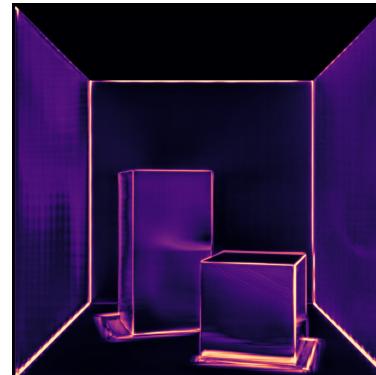
# Generalization – Resolution



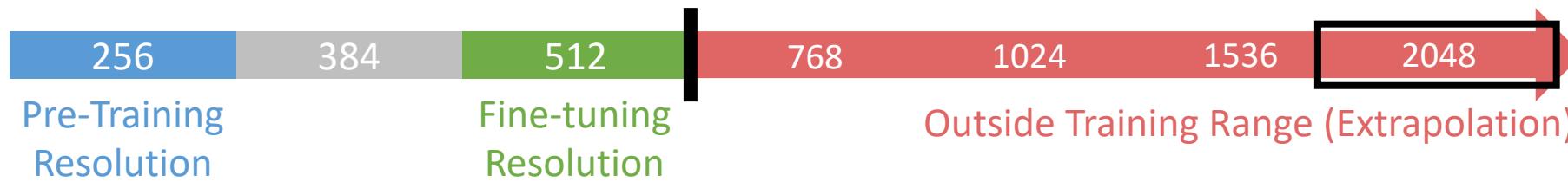
Rendered Image



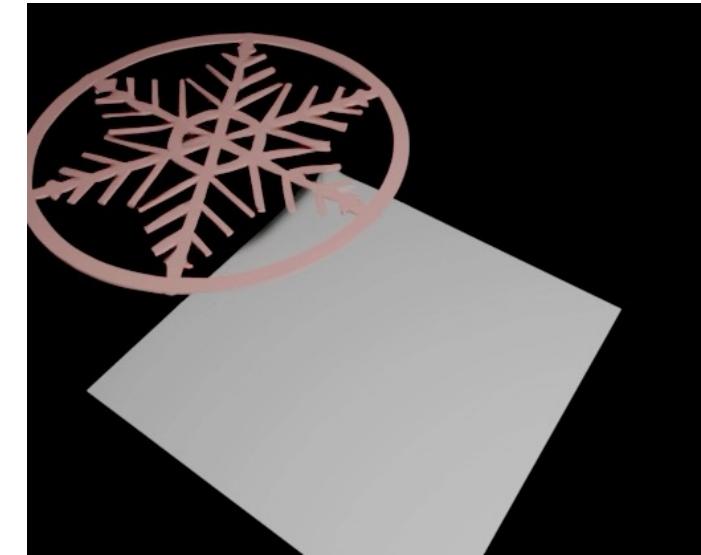
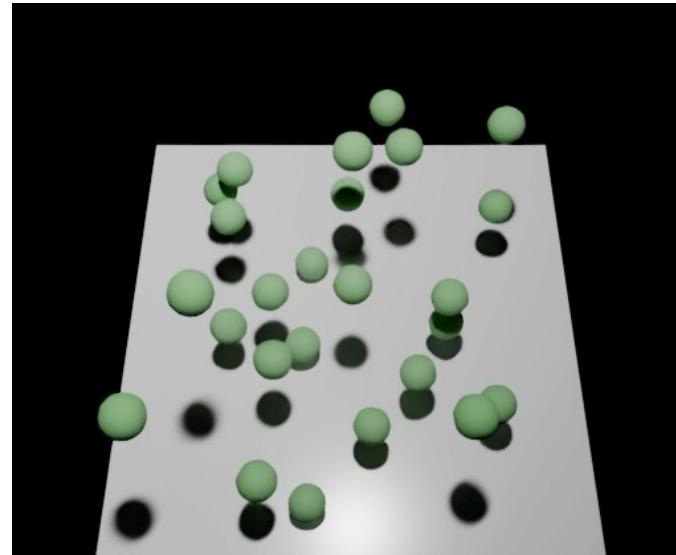
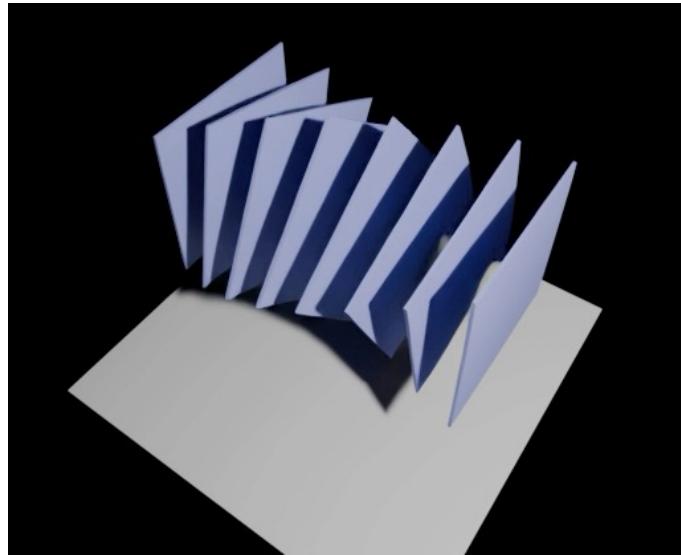
Error Image (x5)



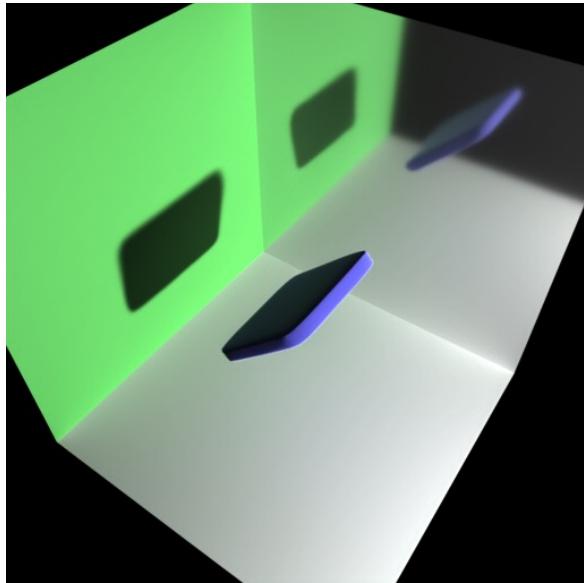
FLIP Error



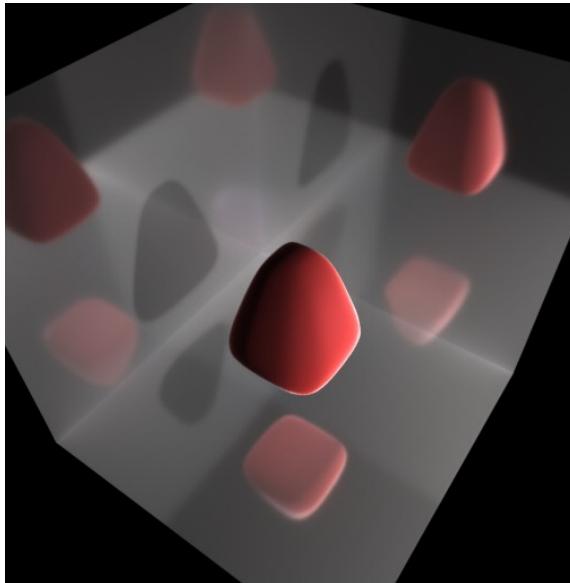
# Generalization – Shadow Complexity



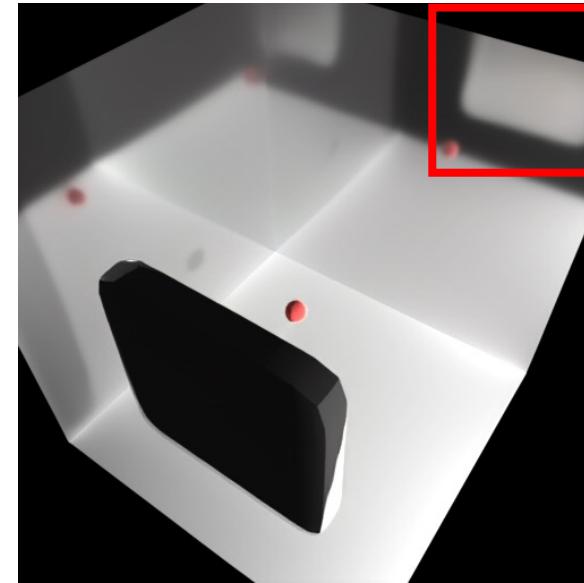
# Generalization – Interreflection Depth



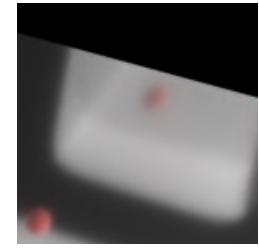
1 Interreflection



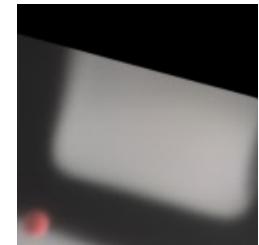
2 Interreflections



3 Interreflections

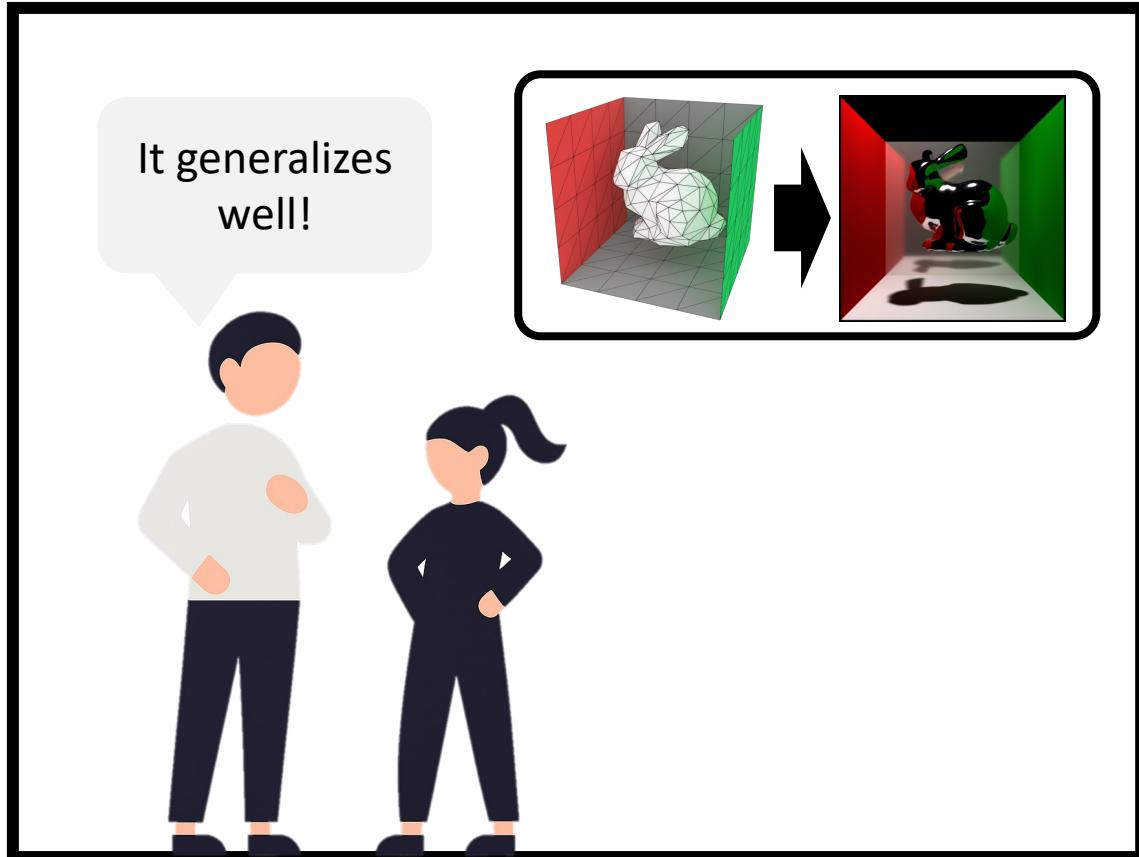


Ref.

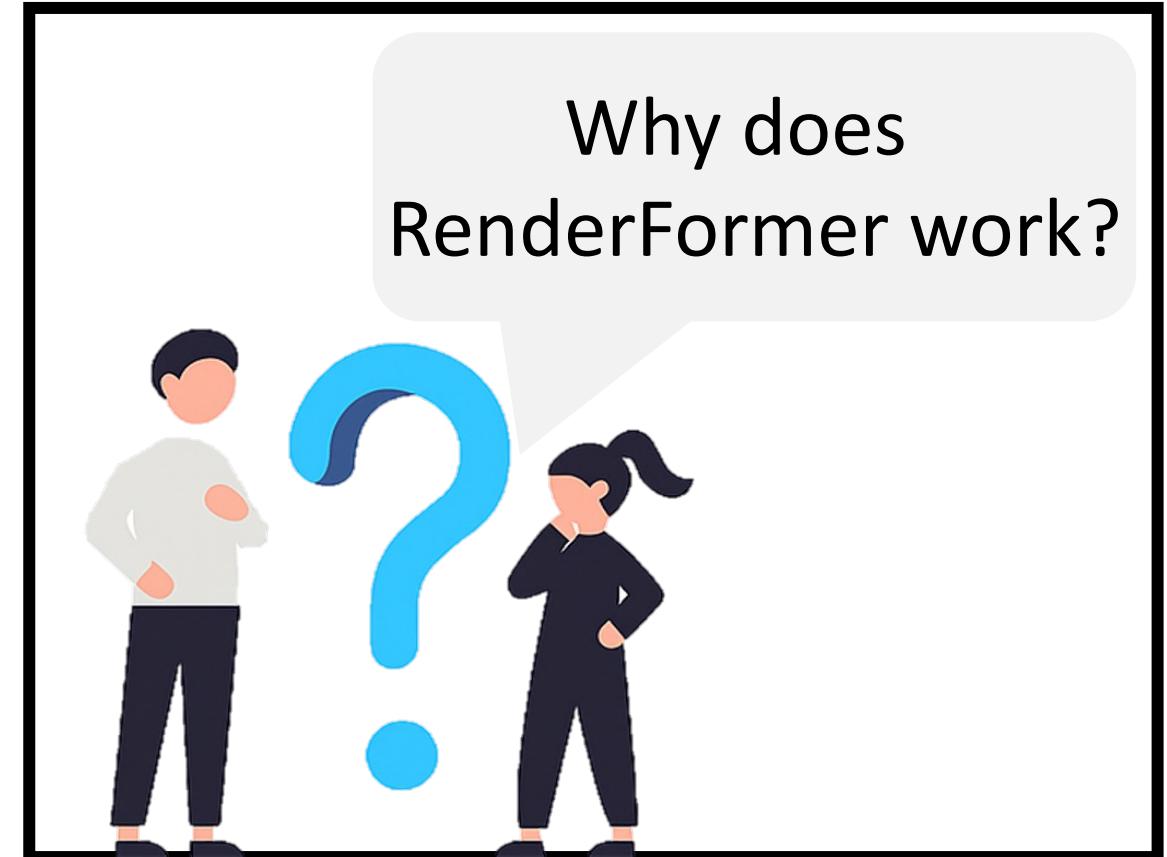


Pred.

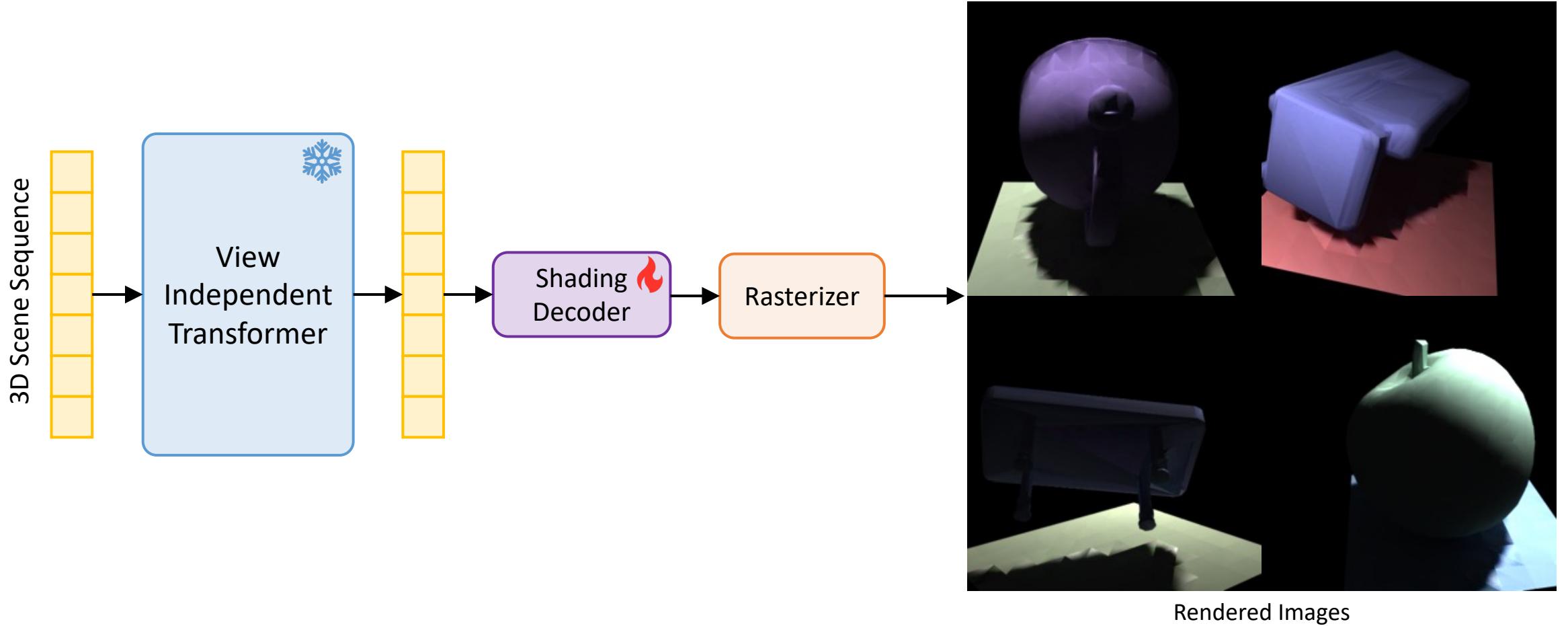
# YES,



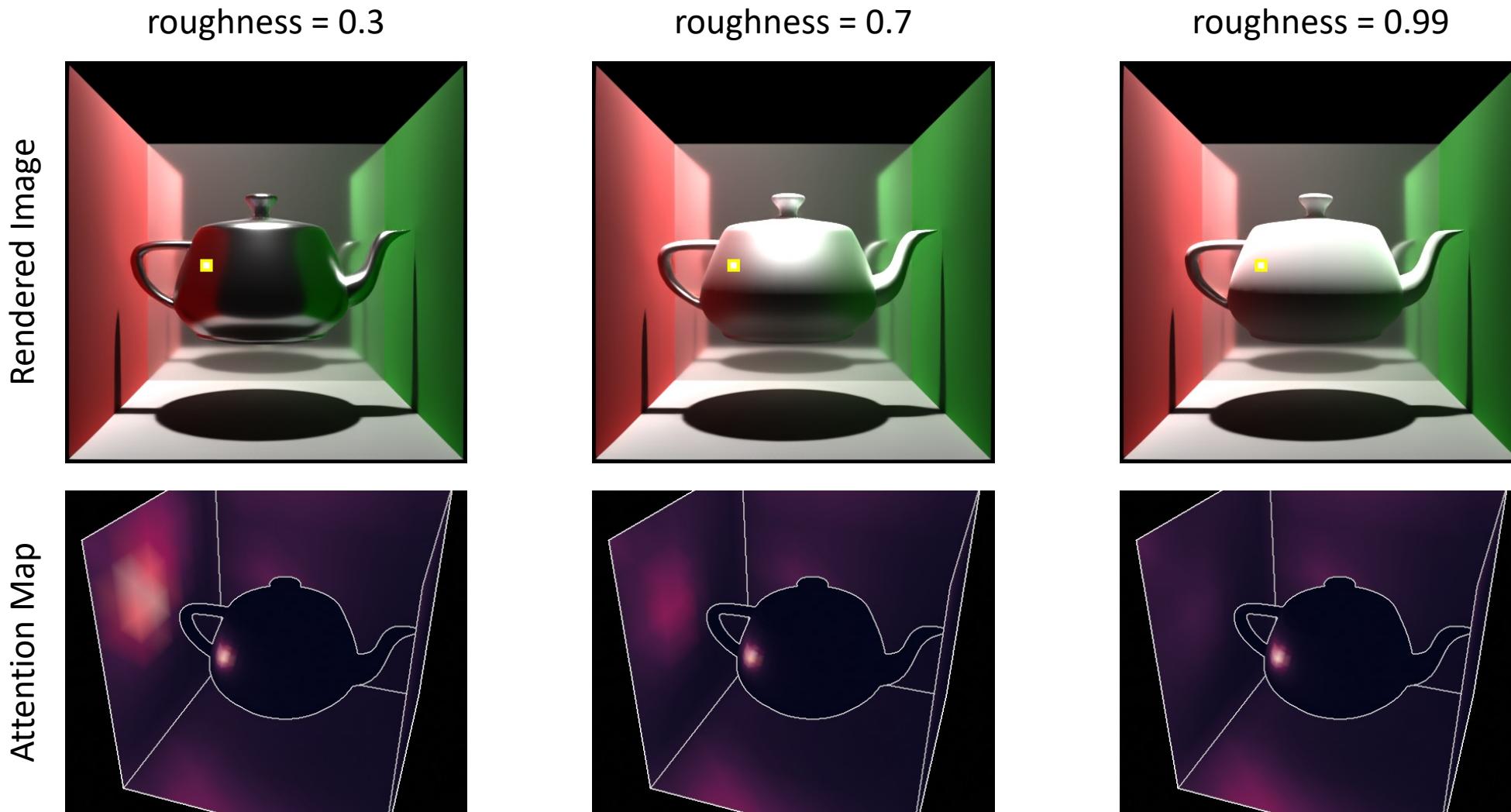
# BUT



# View-Independent Stage Resolves Diffuse Light Transport

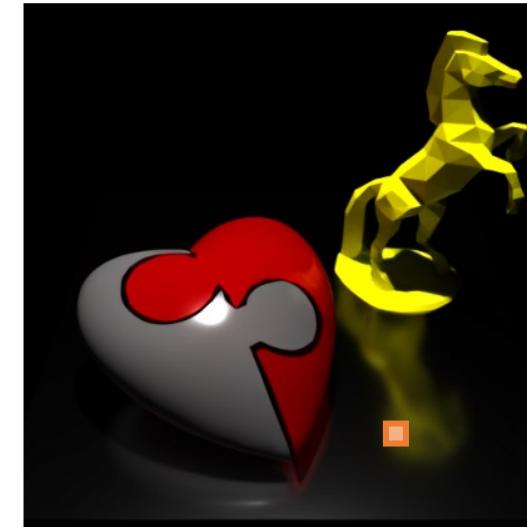
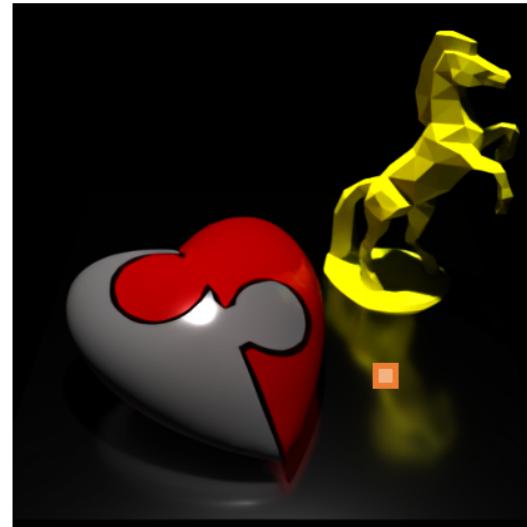
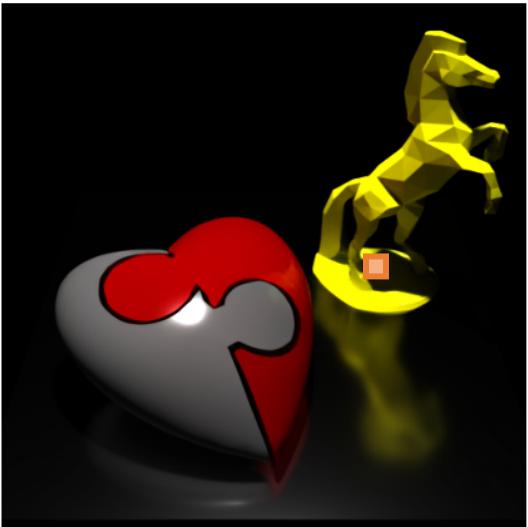


# View-Dependent Stage Resolves Triangle-Ray Intersection

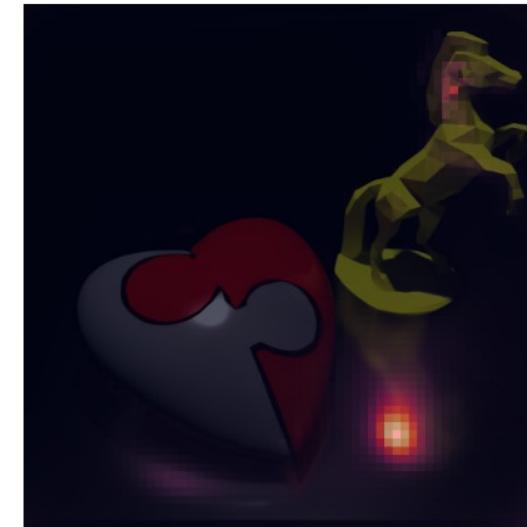
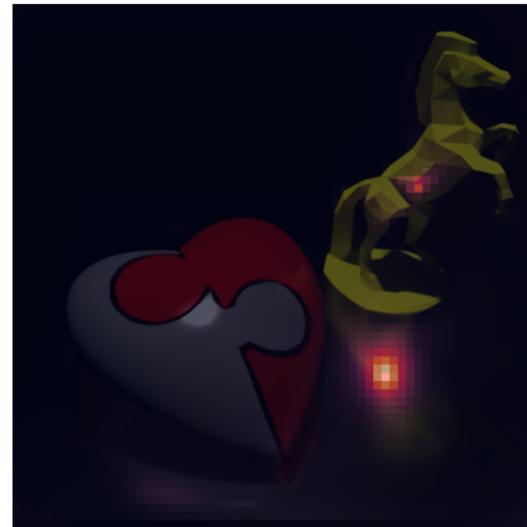


# View-Dependent Stage Learns “Screen-Space Reflection”

Rendered Image

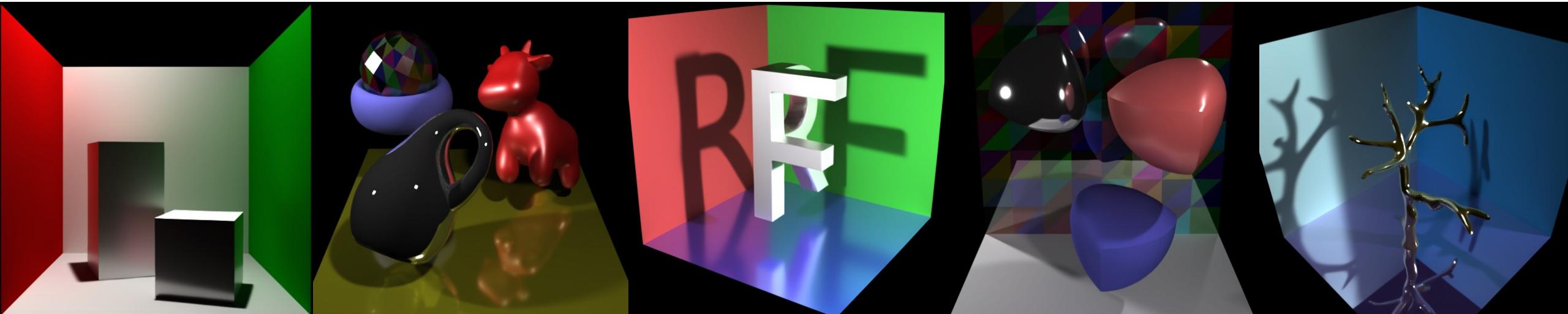


Attention Map



# Conclusion

- RenderFormer: A *Fully Transformer* Rendering Pipeline
  - 3D Mesh Input
  - Full Global Illumination Effects
  - No Per-Scene Training
  - Minimal Prior Constraints
- Attention Is All You Need for Rendering



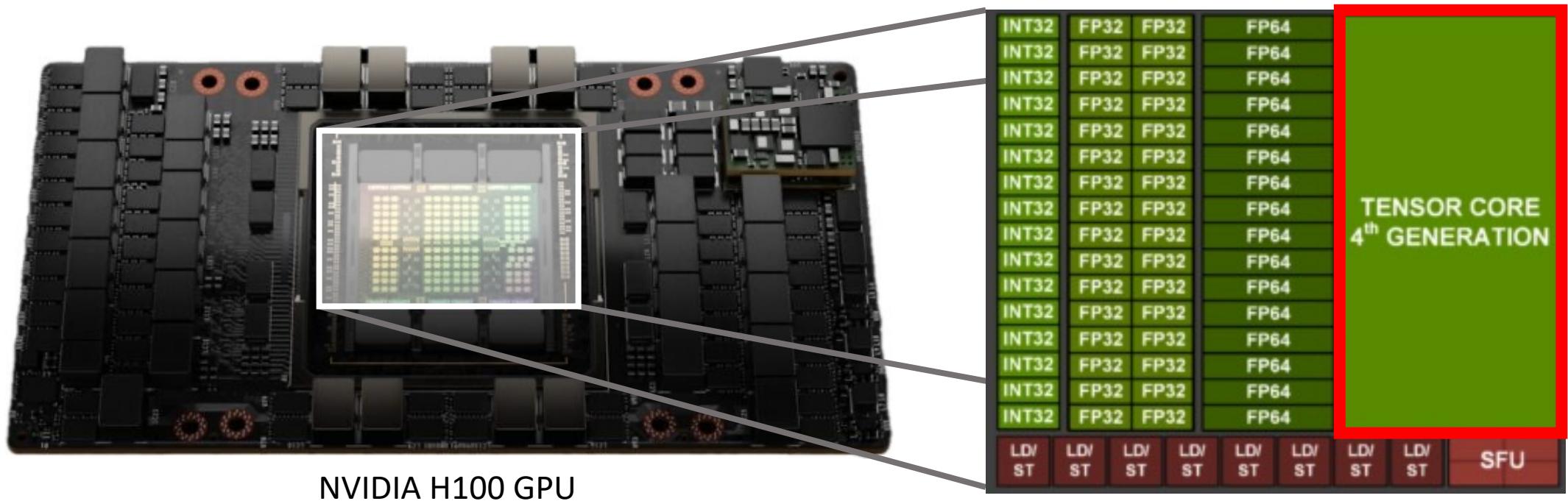
# RenderFormer Today: An Initial Step



## Future Work

- More Generality
  - Larger Scene
  - Texture
  - Complex Material
  - Environmental Lighting
  - ...
- Higher Efficiency

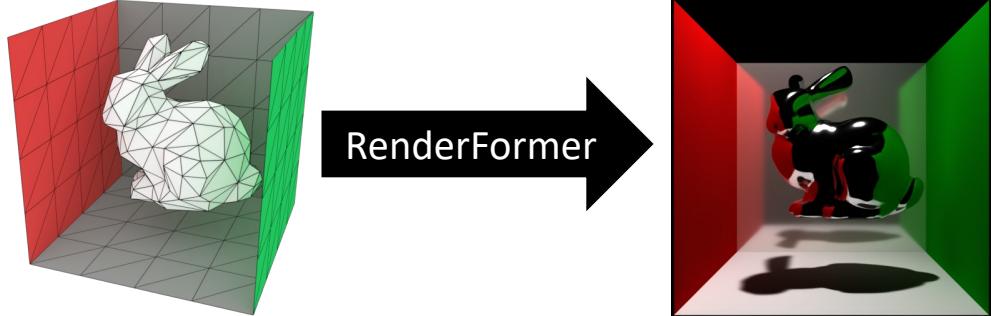
# Modern GPUs are Built for Transformers



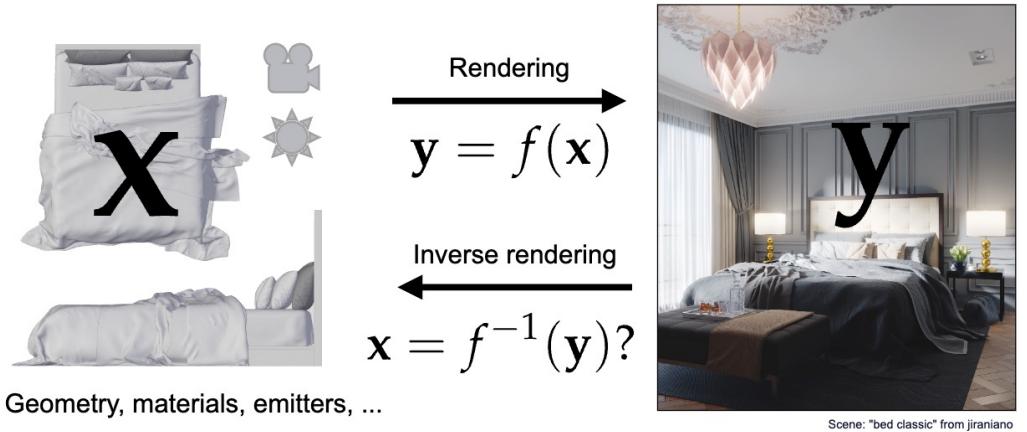
# Unifying Computations on GPU

# The Road Ahead: Exciting Future Applications

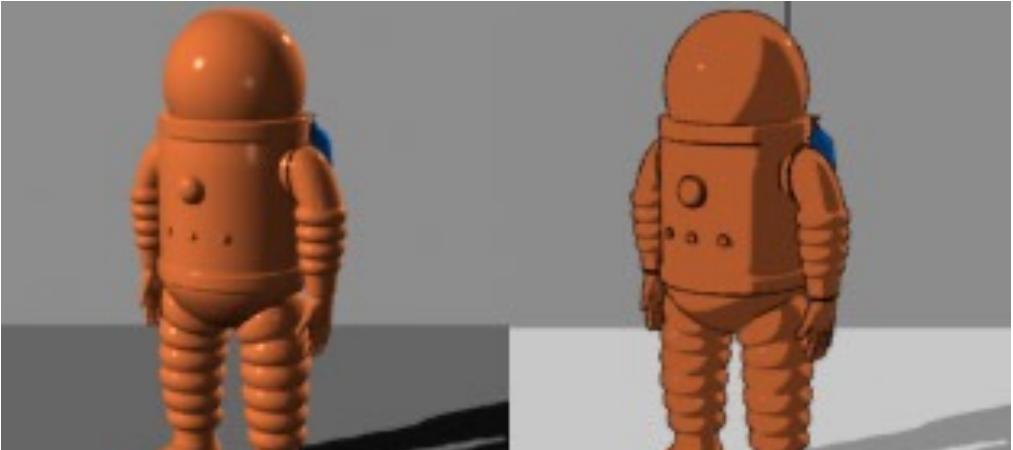
Graphics Foundation Models



Inverse Rendering



Unified Photorealistic & Stylized Rendering



Batched Rendering



# Thank You for Your Attention!

RenderFormer: Transformer-based Neural Rendering of Triangle Meshes with Global Illumination

<https://microsoft.github.io/renderformer/>



Project Page

