



Massachusetts
Institute of
Technology



Learning to Render Novel Views from Wide-Baseline Stereo Pairs

Yilun Du, Cameron Smith, Ayush Tewari[†], Vincent Sitzmann[†]

[†] denotes equal advising

TUE-PM-079

3D Reconstruction from Many Images

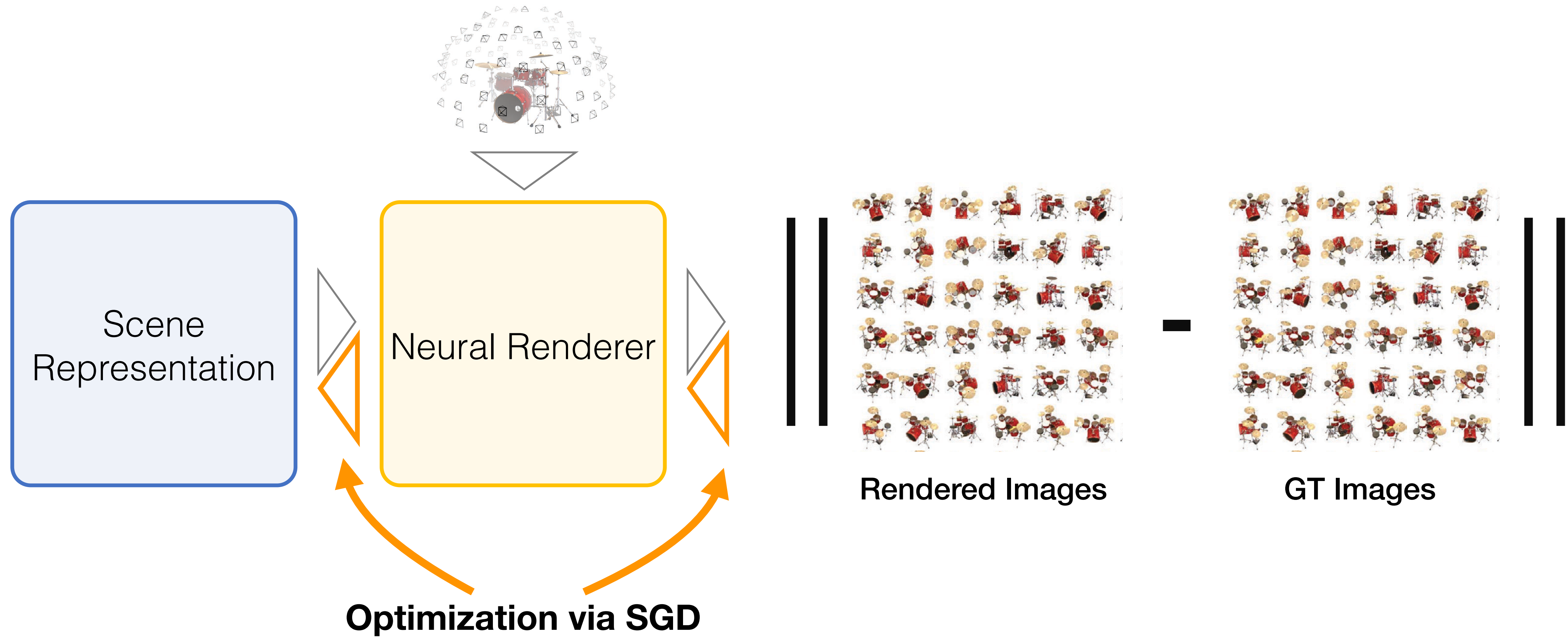


NeRF, ECCV 2021



Plenoxels, CVPR 2022

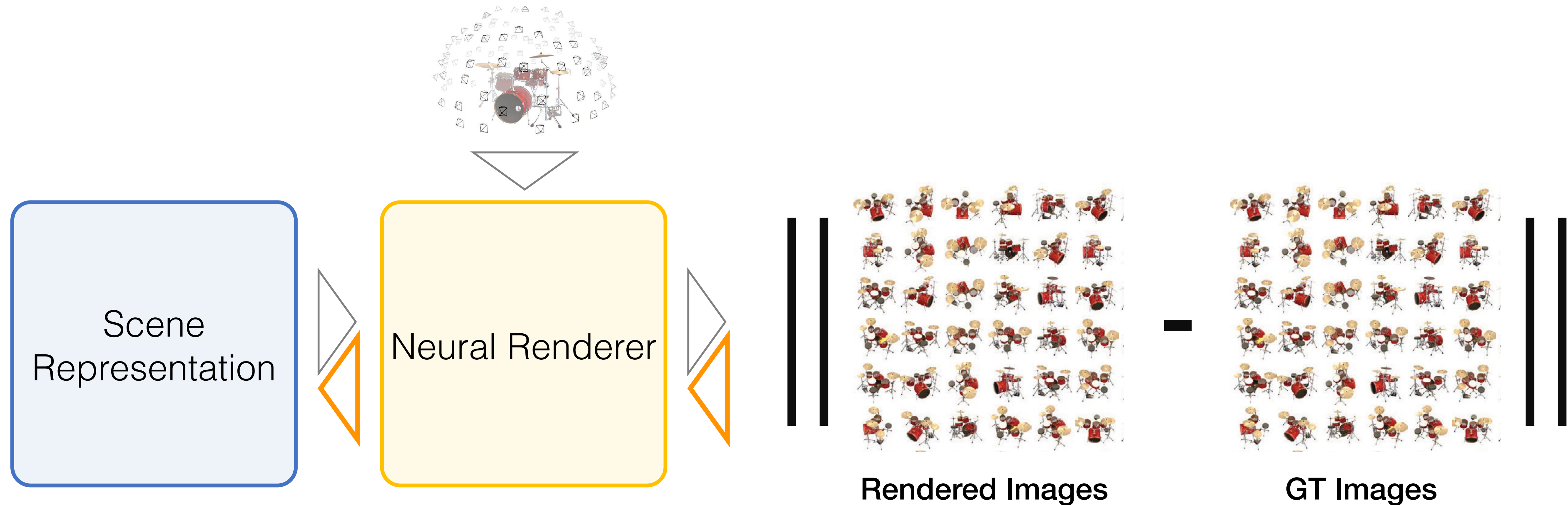
Reconstruction via per-scene Optimization



Reconstructing Static Scene from 3 images



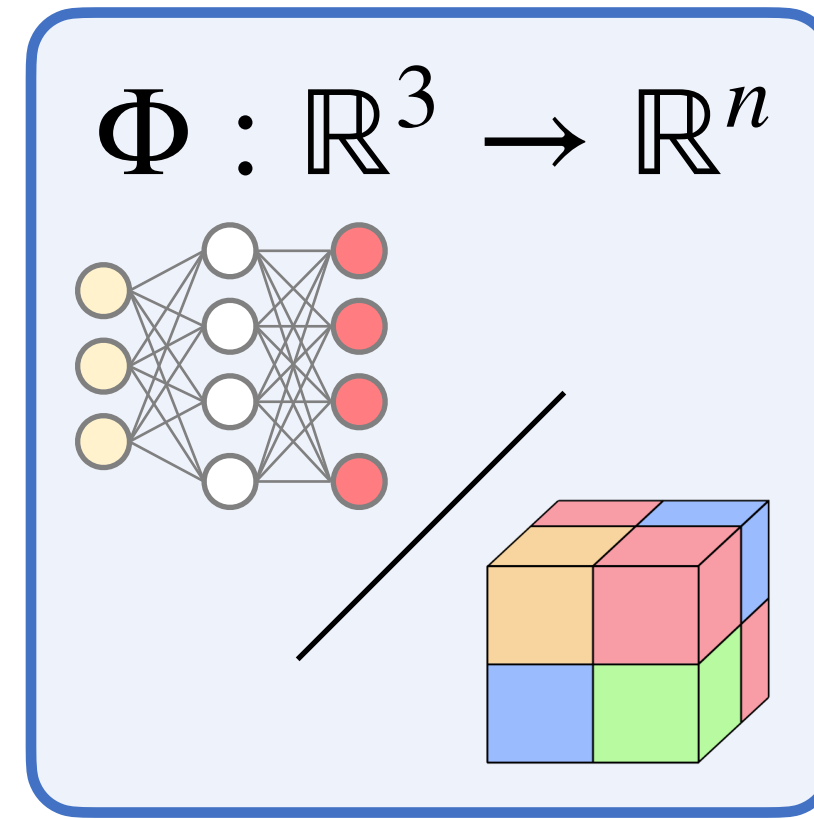
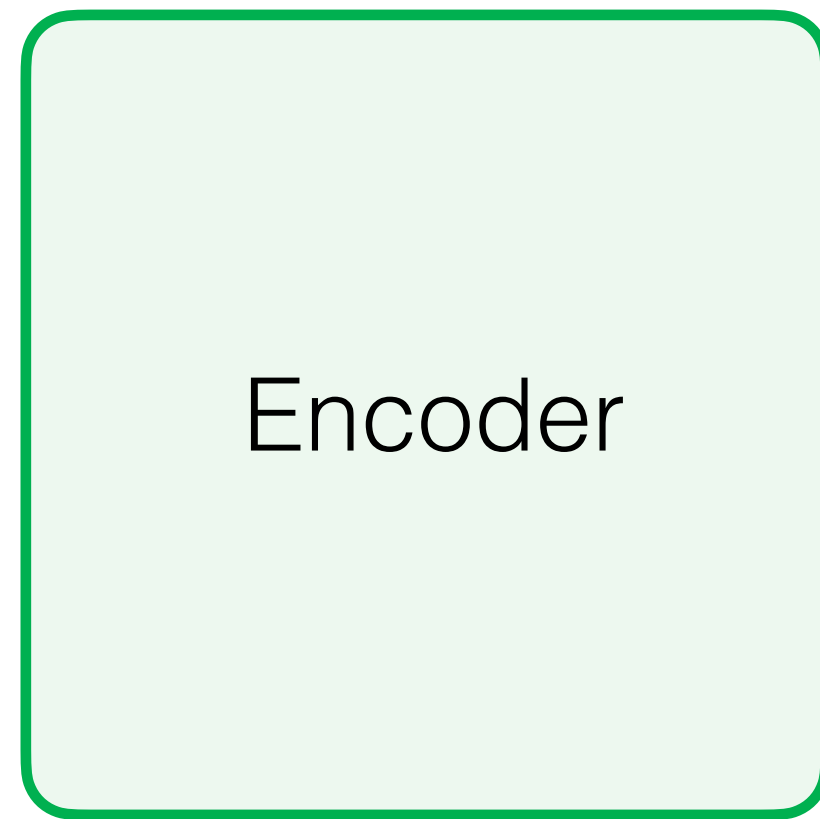
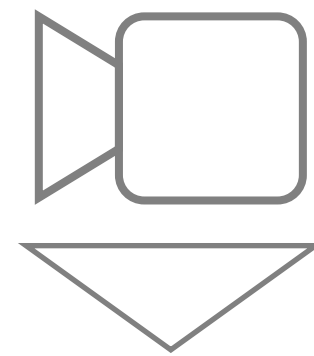
Reconstruction via per-scene Optimization



Can't deal with incomplete observations: Need *learned priors*.
Not scalable: Need *online, real-time* reconstruction.

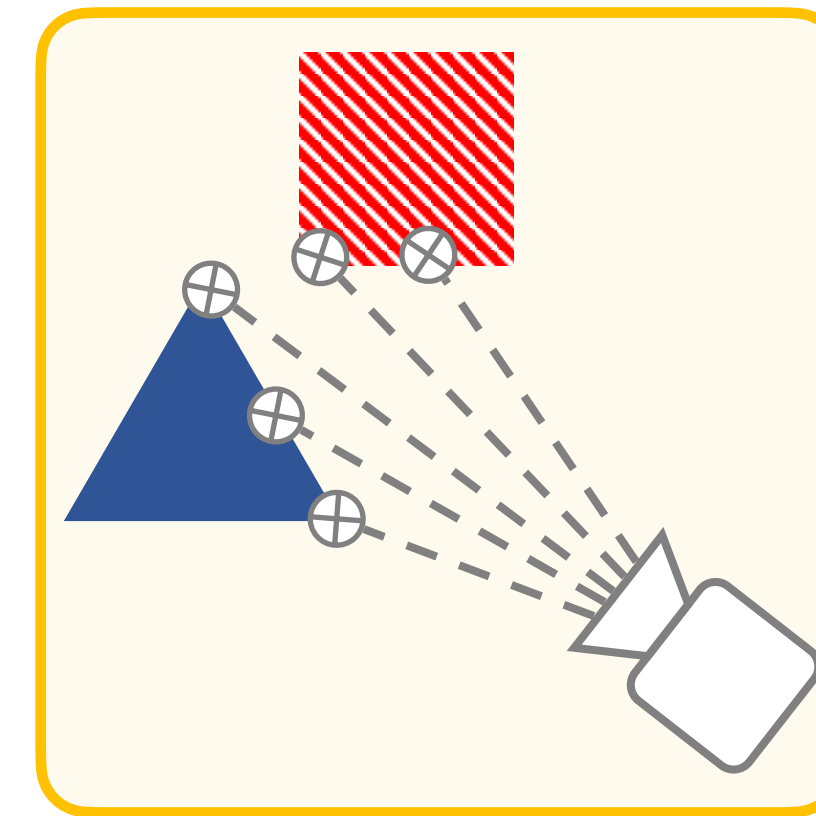
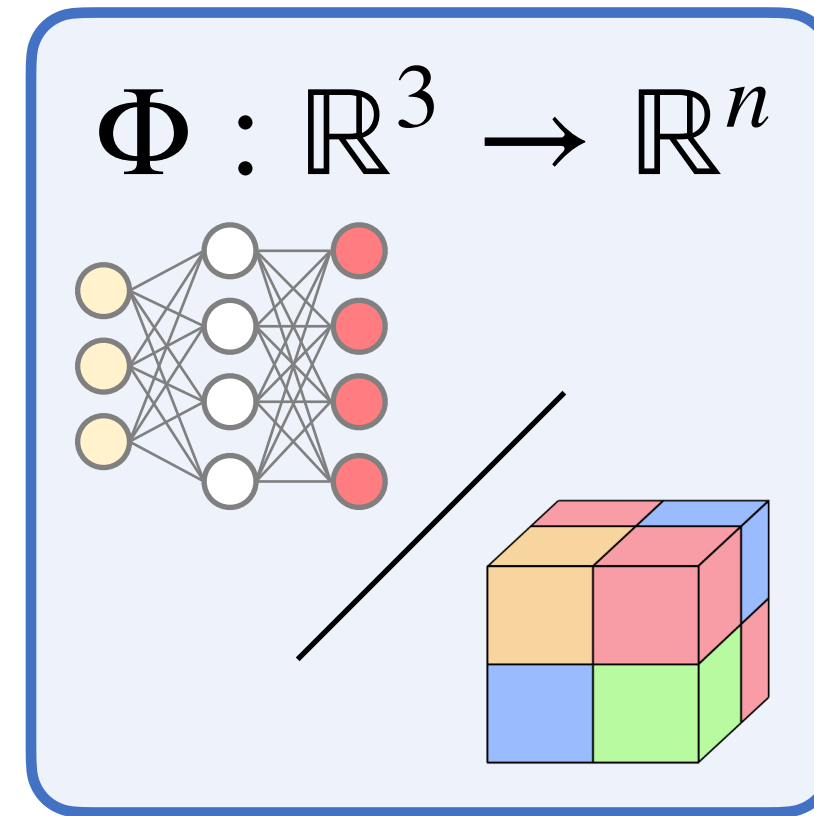
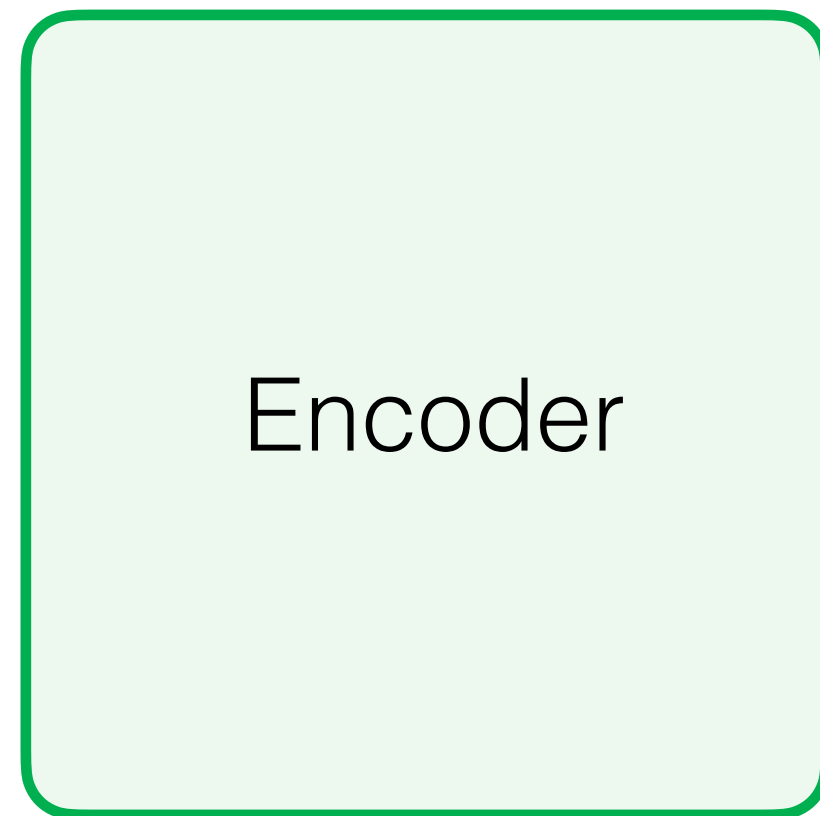
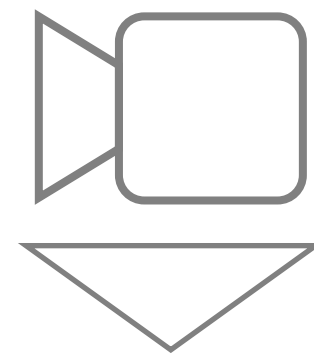
Amortized (=feedforward, generalizable) 3D Reconstruction

Observations



Amortized (=feedforward, generalizable) 3D Reconstruction

Observations



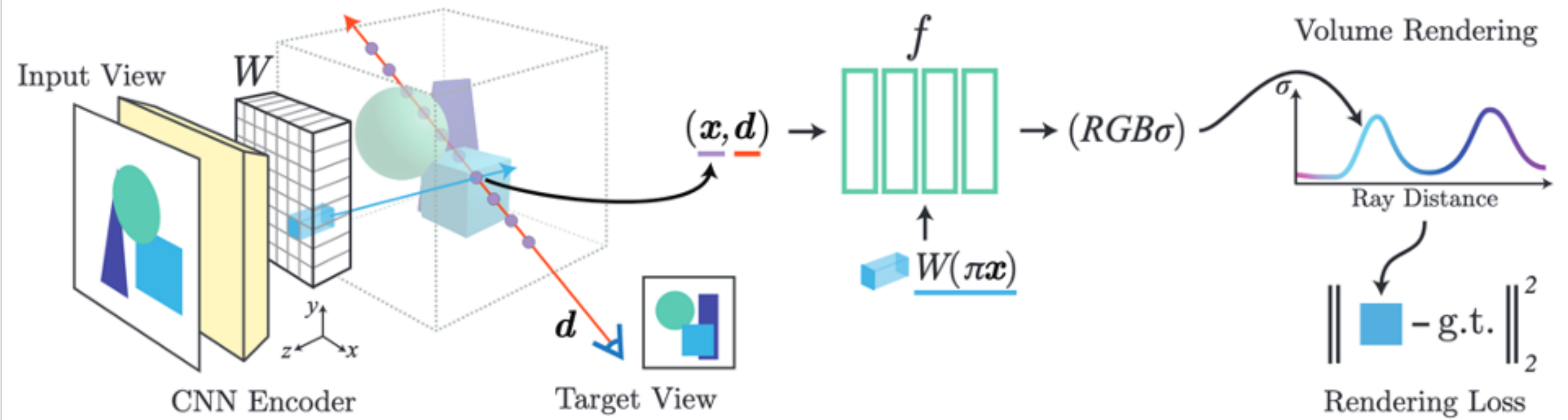
Renderings



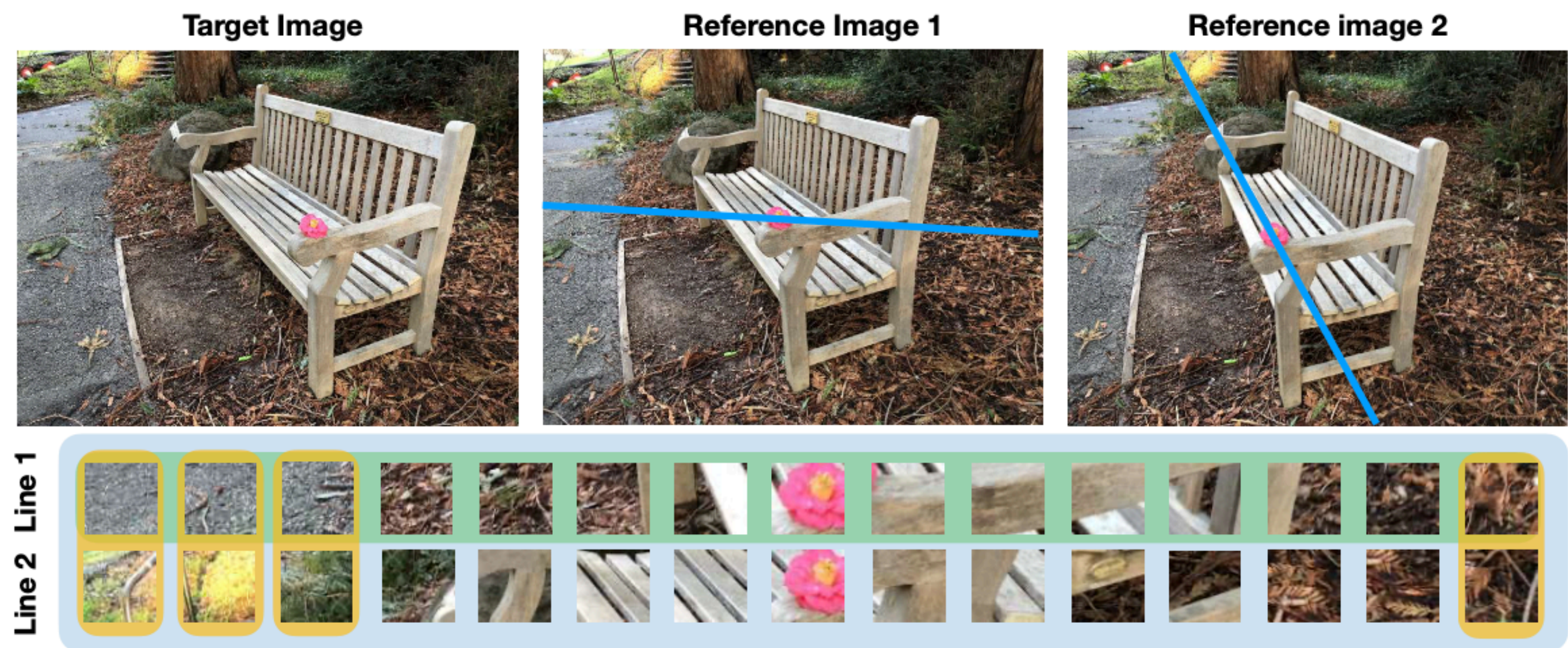
Learning to Render Novel Views from Wide-Baseline Stereo Pairs

Prior Work

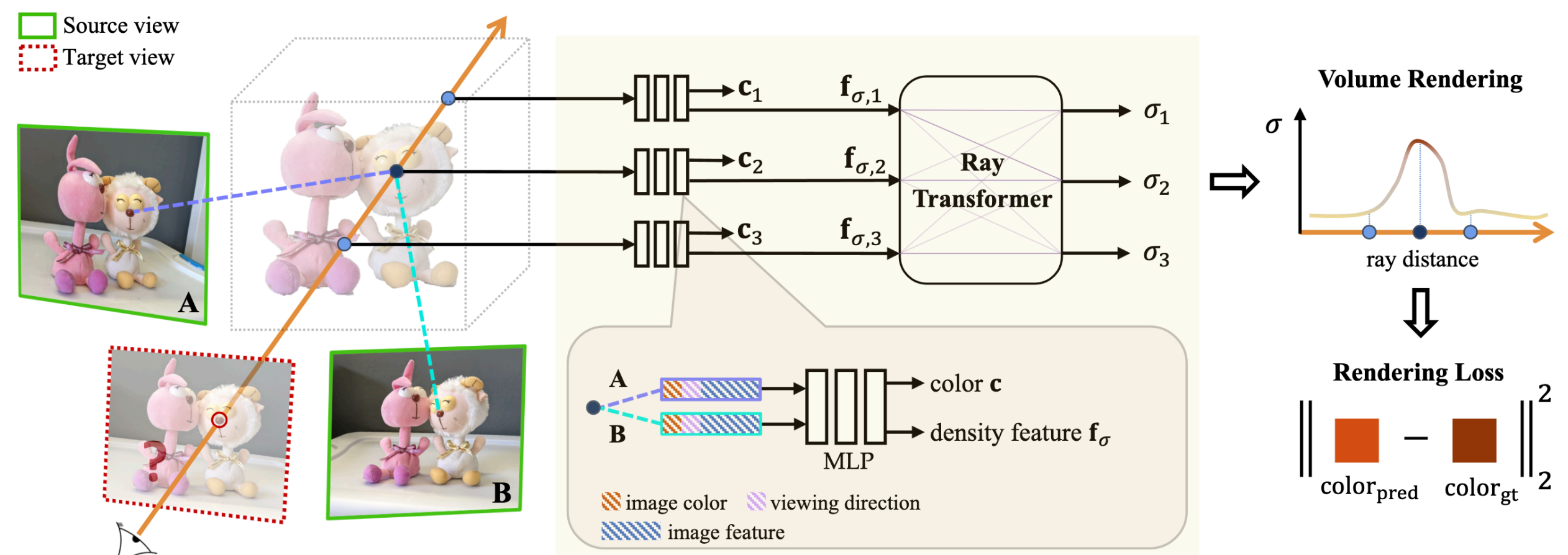
PixelNeRF, Yu et al. 2021



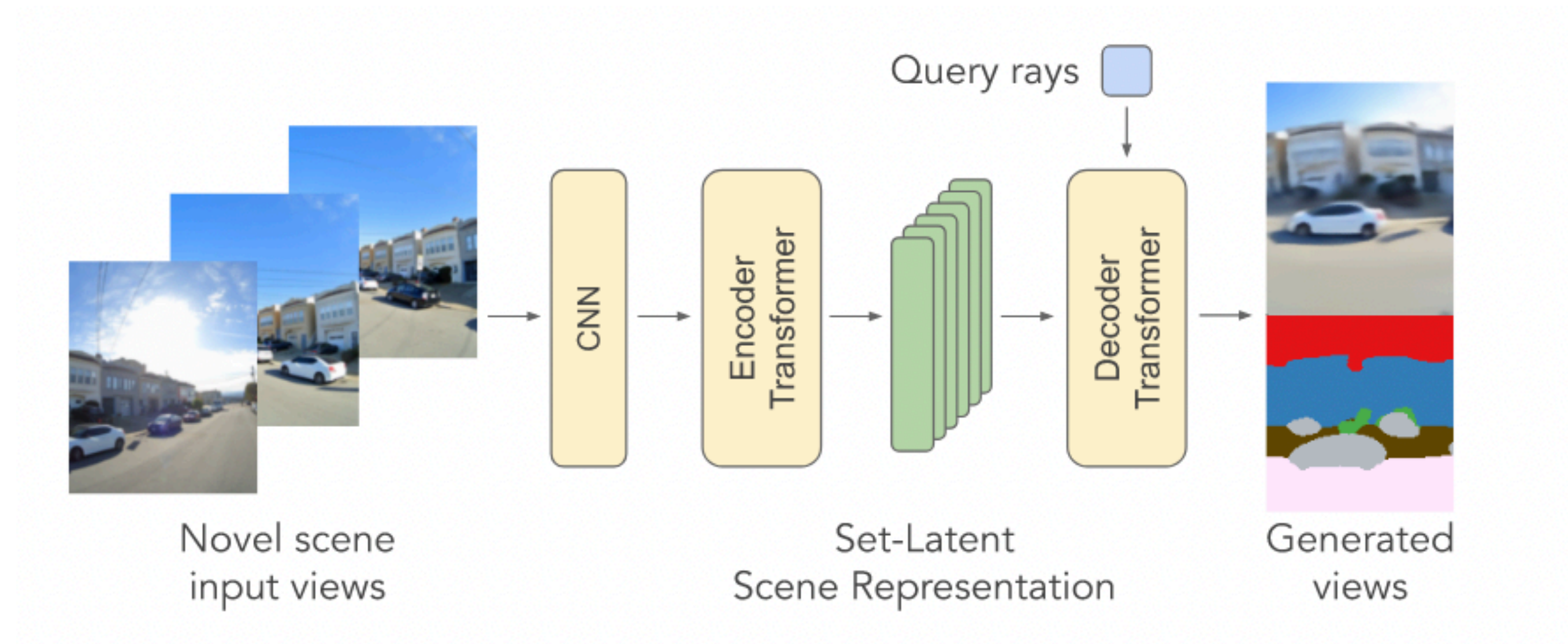
Generalized Patch-Based Neural Rendering, Suhail et al. 2022



IBR-Net, Wang et al. 2021



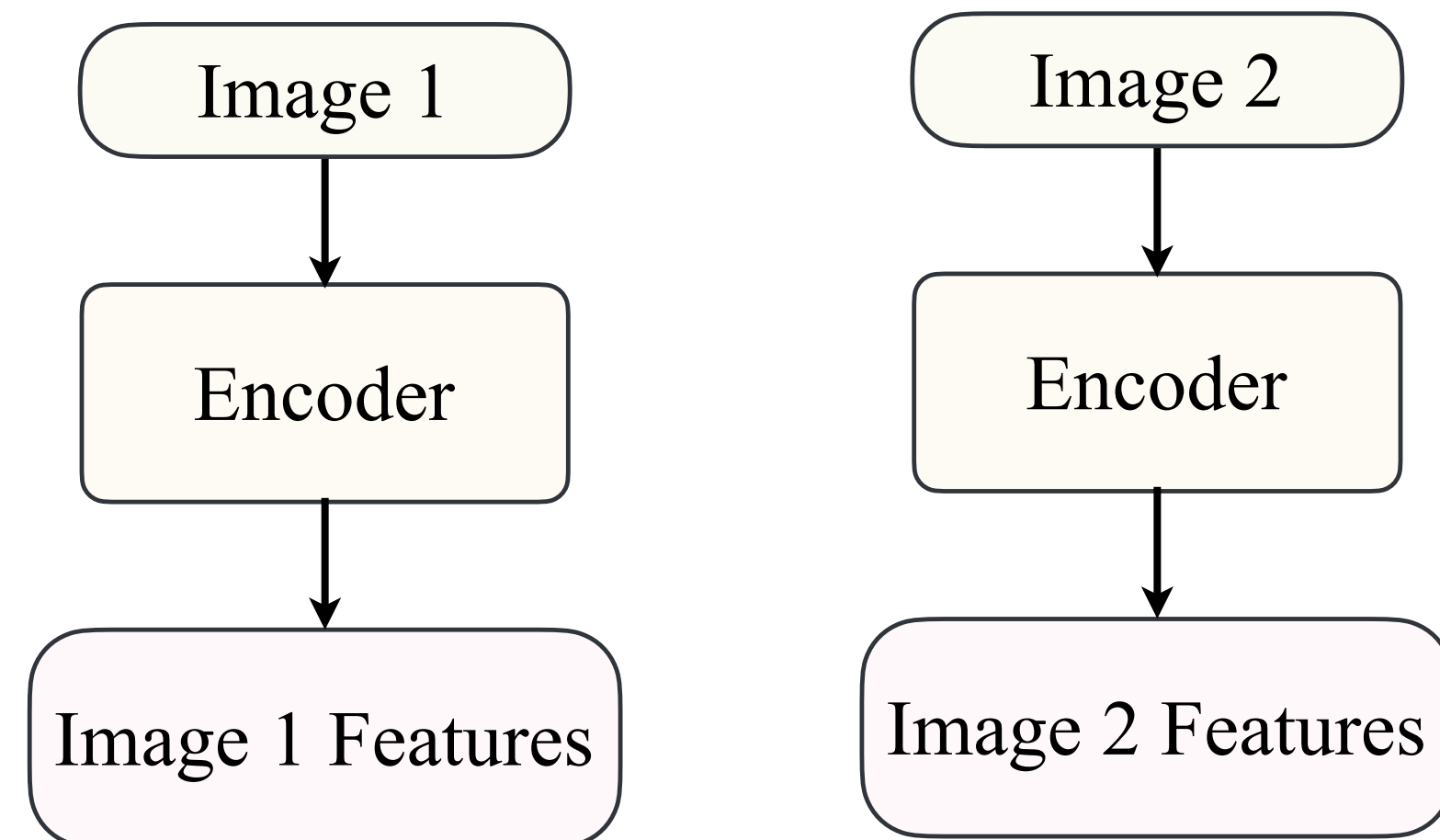
Prior Work



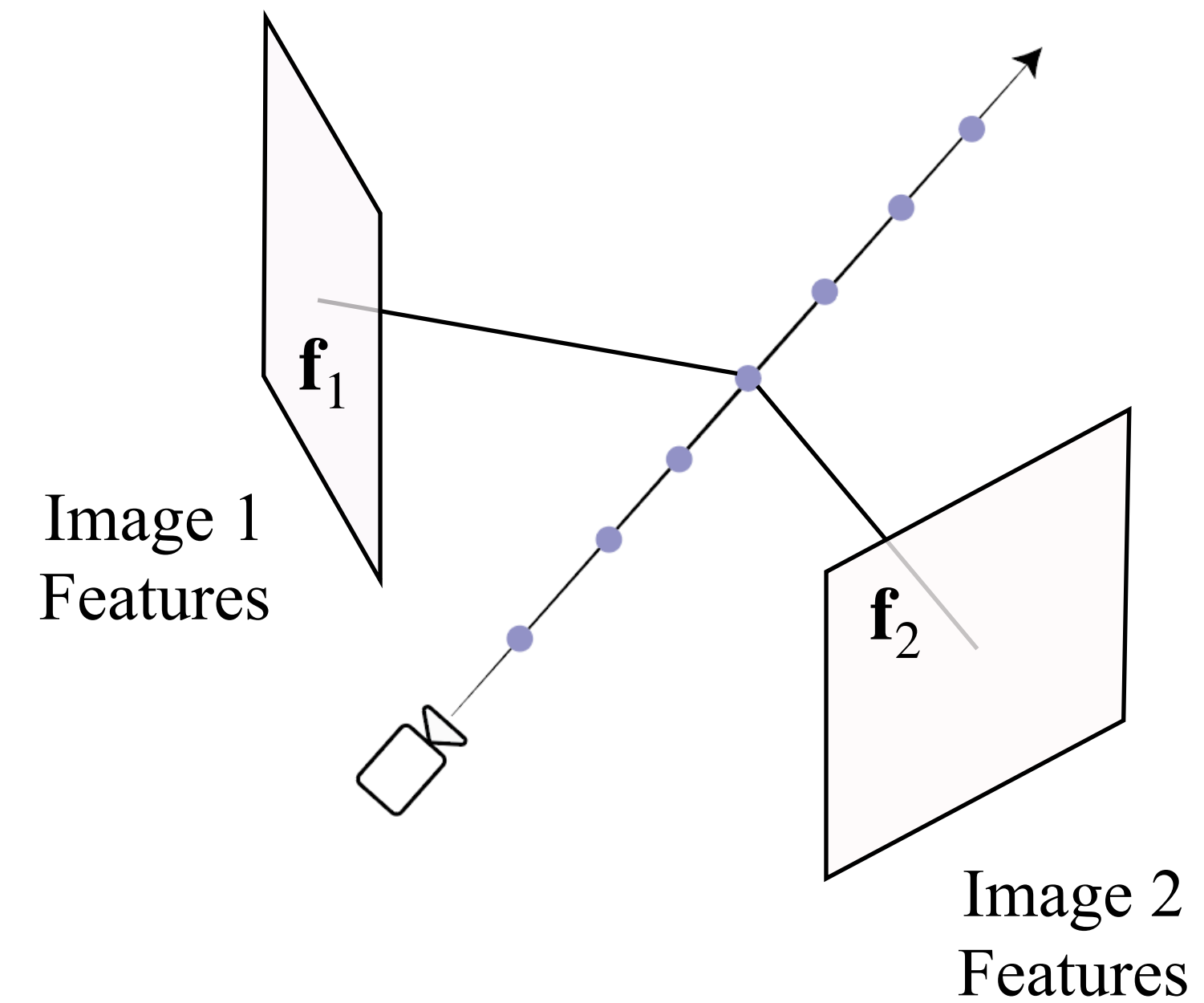
Scene Representation Transformer,
Sajjadi et al. 2022

Two Core Challenges

Per-Image Feature Extraction

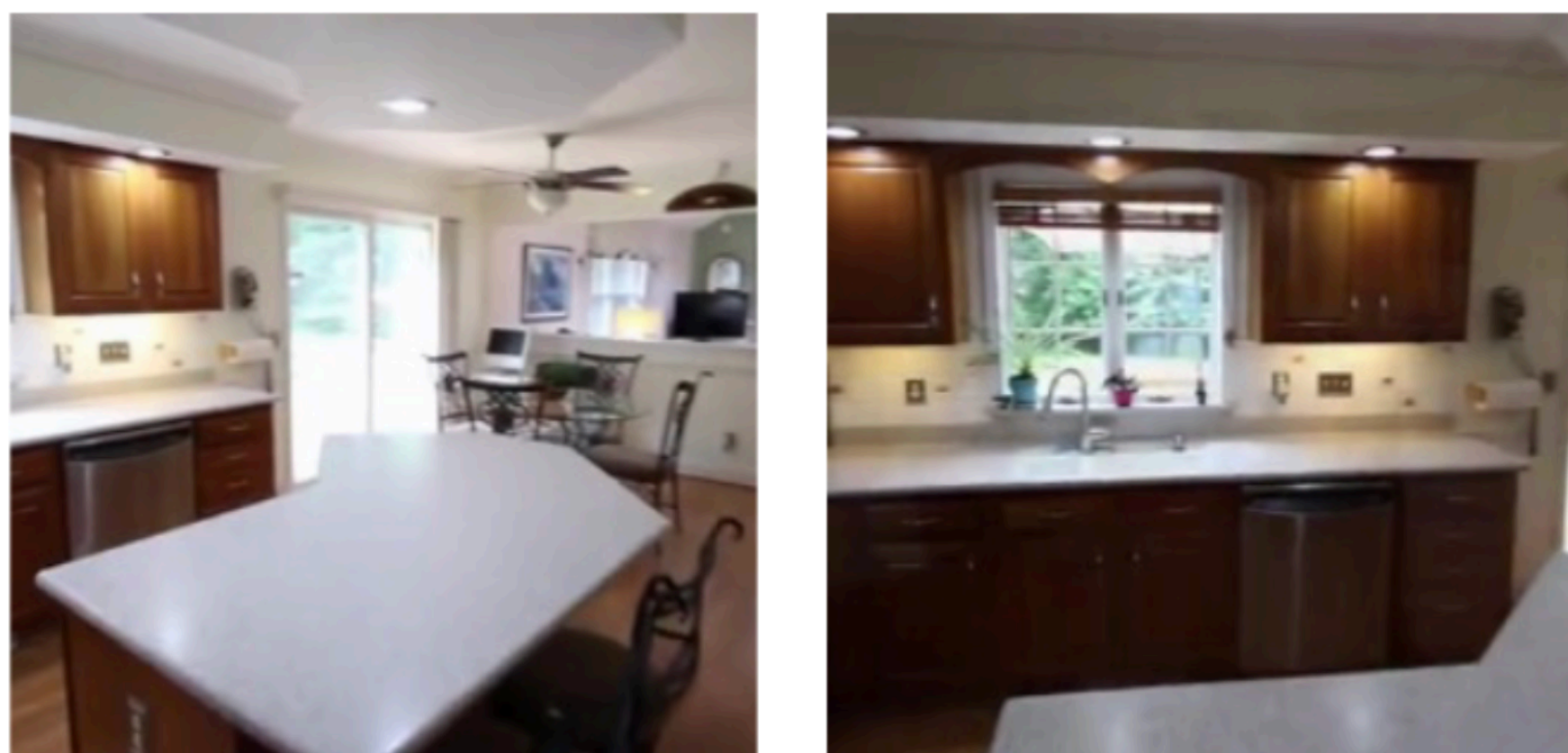
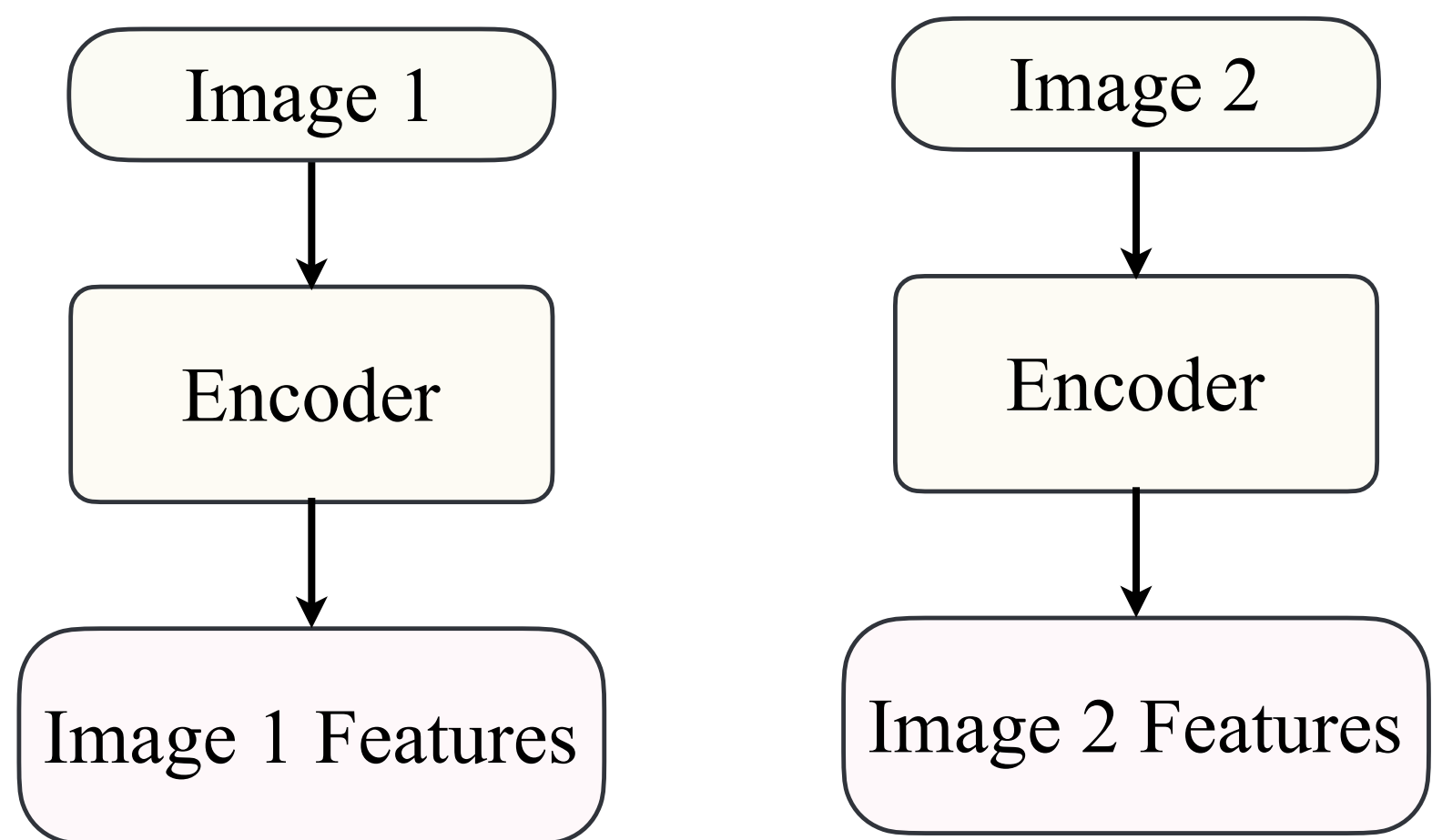


Sampling of pixel-aligned features, Volume Rendering



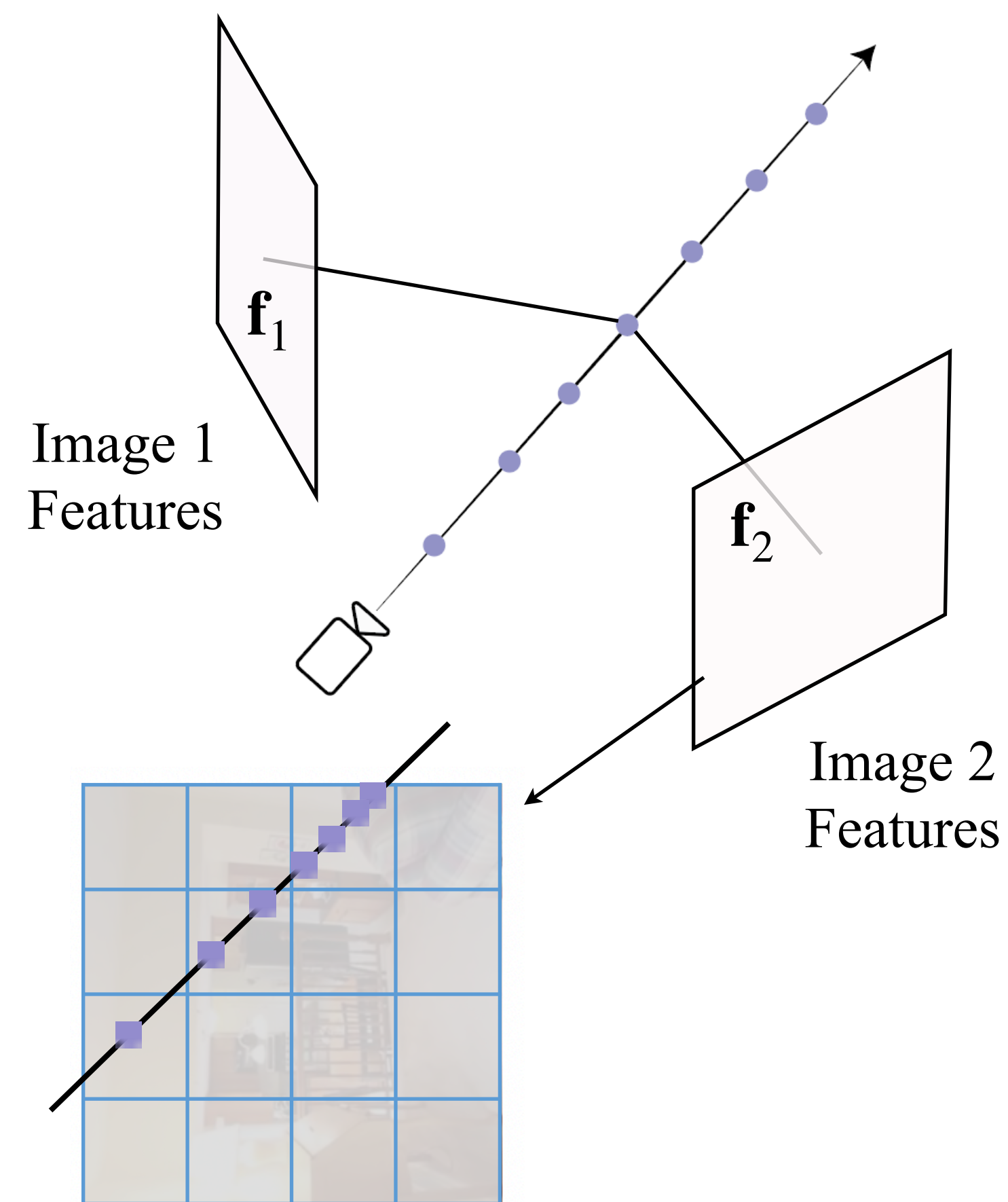
Two Core Challenges

Per-Image Feature Extraction



Depths estimated for view 1 and view 2 might be inconsistent!

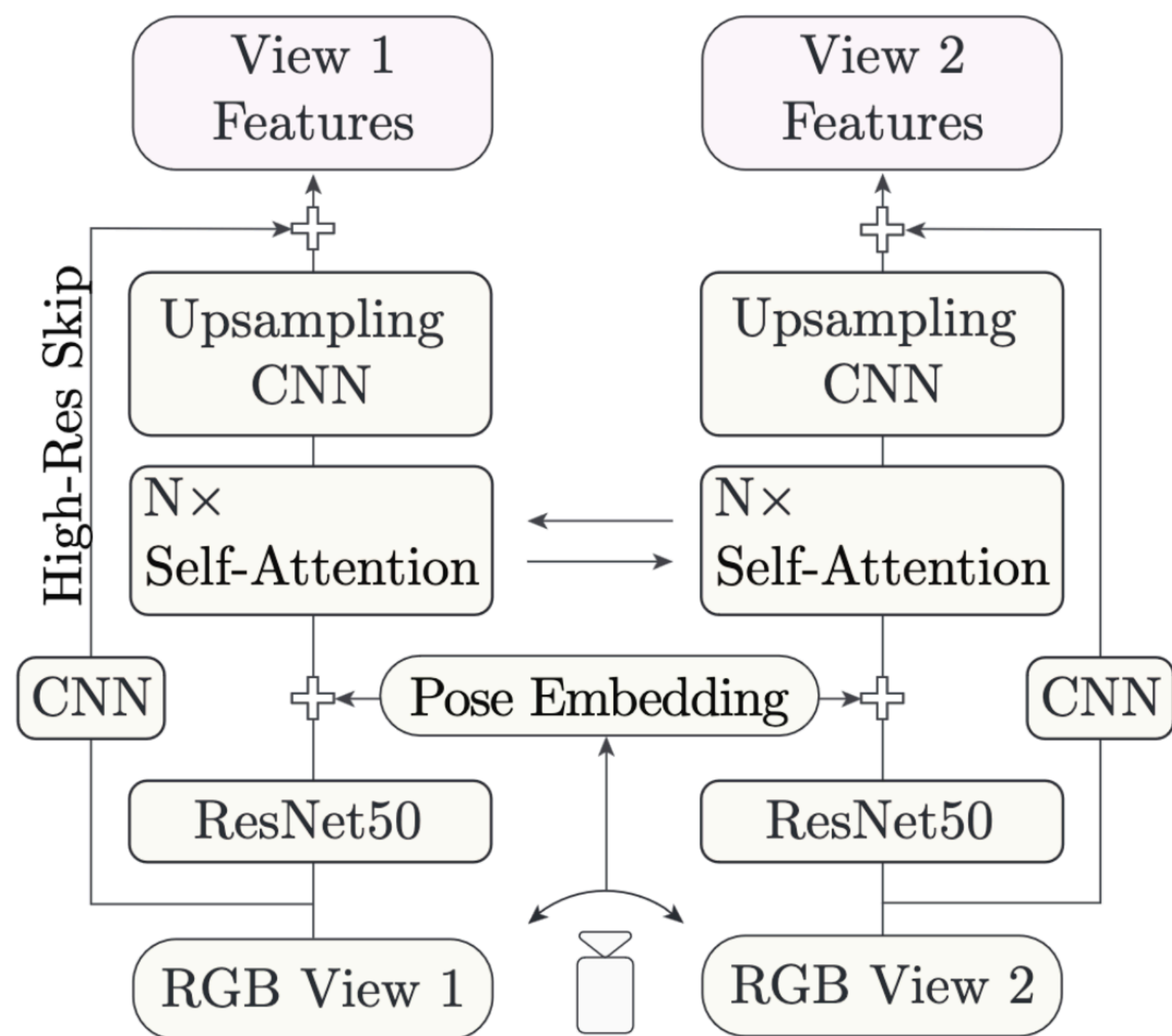
Sampling of pixel-aligned features, Volume Rendering



There's no point in sampling a pixel more than once, we're not gaining information!

Piece 1: Multi-View Encoder

(a) Multi-View ViT Encoder

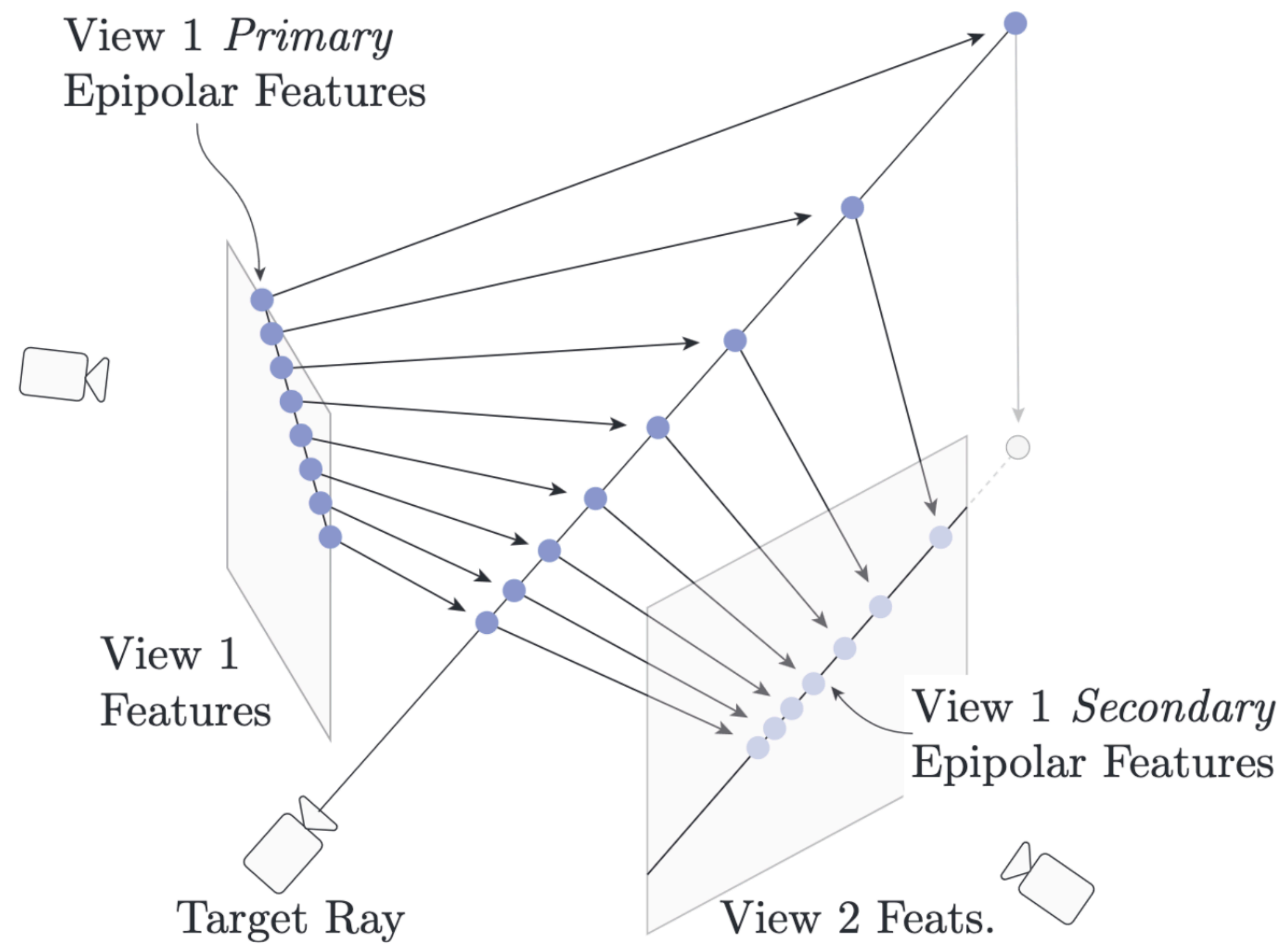


Input Images



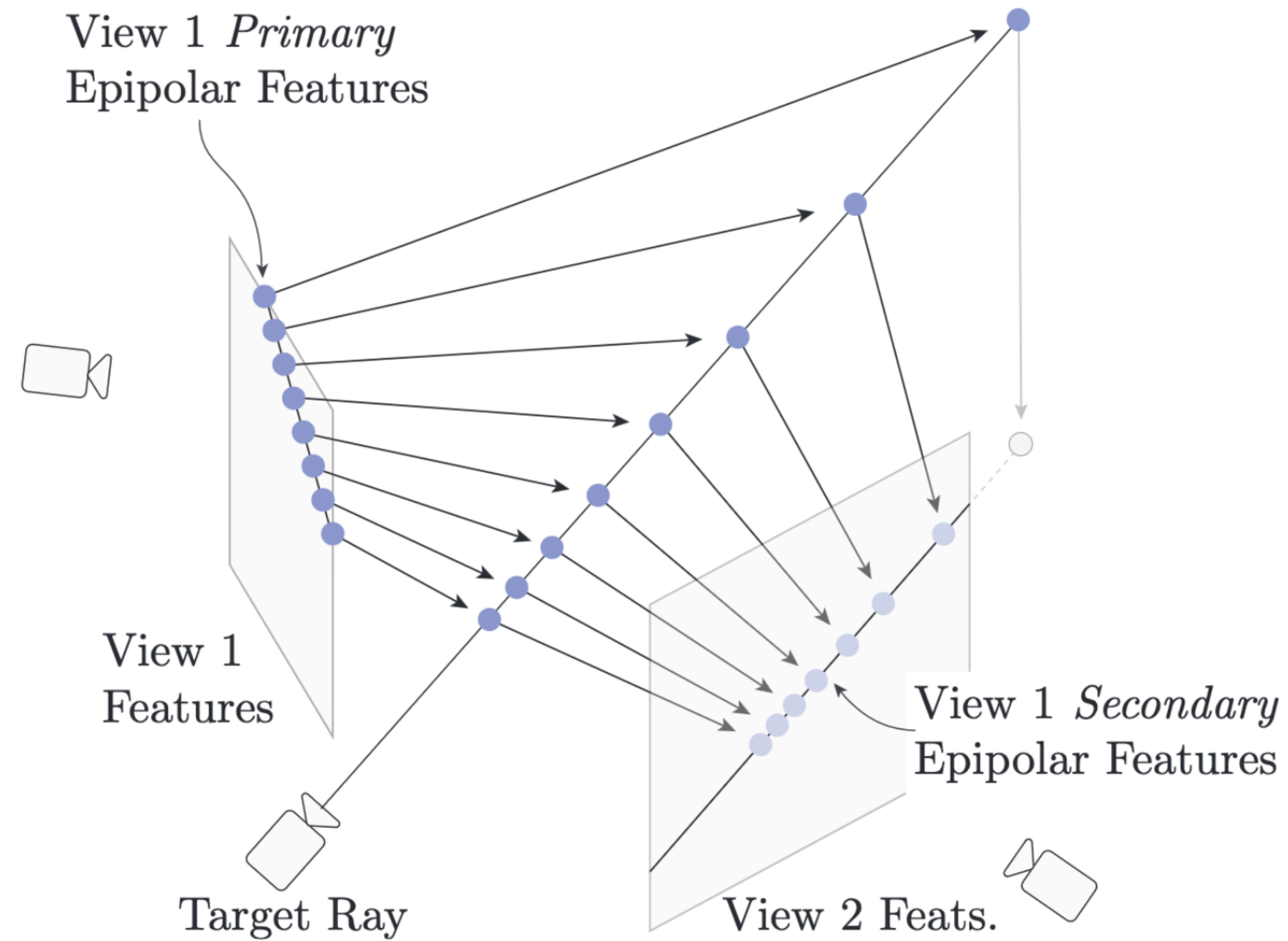
Piece 2: Image-Centric Epipolar Line Sampling & Attention-based Rendering

(b) Image-Centric Epipolar Line Sampling

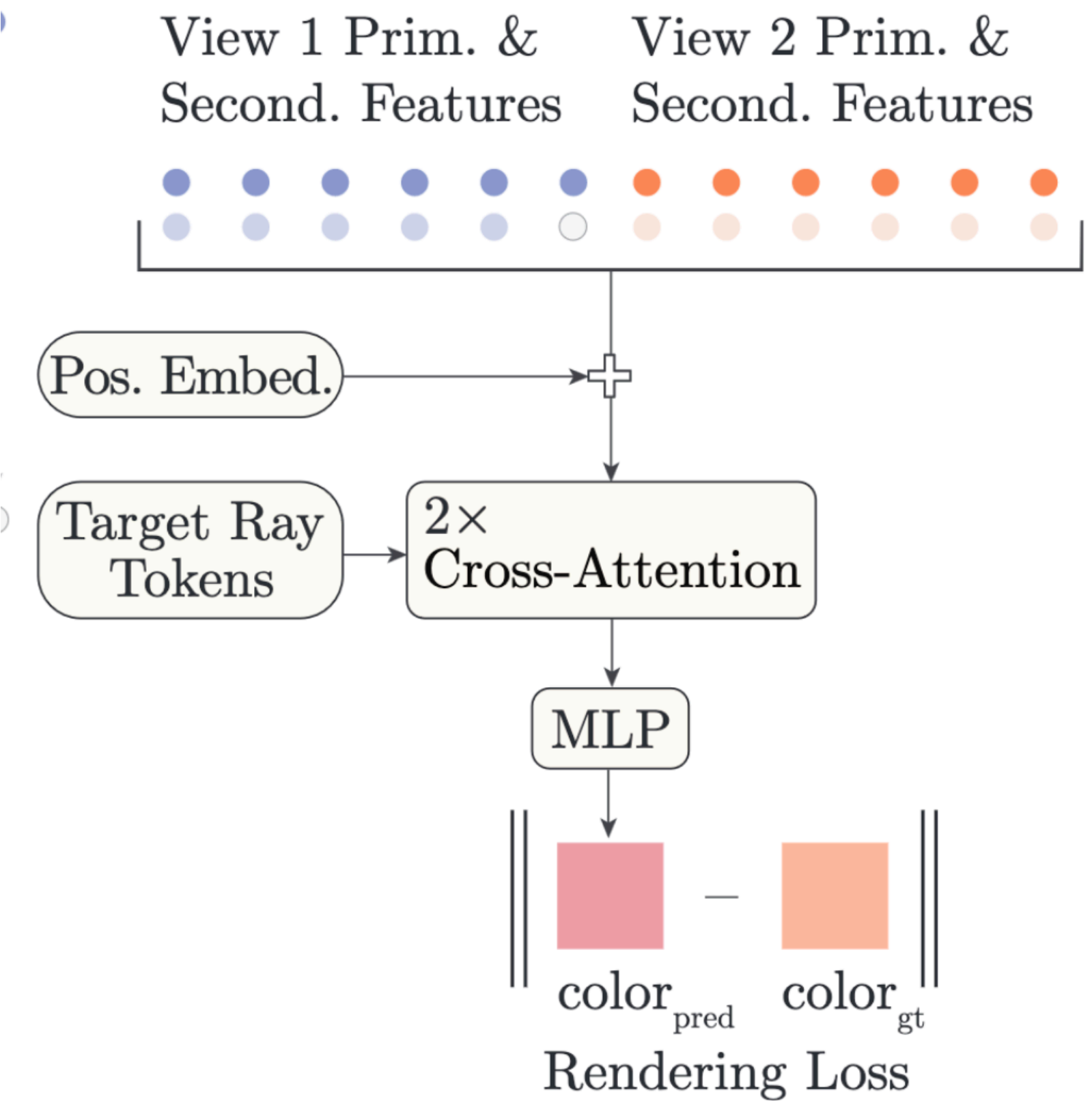


Piece 2: Image-Centric Epipolar Line Sampling & Attention-based Rendering

(b) Image-Centric Epipolar Line Sampling

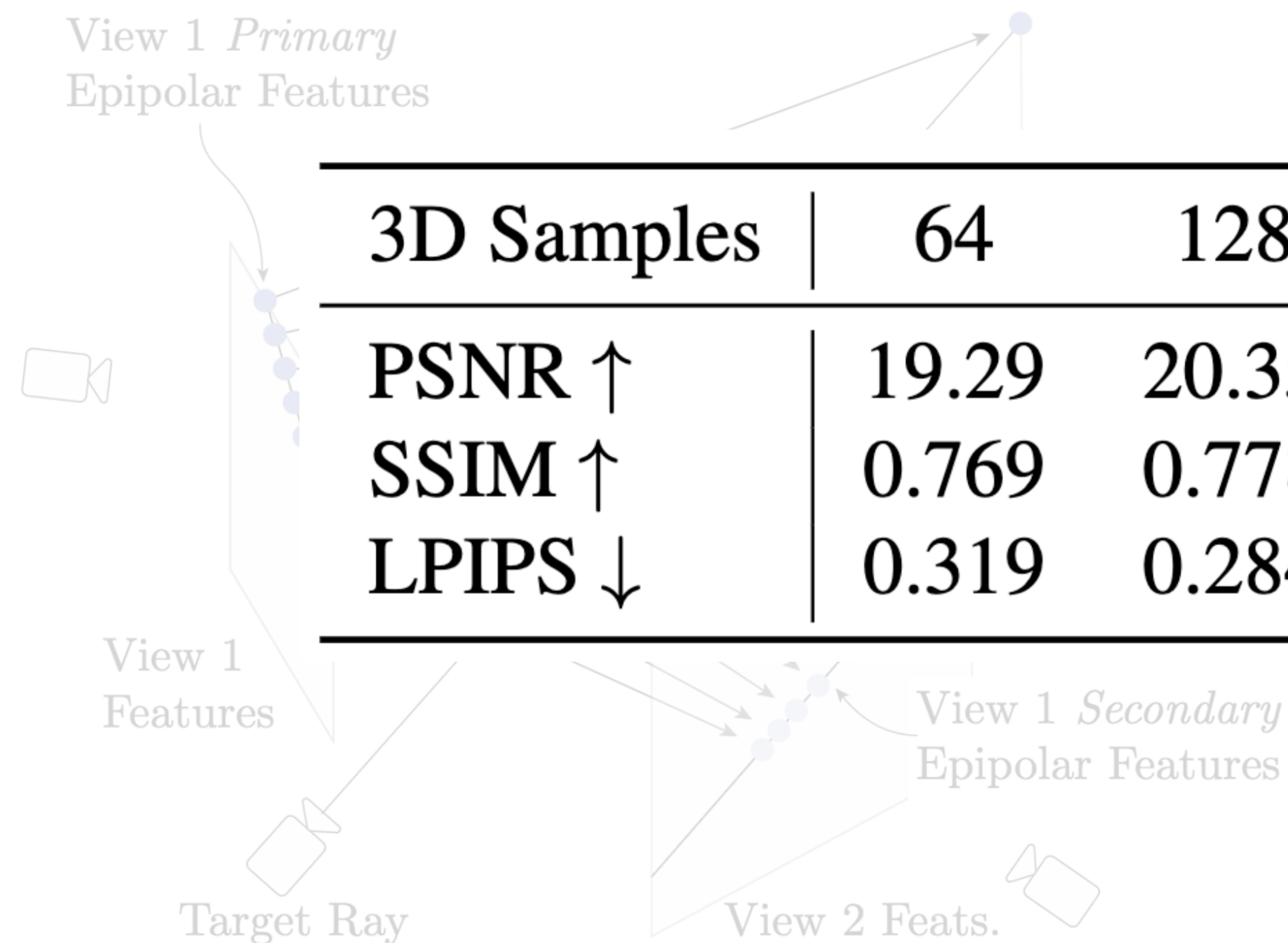


(c) Neural Rendering



Piece 2: Image-Centric Epipolar Line Sampling & Attention-based Rendering

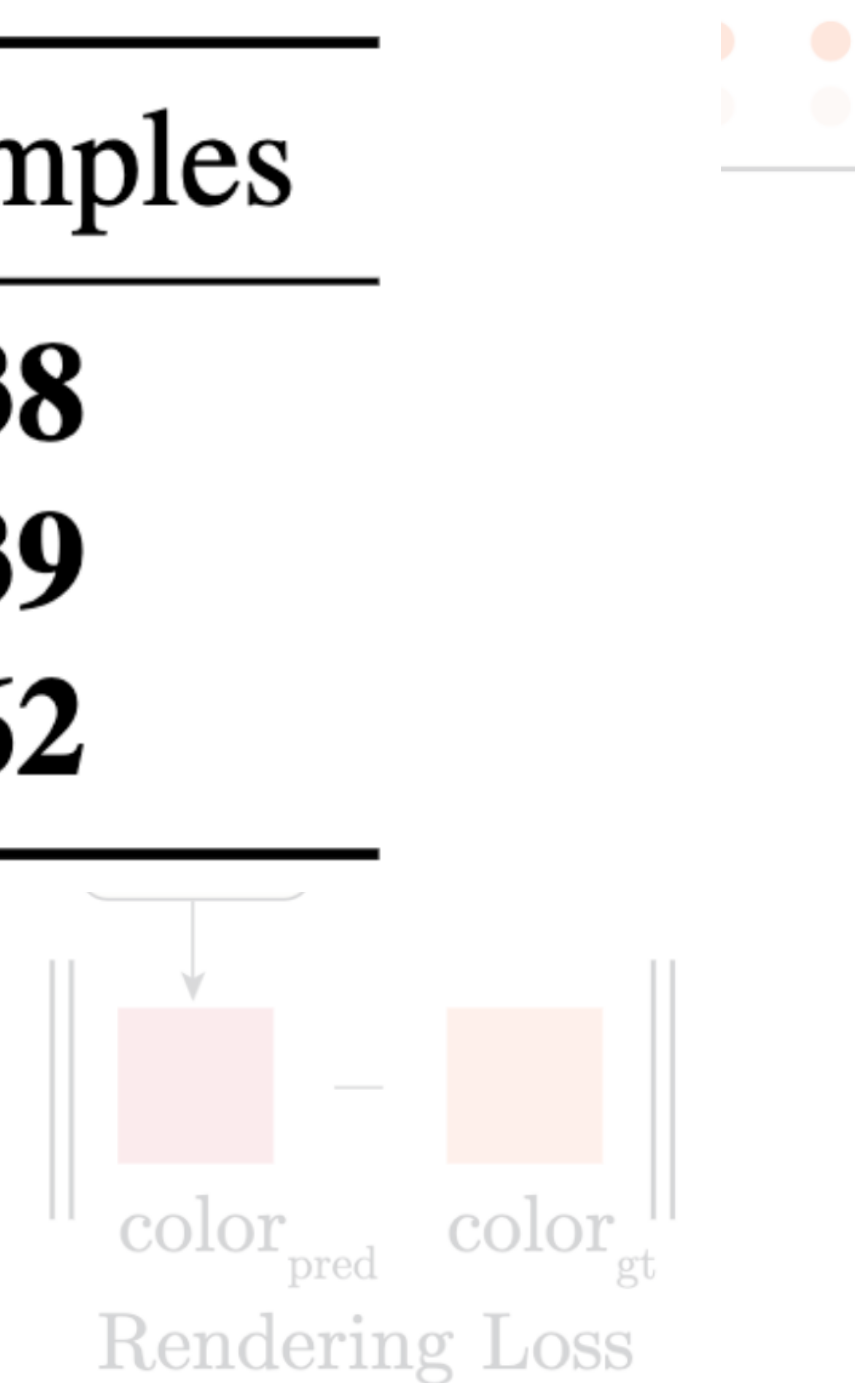
(b) Image-Centric Epipolar Line Sampling



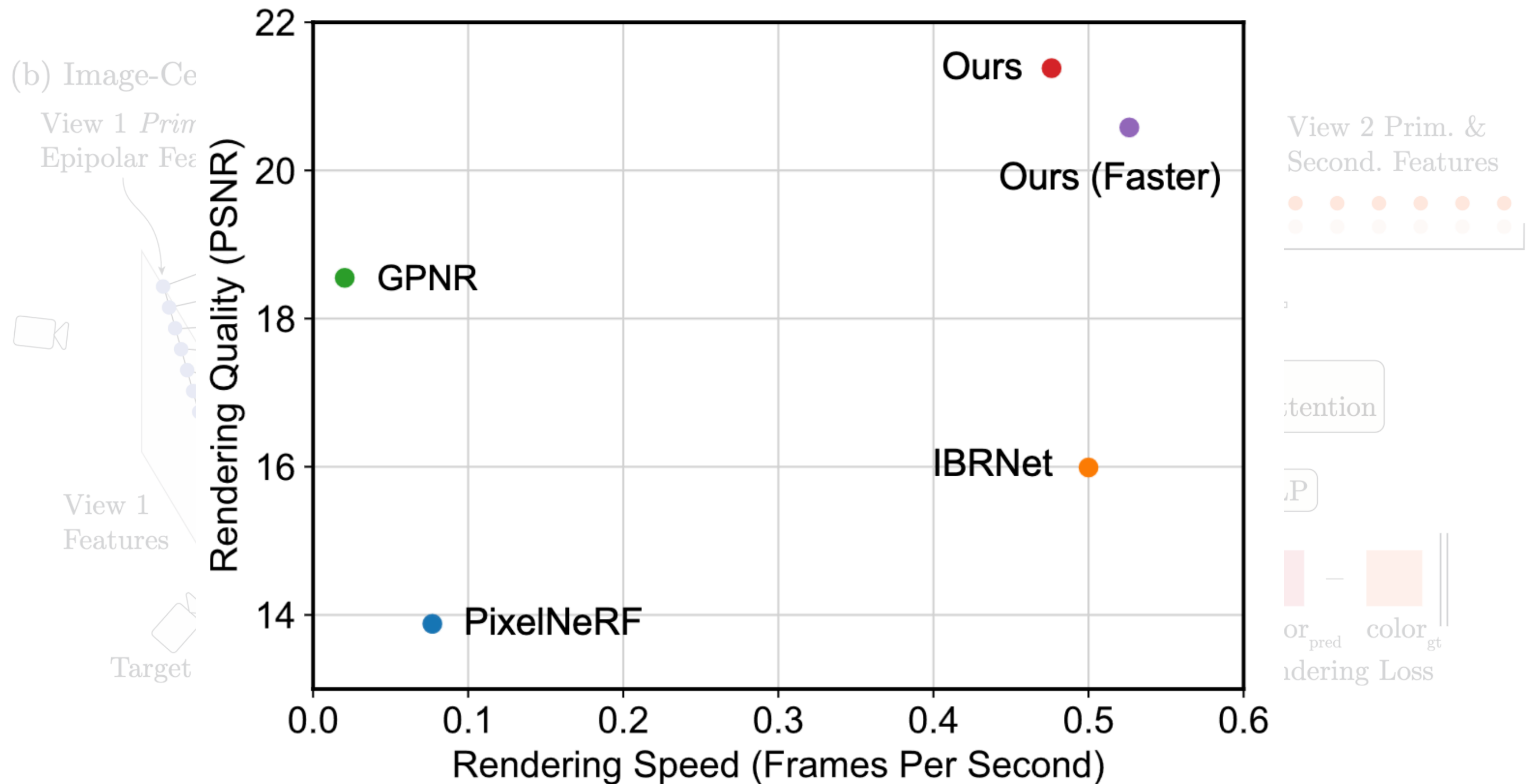
(c) Neural Rendering

View 1 Prim. & Second. Features View 2 Prim. & Second. Features

3D Samples	64	128	192	Epi. Samples
PSNR \uparrow	19.29	20.35	20.60	21.38
SSIM \uparrow	0.769	0.778	0.790	0.839
LPIPS \downarrow	0.319	0.284	0.273	0.262



Piece 2: Image-Centric Epipolar Line Sampling & Attention-based Rendering



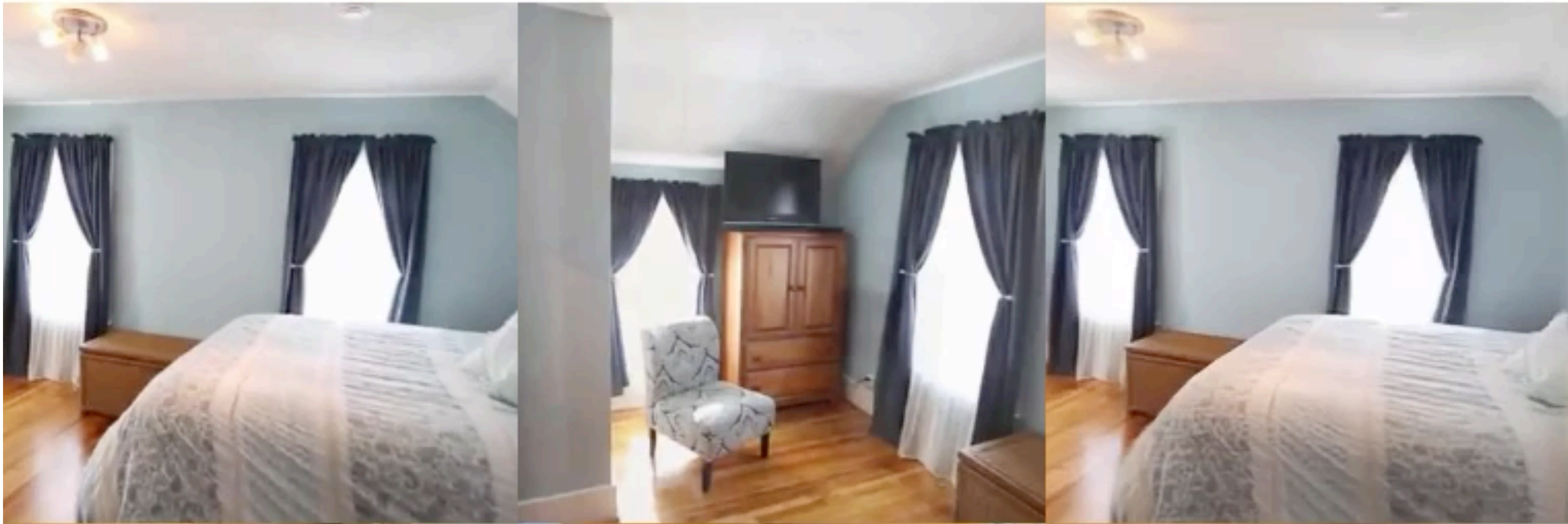
Input

Rendering



Input

Rendering



Baseline Comparison



Unposed Images

Input

Rendering



Input

Rendering





Massachusetts
Institute of
Technology



Learning to Render Novel Views from Wide-Baseline Stereo Pairs

Yilun Du, Cameron Smith, Ayush Tewari[†], Vincent Sitzmann[†]

[†] denotes equal advising

TUE-PM-079