Learning to Predict Scene-Level Implicit 3D from Posed RGBD Data

Nilesh Kulkarni, Linyi Jin, Justin Johnson, David Fouhey

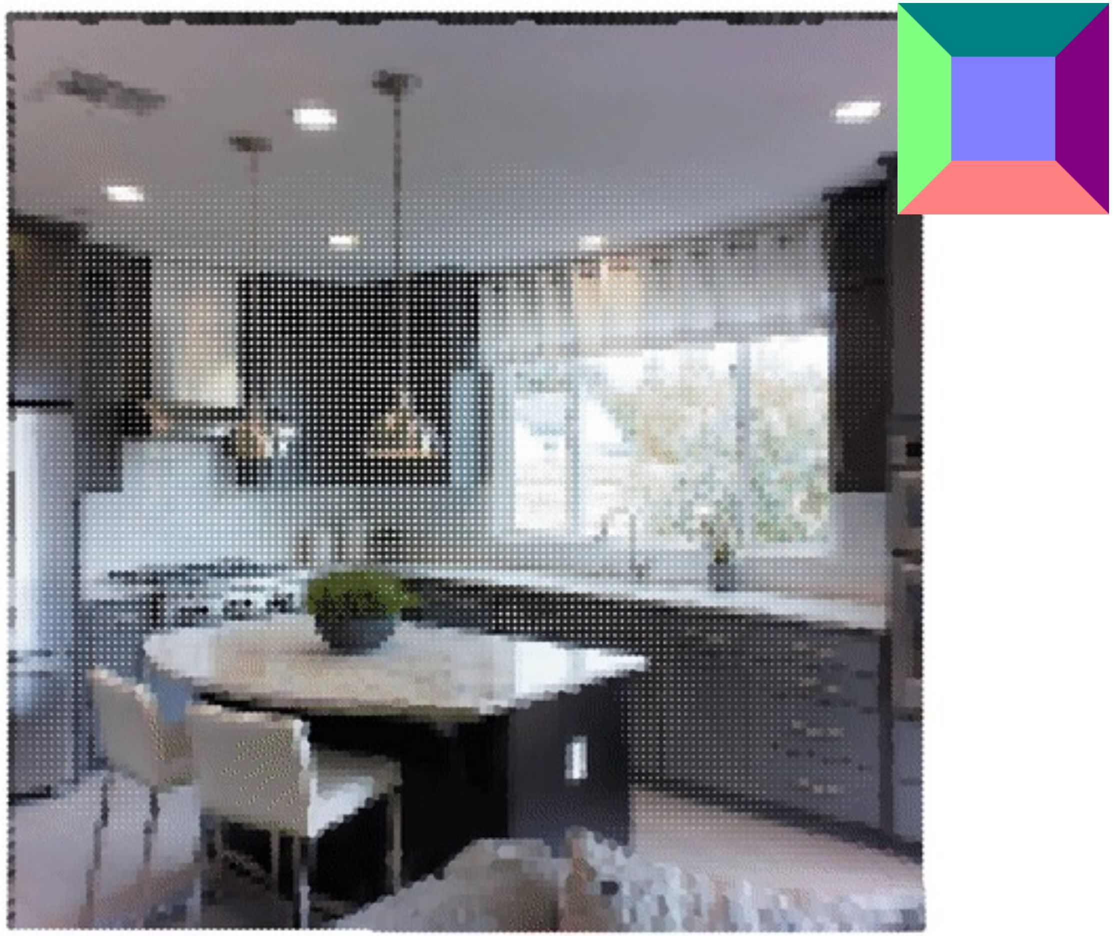
https://nileshkulkarni.github.io/d2drdf




Problem Statement (Goal)

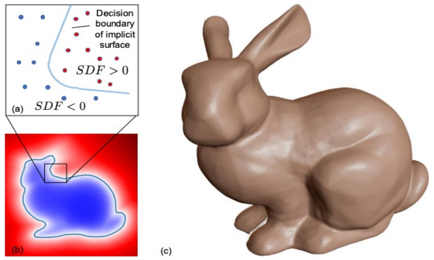
Reconstruct scene, including invisible surfaces

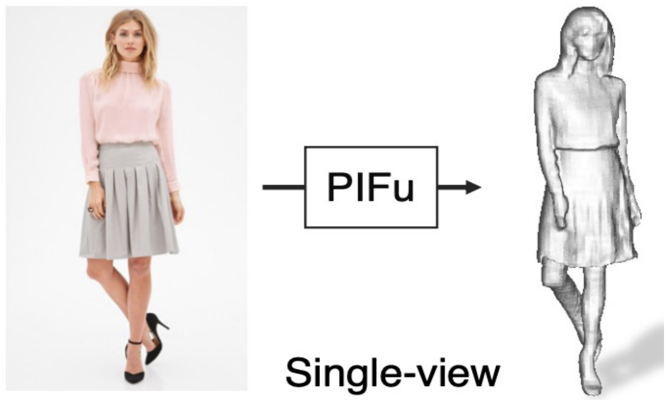
- *Test time:* single, previously unseen RGB image
- *Training time*: Posed RGBD data
- *Method*: implicit functions



Output from our system (D2-DRDF)

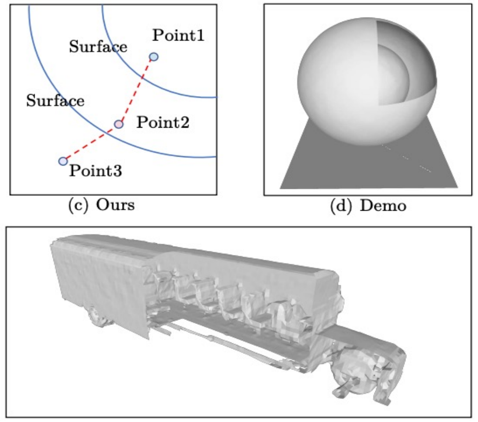
Prior Work

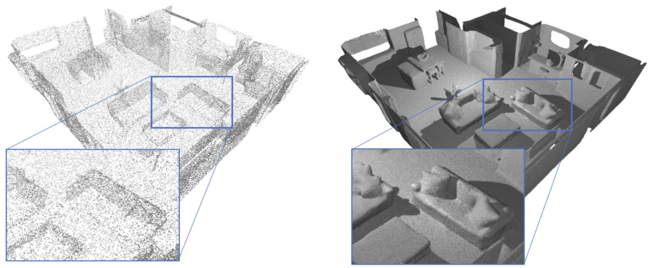




Learn using Watertight 3D Data.

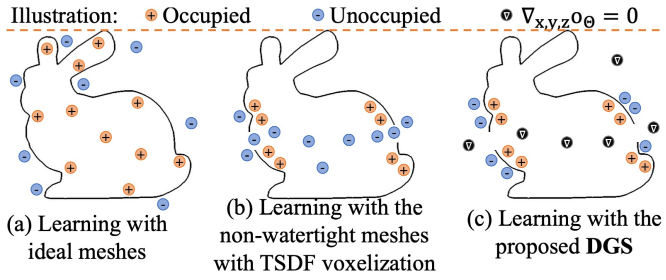
DeepSDF, Park et. al, Chen et. al, Michalkiewicz et al, Mescheder et. al
PiFU Saito et. al
SIREN, Sitzmann et. al

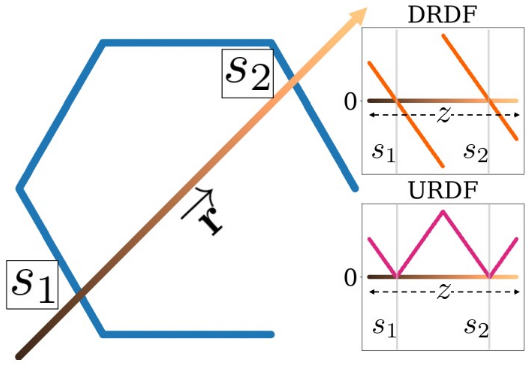




Learn on Non-Watertight 3D Data.
3D from point clouds

GIFS, Ye et. al
UNDF, Chibane et. al





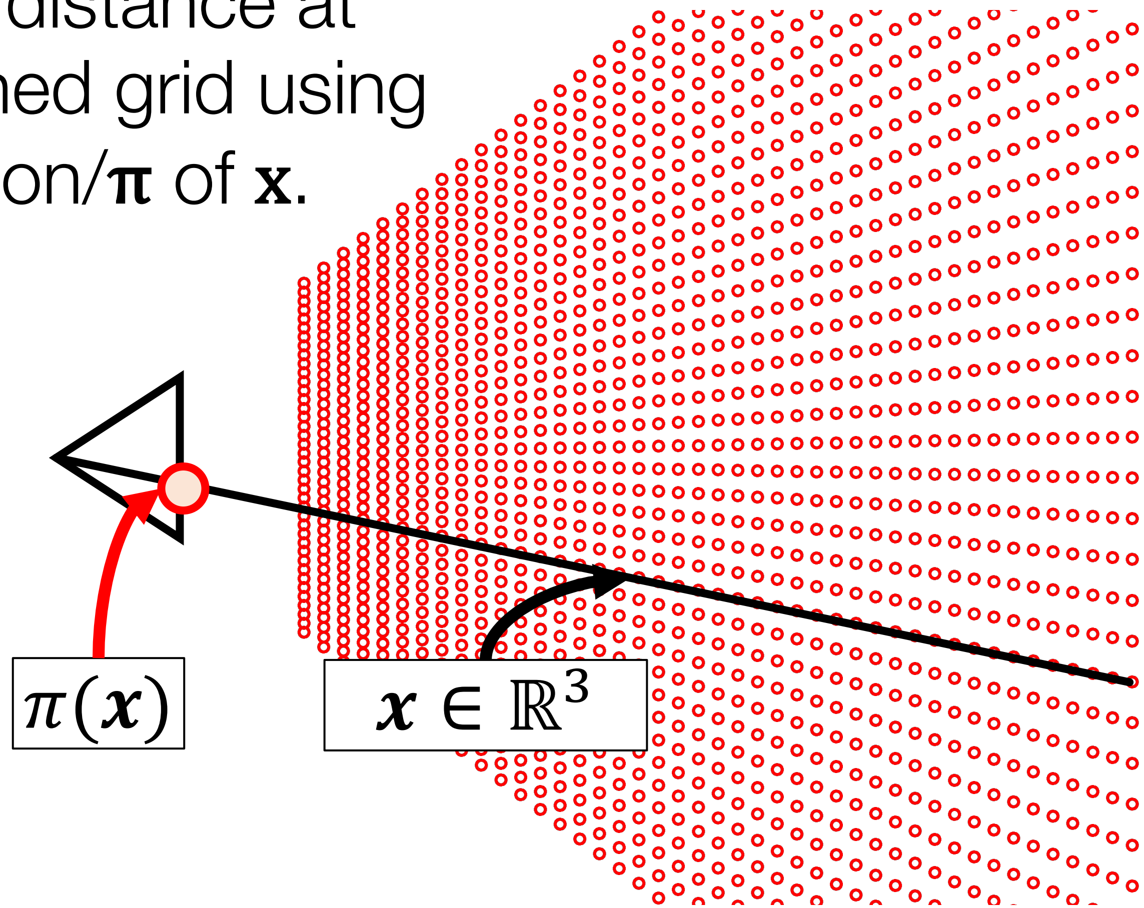
Learn on Non-Watertight 3D Data.
3D from single input image.

DGS, Zhu et. al
DRDF, Kulkarni et. al

D2-DRDF Method

D2-DRDF Method

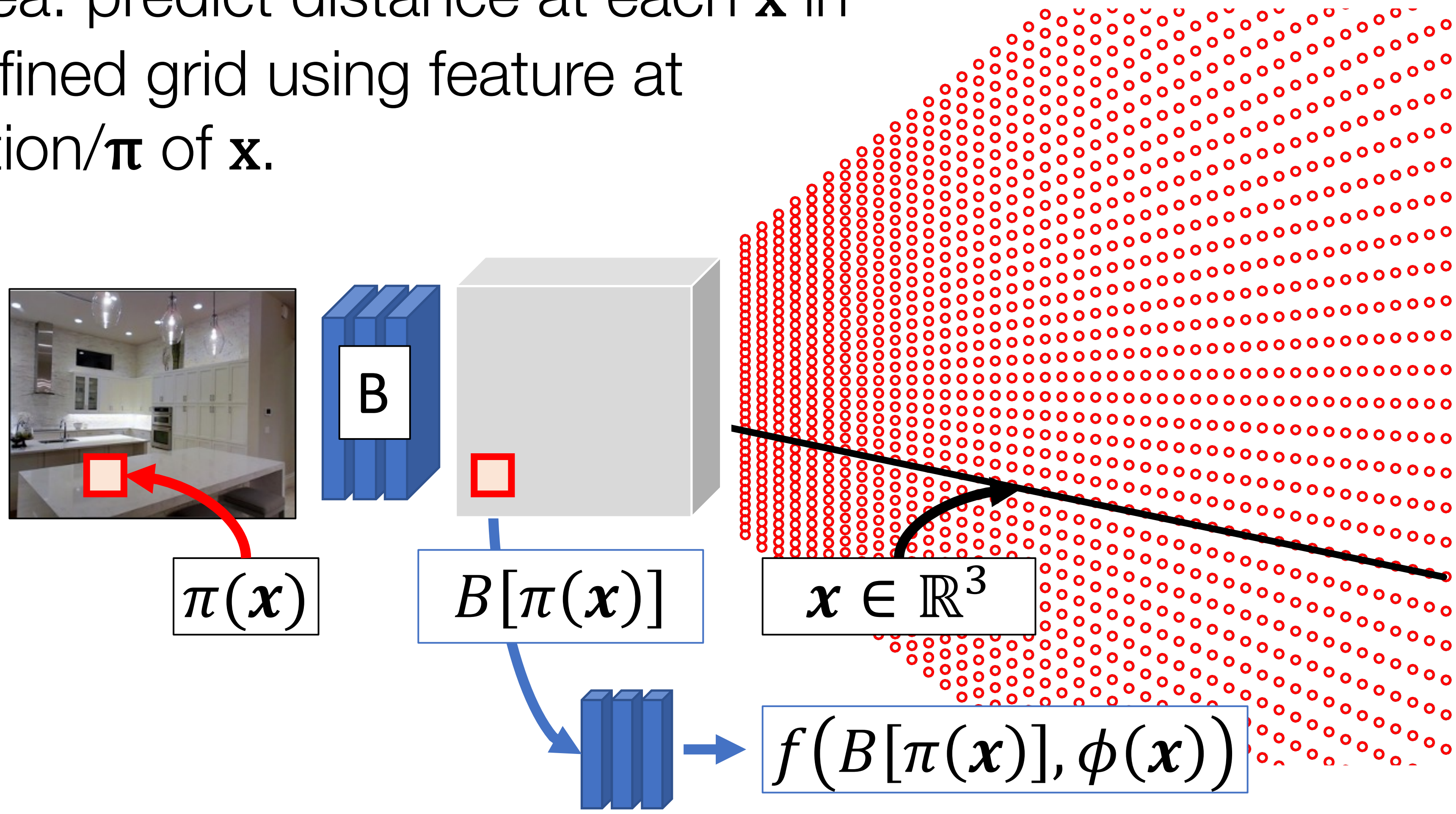
Key idea: predict distance at each $\mathbf{x}$ in predefined grid using feature at projection/$\boldsymbol{\pi}$ of $\mathbf{x}$.



See *pixelNeRF: Neural Radiance Fields from One or Few Images*, Yu et al. CVPR 2021

D2-DRDF Method

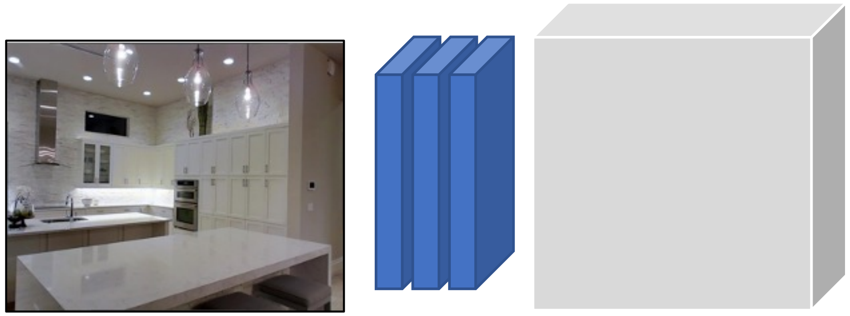
Key idea: predict distance at each $\mathbf{x}$ in pre-defined grid using feature at projection/$\boldsymbol{\pi}$ of $\mathbf{x}$.

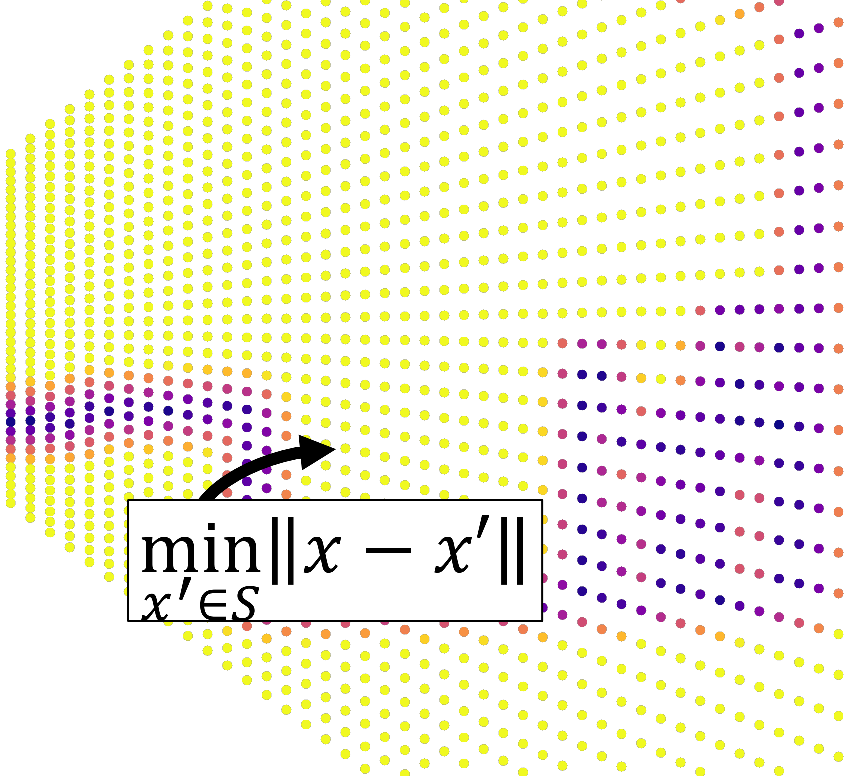


See *pixelNeRF: Neural Radiance Fields from One or Few Images*, Yu et al. CVPR 2021

Example Output

Output: Grid of unsigned distance function (UDF) to nearest surface in 3D



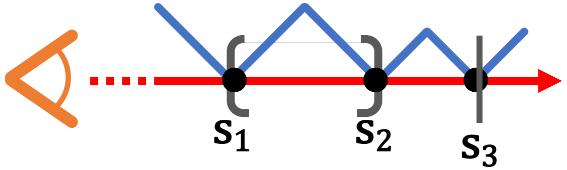


Key detail: Decoding function to convert distances to surfaces.

See *pixelNeRF: Neural Radiance Fields from One or Few Images*, Yu et al. CVPR 2021

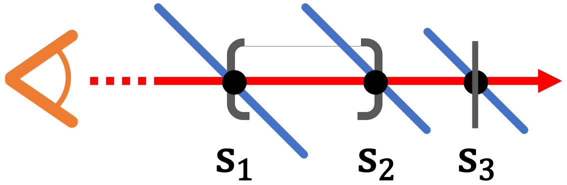
Ray Distance Functions

Unsigned Ray Distance Function

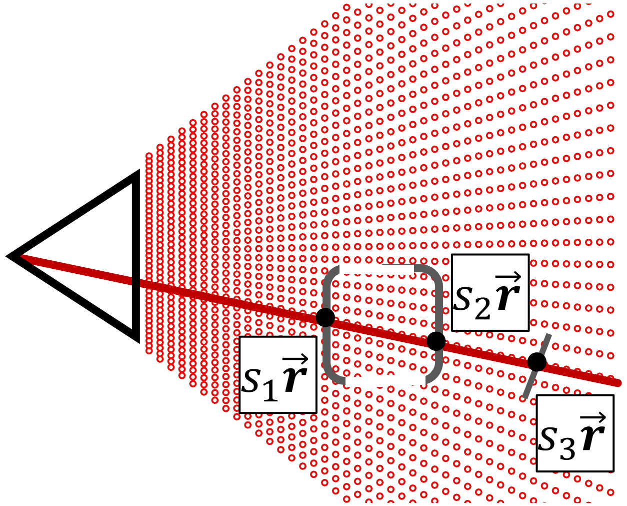


Decoding: Values within ϵ of 0

Directed Ray Distance Function



Decoding: Positive to Negative Zero crossings



Pros of using DRDF (*Kulkarni et. al*)

1. Ray *v.s.* Scene (UDF)
2. Ease of decoding
3. Better behavior under uncertainty

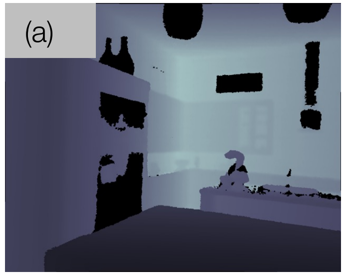
D2-DRDF Supervision with Posed Depth Data

Reference View

We wish to supervise the DRDF functions for all ray originating form this *reference camera* ; using the posed depth data

D2-DRDF Supervision with Posed Depth Data

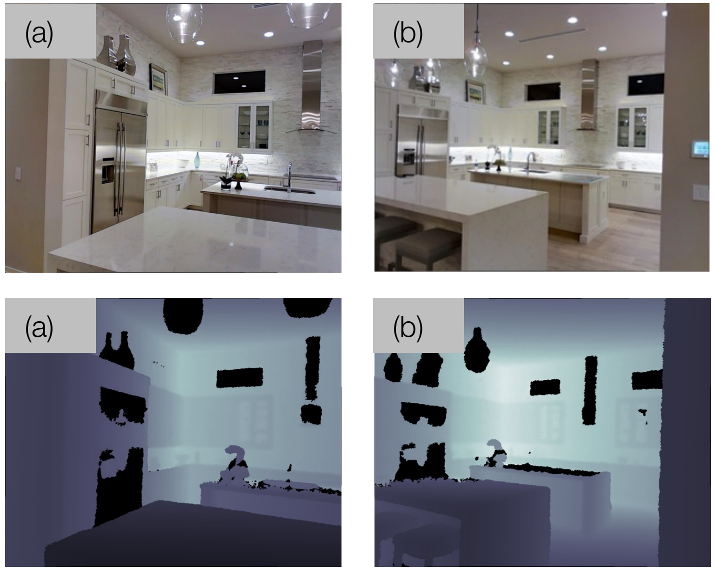
Reference View



Auxiliary Views

D2-DRDF Supervision with Posed Depth Data

Reference View



Auxiliary Views

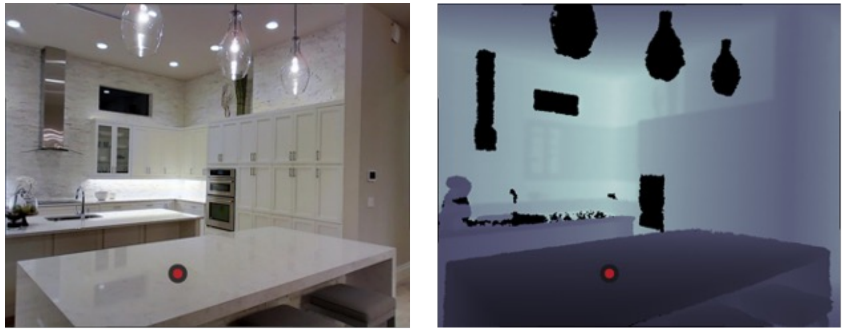
D2-DRDF Supervision with Posed Depth Data

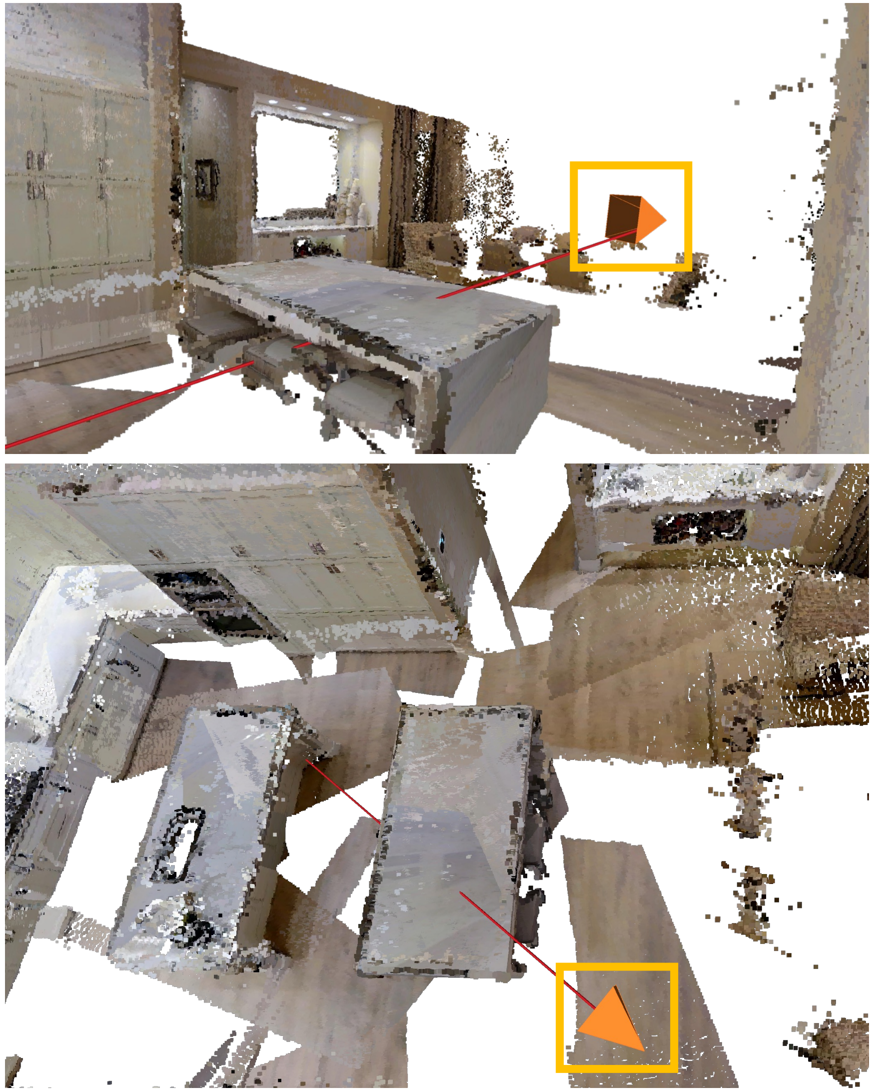
Reference View

Auxiliary Views

Ray based Supervision for the Reference View

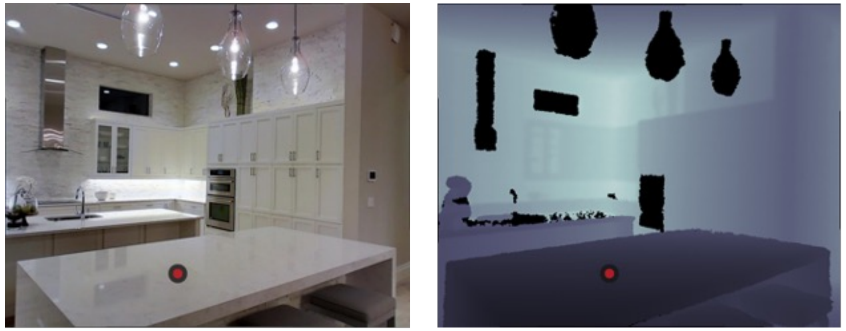
Reference View

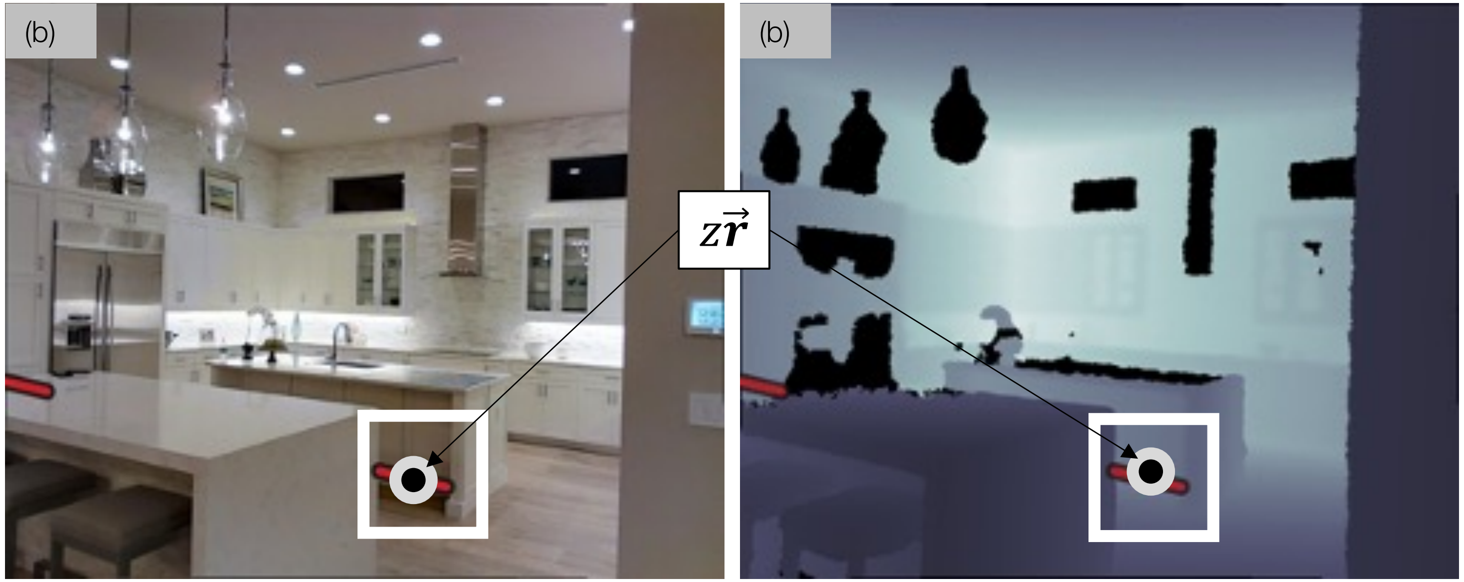


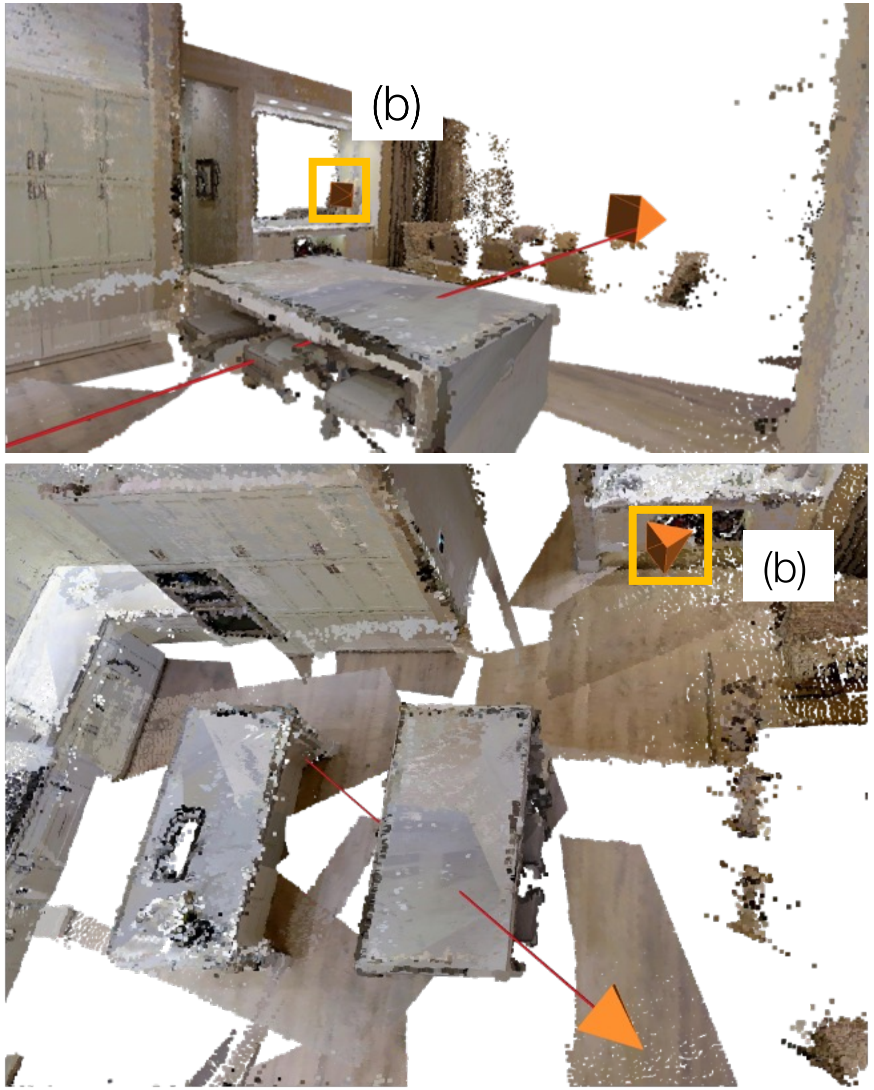


Ray based Supervision for the Reference View

Reference View

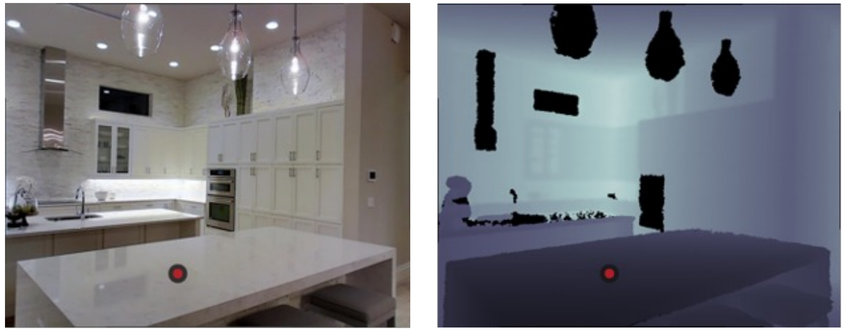


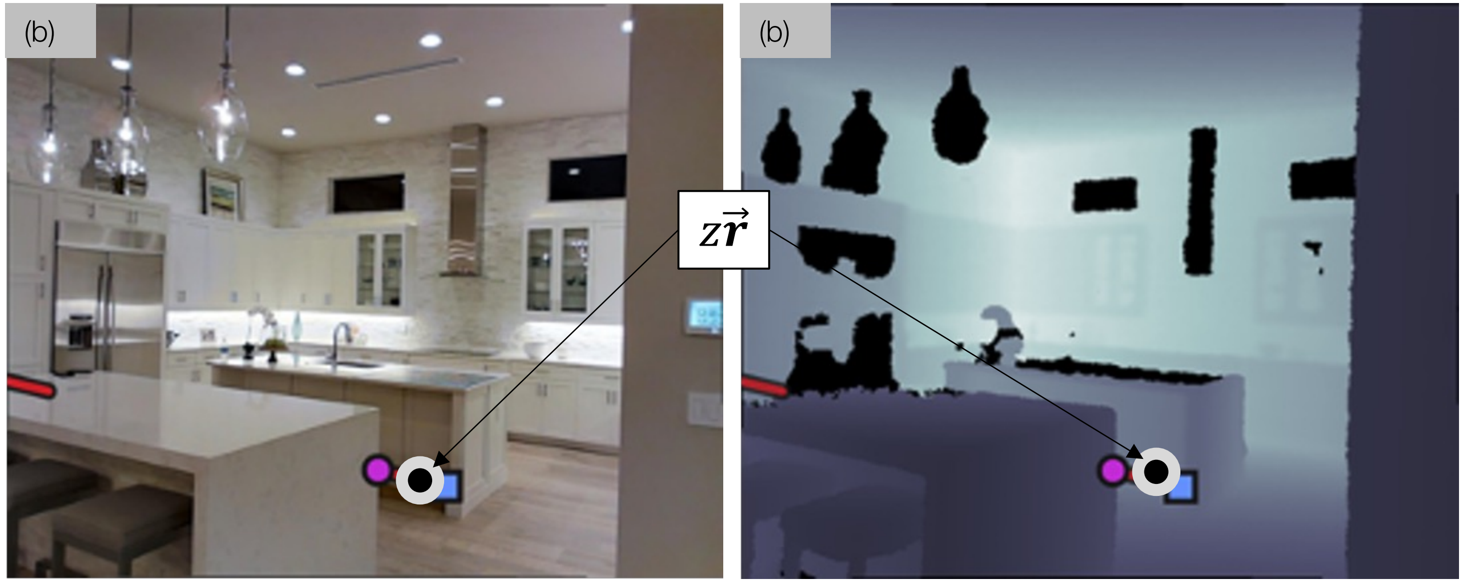




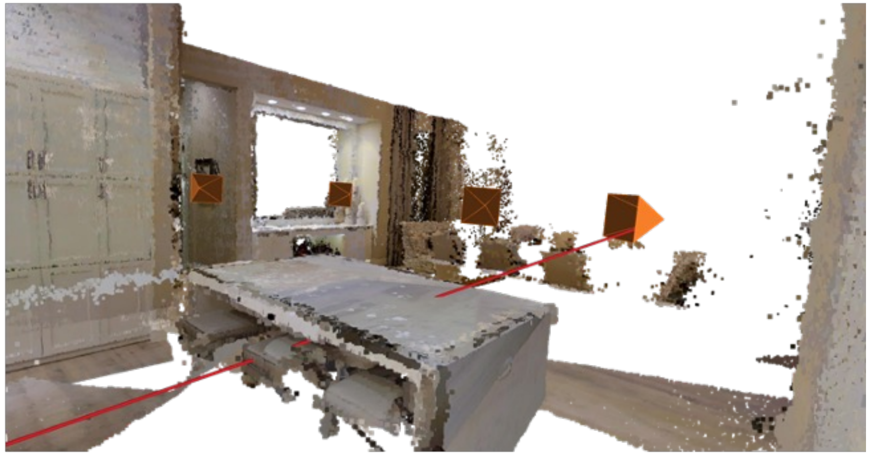
Ray based Supervision for the Reference View

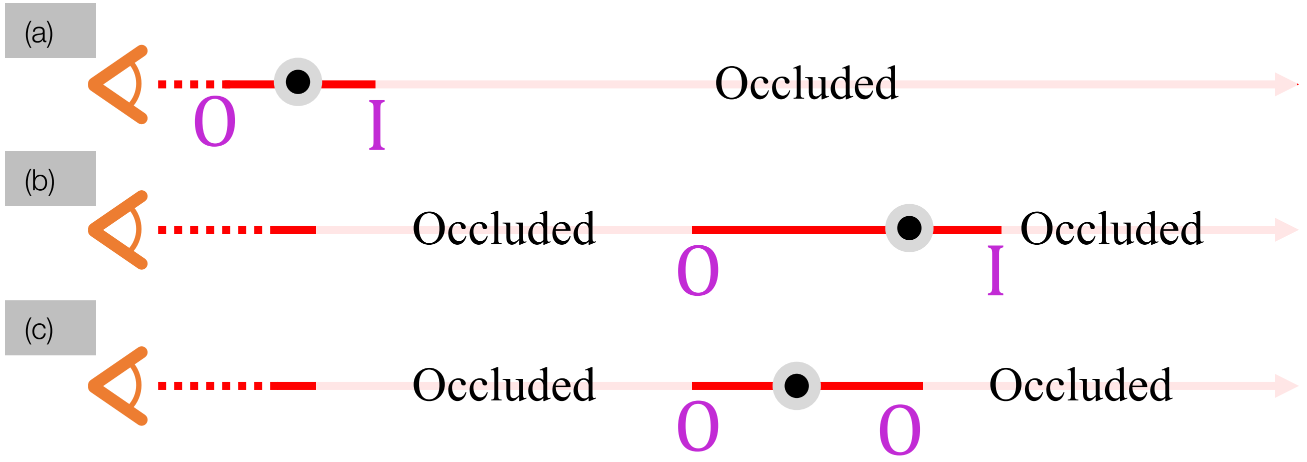
Reference View



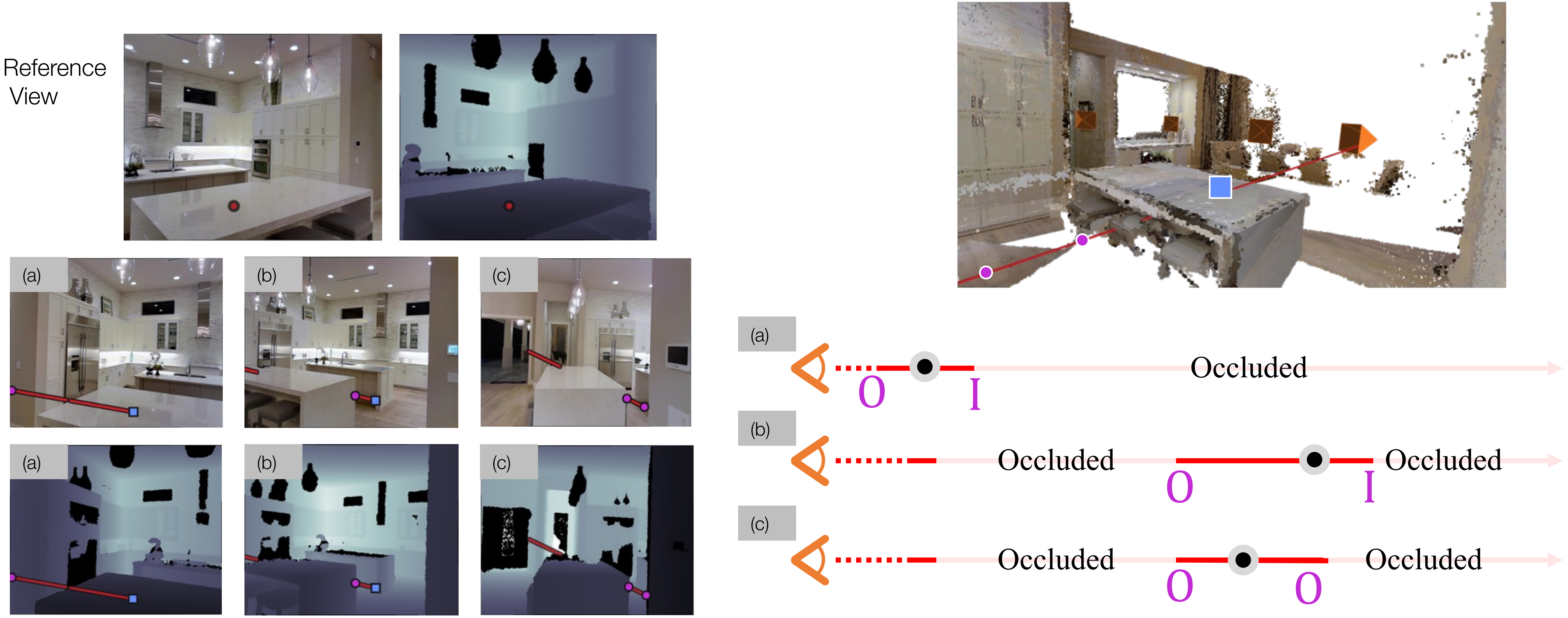


Segments along the red ray

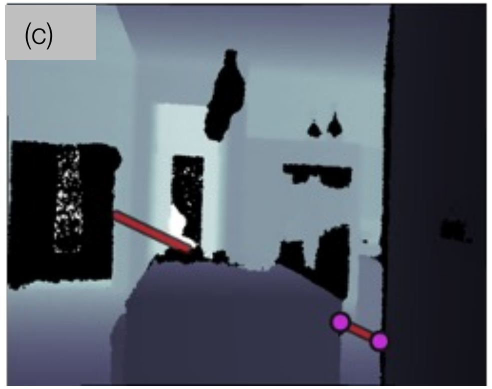


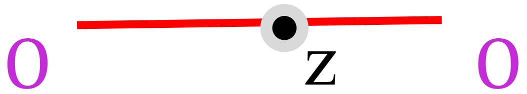


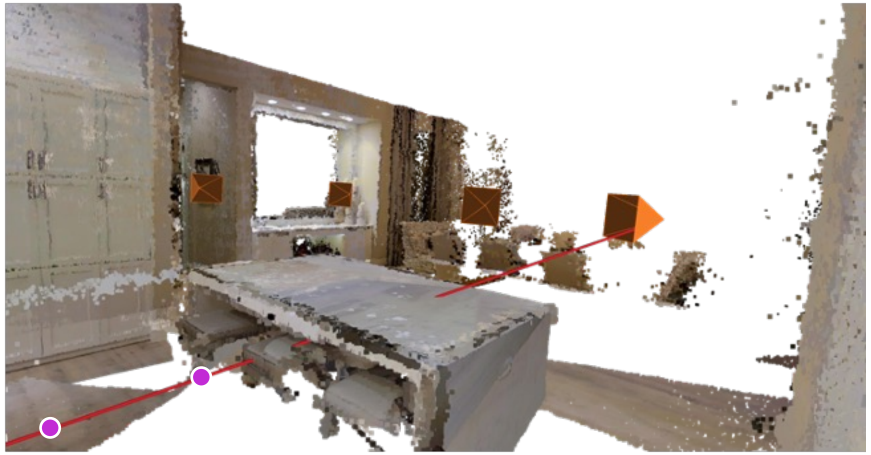
Segments along the red ray



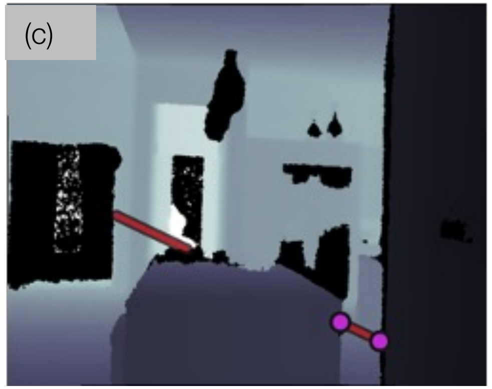
Segment 00

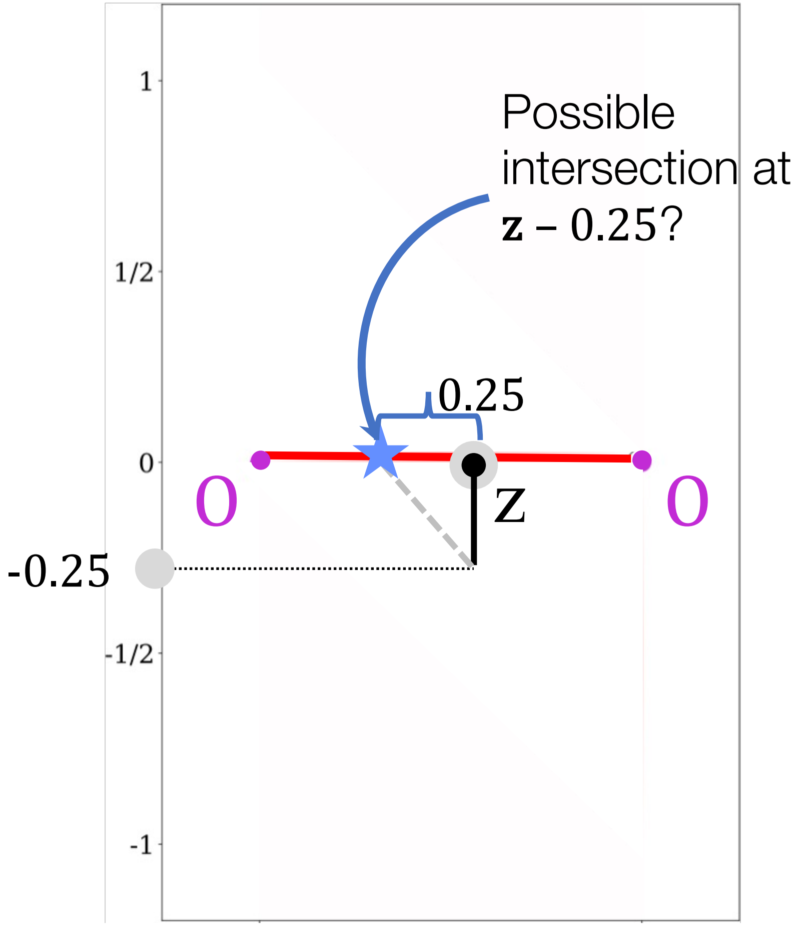


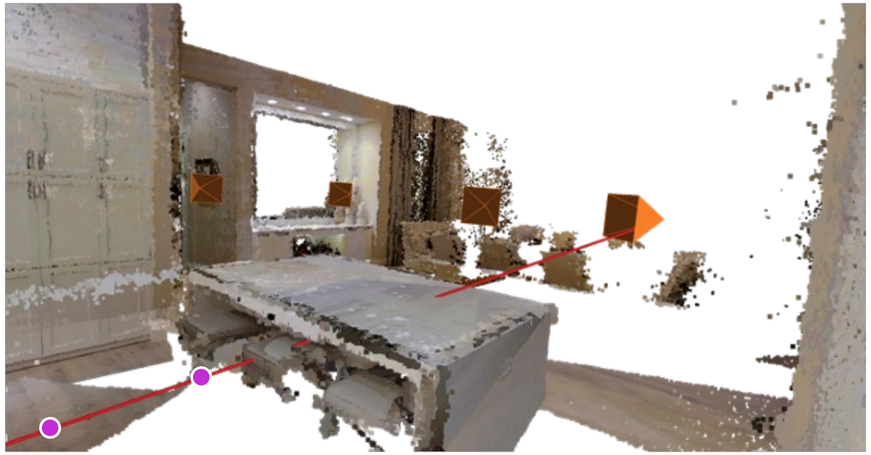




Segment 00

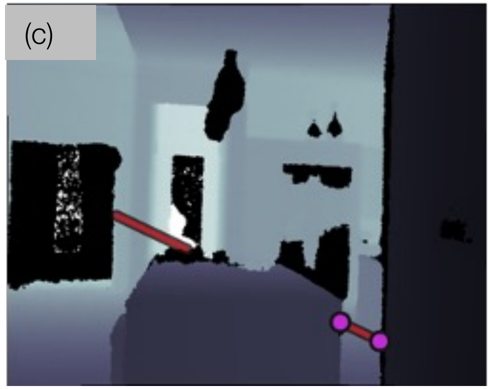


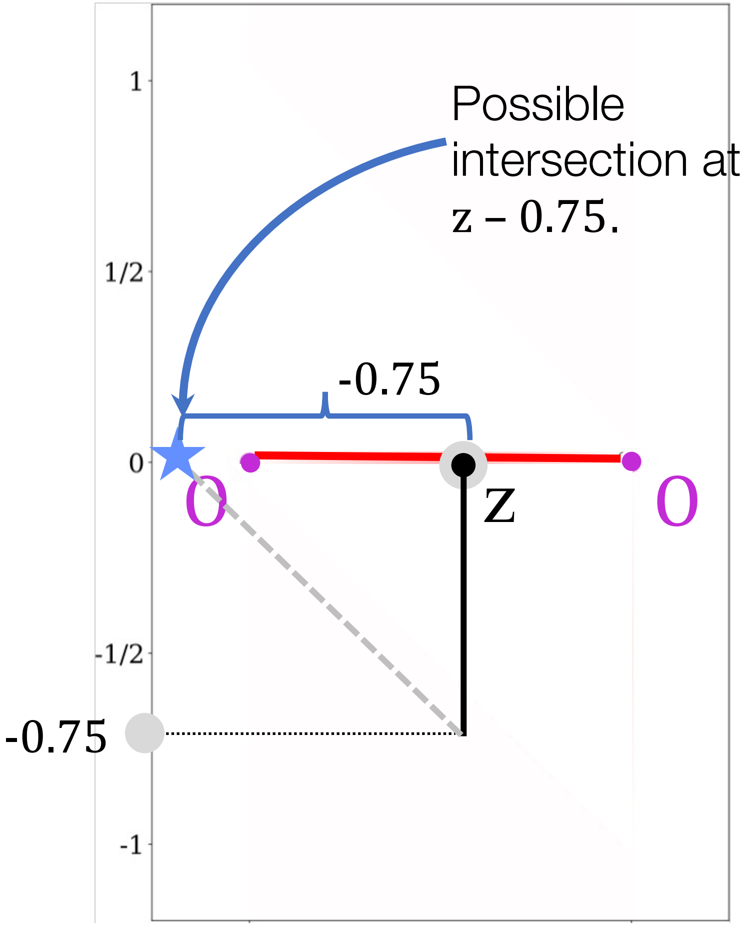


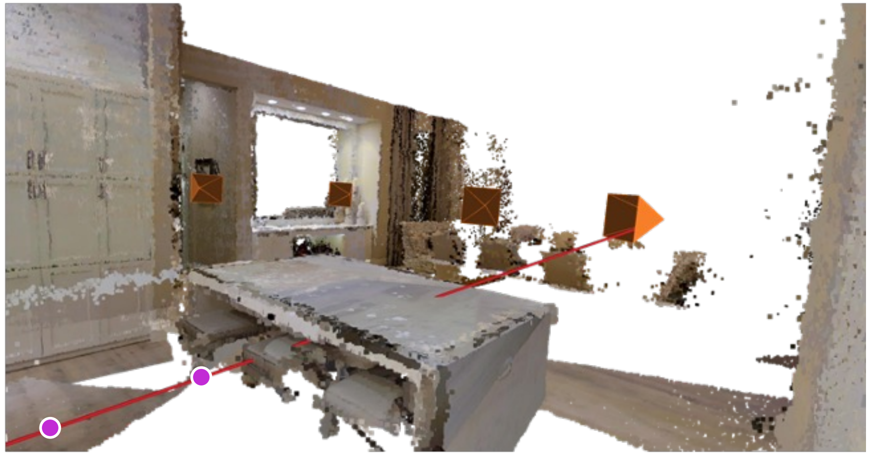


Disagrees with evidence in view (c)

Segment 00

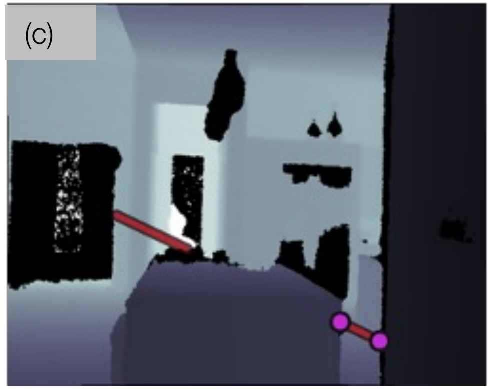


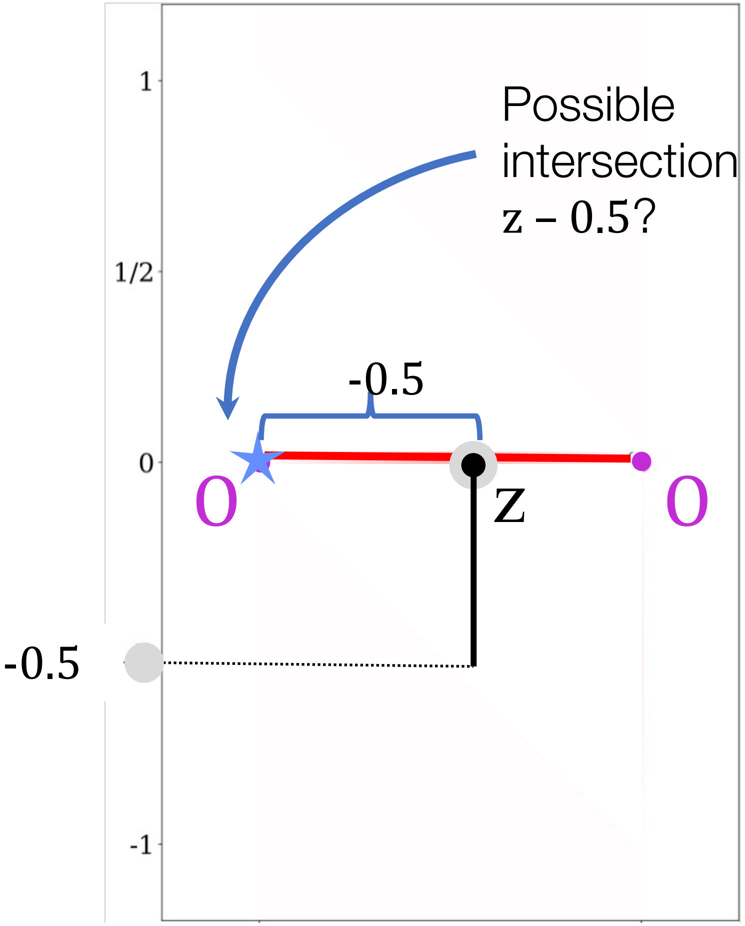


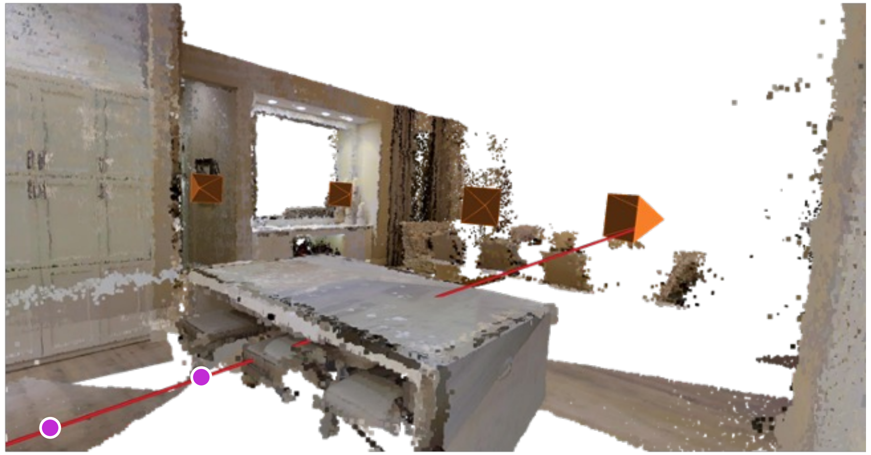


No evidence in view (c) to refute this intersection

Segment OO

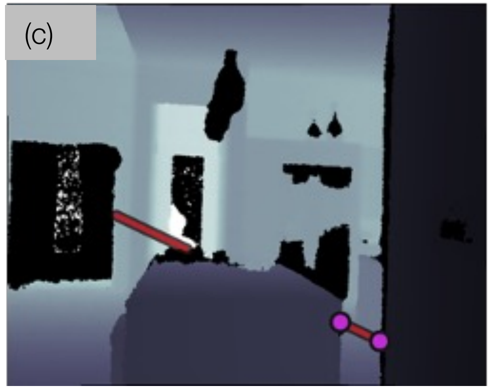


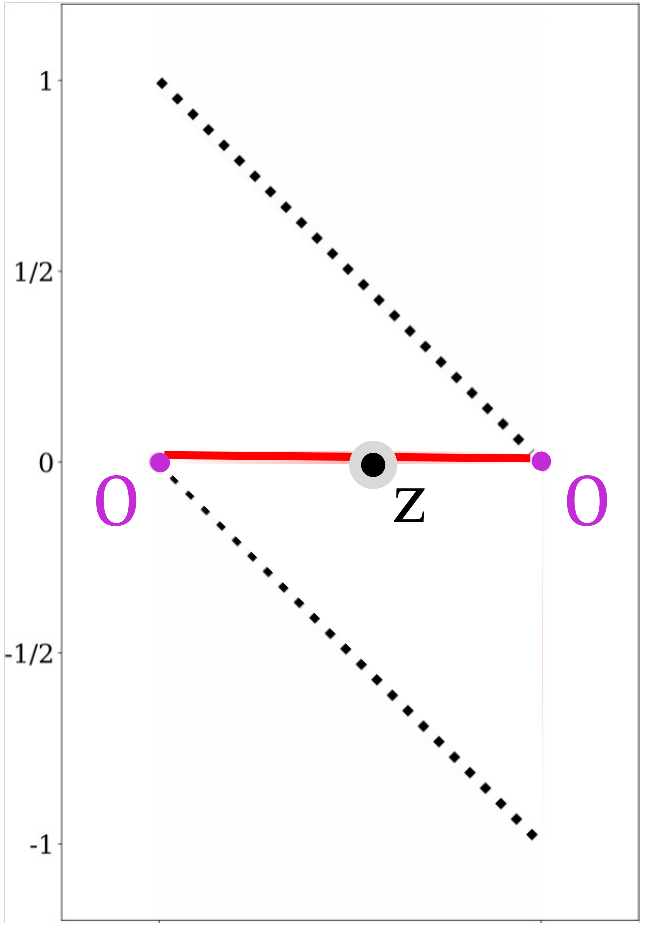


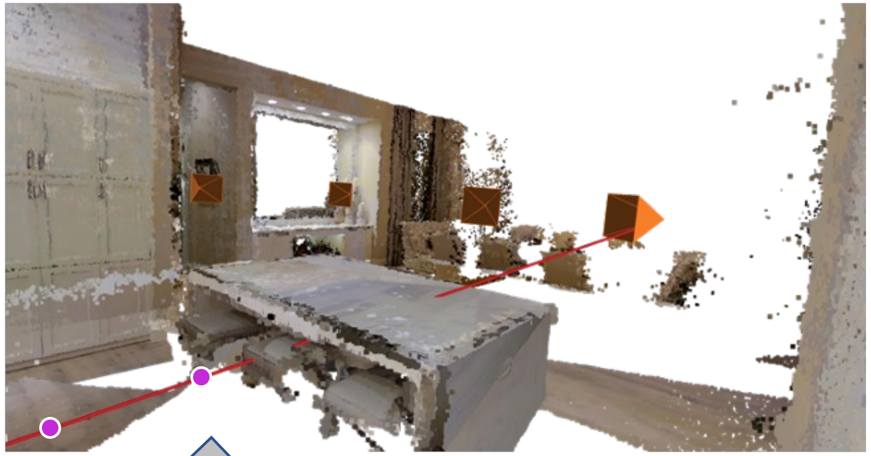


This is the limiting case of our analysis

Segment 00

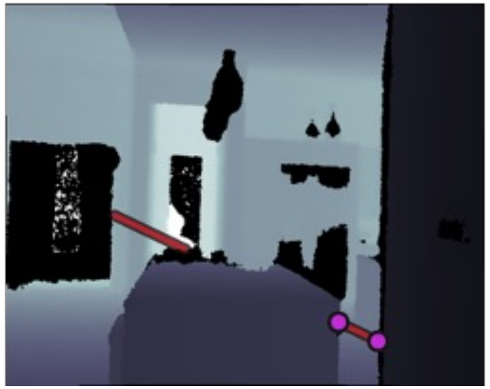


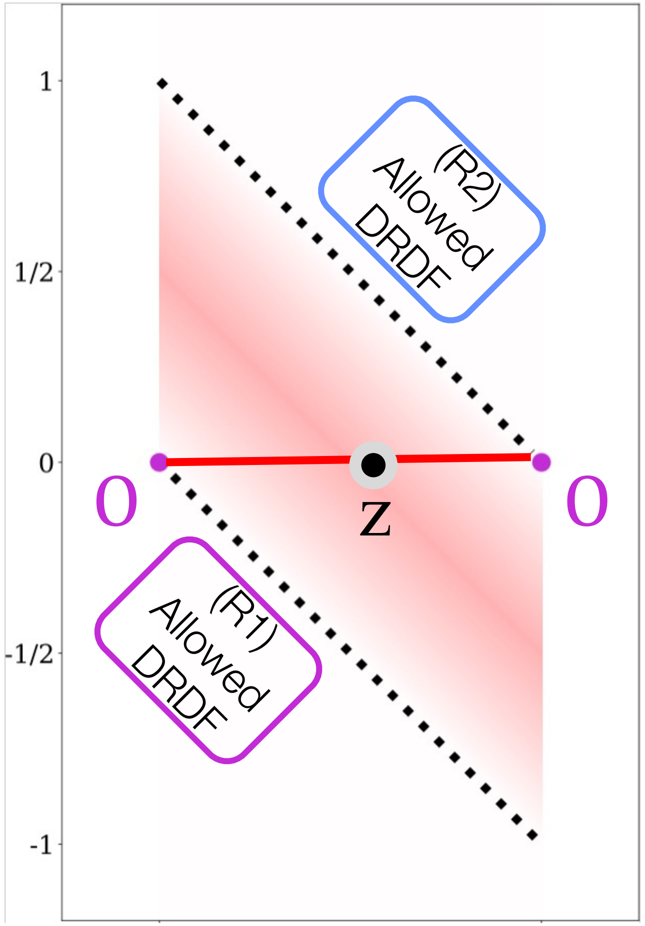


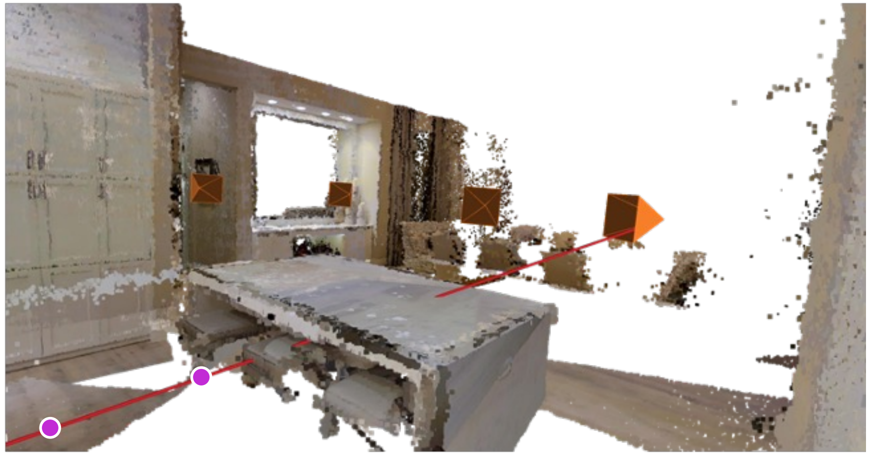


DRDF Values must be beyond the lines passing through the 0 events with a slope of -1

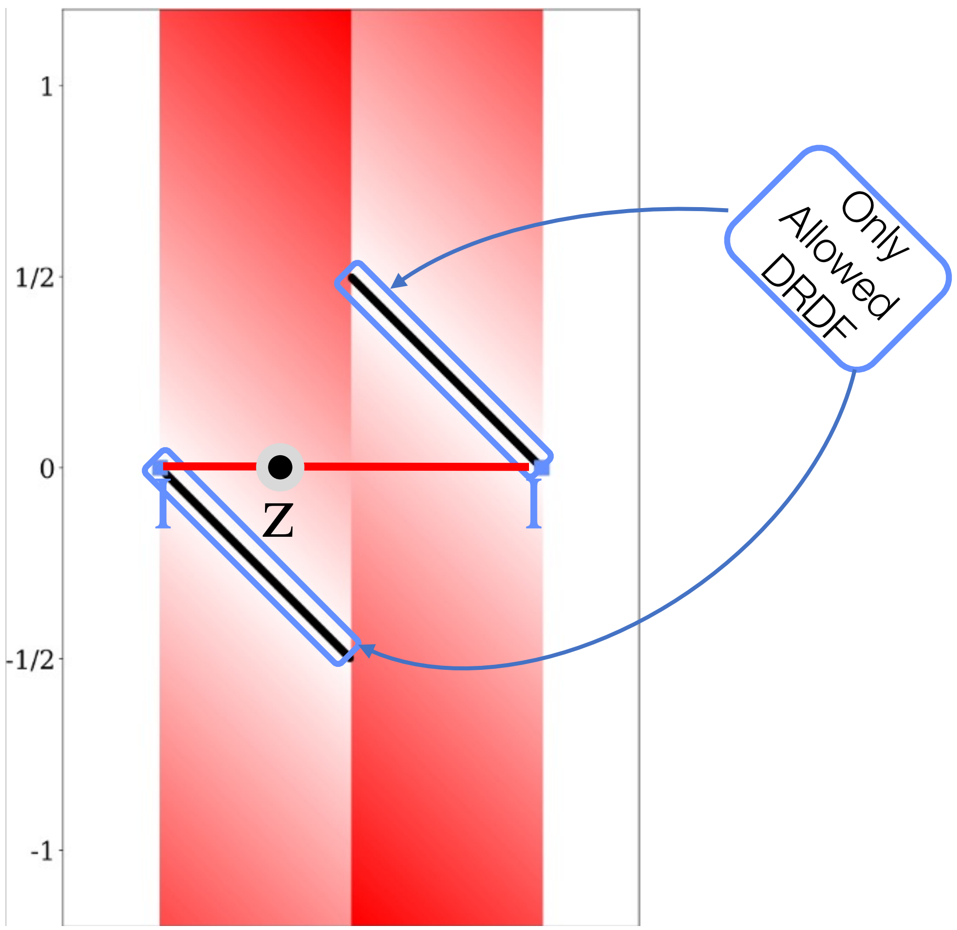
Segment 00



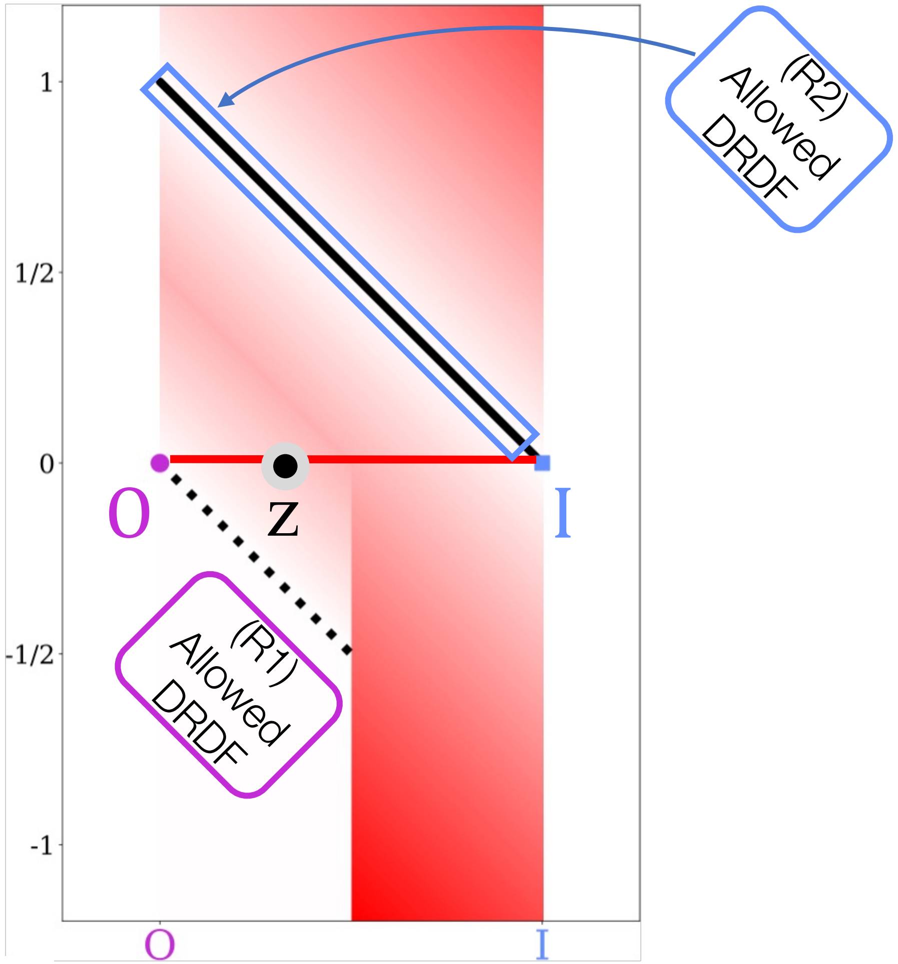




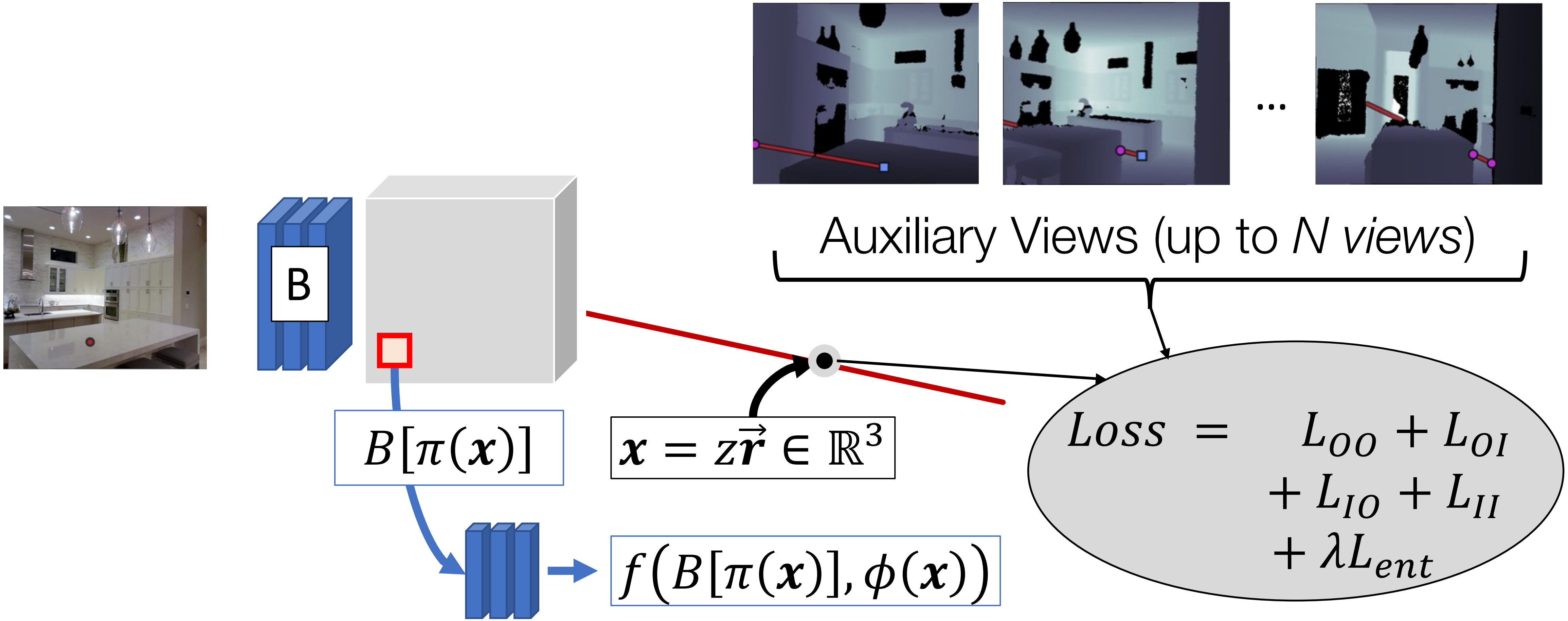
Segment II



Segment OI

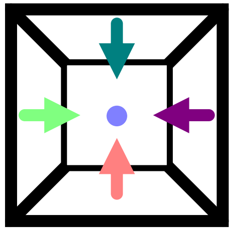


Training: Putting it all together

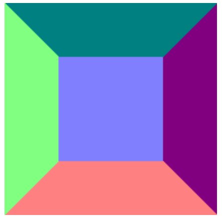


Matterport Qualitative Results

Cube



Legend



Quantitative Numbers (Matterport)

Supervision	Method		Scene Acc	Scene Cmp	Scene F1	Ray Acc	Ray Cmp	Ray F1
3D Mesh	Mesh	LDI [46]	66.2	72.4	67.4	13.9	42.8	19.3
		UDF [5]	58.7	76.0	64.7	15.5	23.0	16.6
		ORF	73.4	69.4	69.6	26.2	20.5	21.6
		URDF [5]	74.5	67.1	68.7	24.9	20.6	20.7
		DRDF [27]	75.4	72.0	71.9	28.4	30.0	27.3
Posed RGBD	D2-DRDF		73.7	73.5	72.1	28.2	22.6	25.1
	Density Field [57]		45.8	80.2	57.5	24.8	14.0	17.9

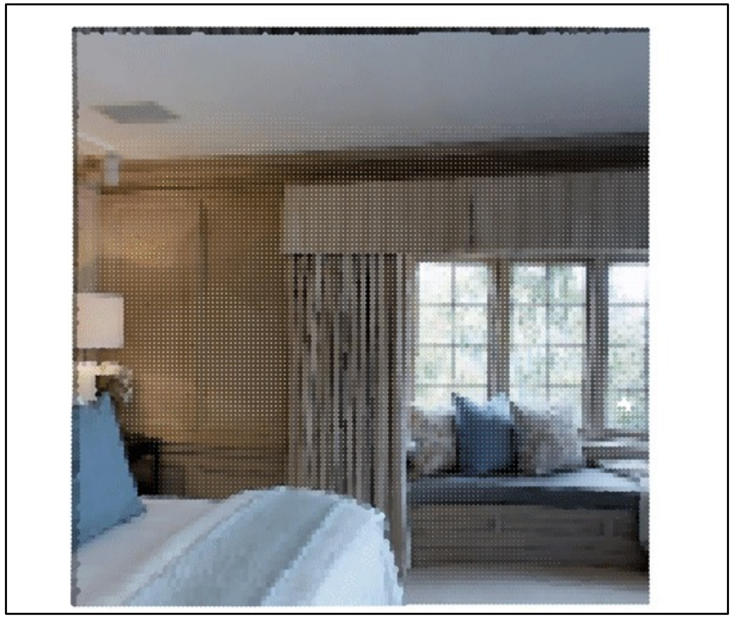
D2-DRDF is competitive , is the second-best method in the table

Robustness to Sparse Data (Matterport)

		Scene F1		Ray Occ F1	
Im%	3D%	Mesh [27]	Depth (ours)	Mesh [27]	Depth (ours)
100	100	71.9	72.1	27.3	25.1
50	56	68.4 (-3.5)	70.0 (-2.1)	23.6 (-3.7)	24.4 (-0.7)
25	43	66.8 (-5.1)	70.0 (-2.1)	21.2 (-6.1)	24.9 (-0.2)

D2-DRDF (ours) is much more robust and has 0.2% drop in F1 as compared to 6.1% for Mesh-DRDF [27] on 25% Image data.

Adaptation to New Scenes



D2 - DRDF

D2-DRDF + Adapt

Adaptation allows D2-DRDF to get finer details accurately
(like the green colored desk)

Thank you