UC San Diego

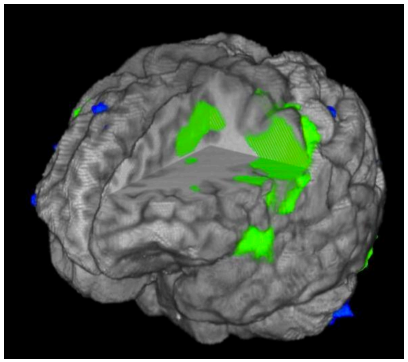
Lecture 12:

Deep Learning on Volumetric Representation

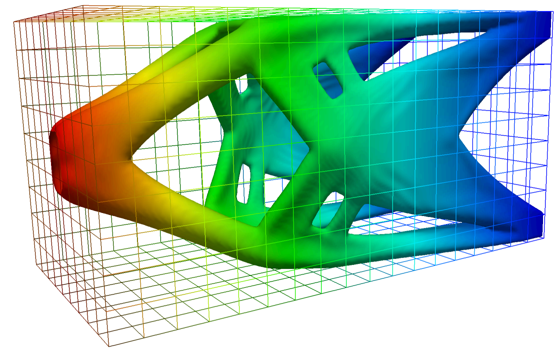
Instructor: Hao Su

Feb 19, 2018

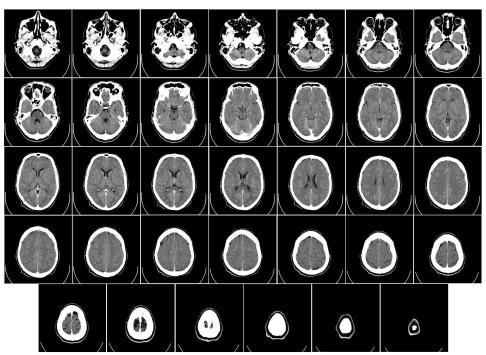
Popular 3D volumetric data



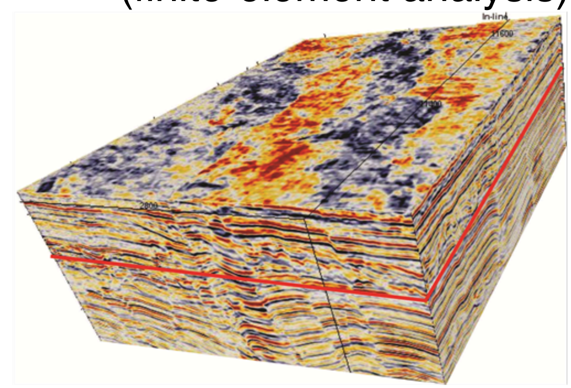
fMRI



Manufacturing
(finite-element analysis)



CT



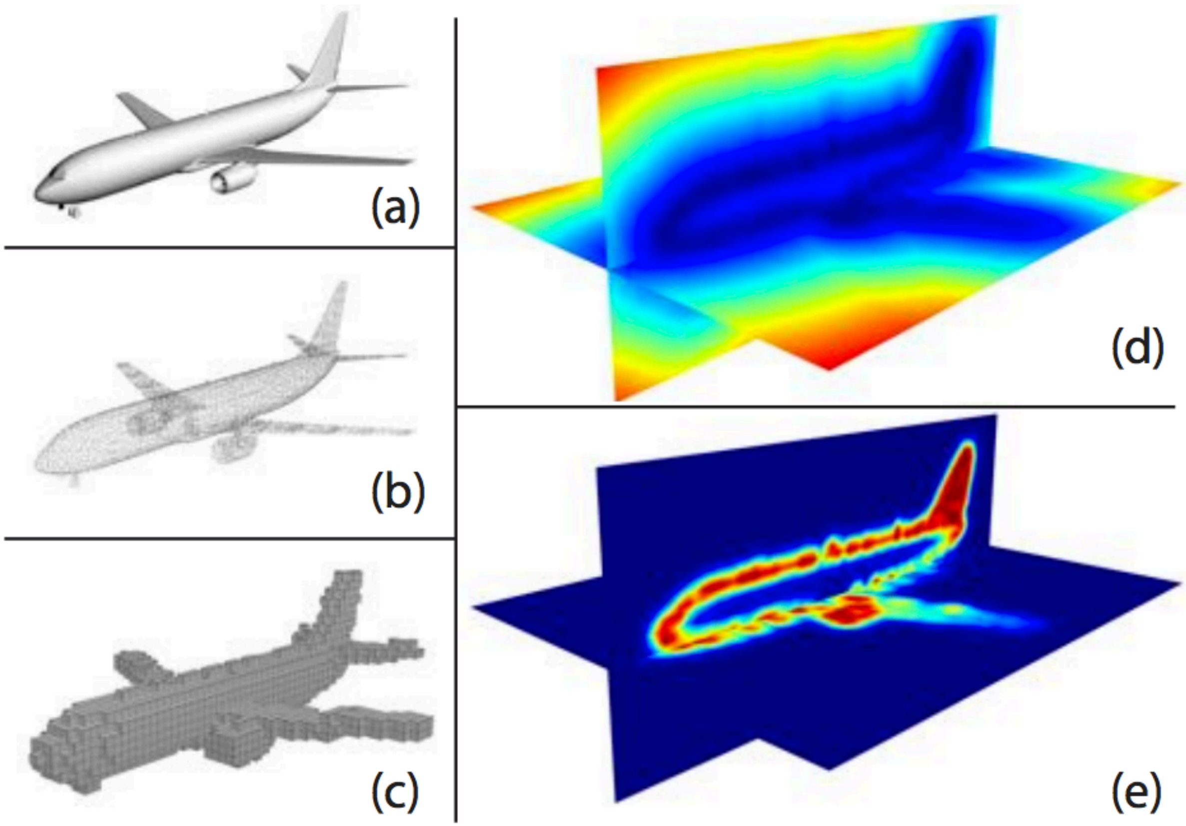
Geology

3D volumetric representations

Mesh

Point cloud

Volumetric occupancy

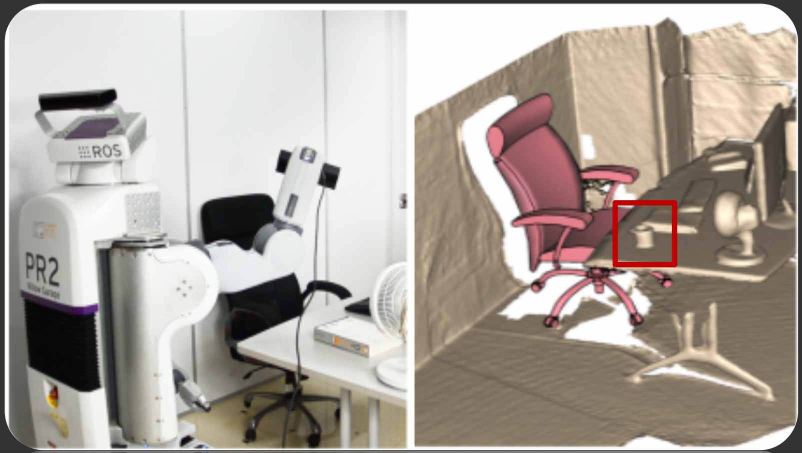


Volumetric distance field

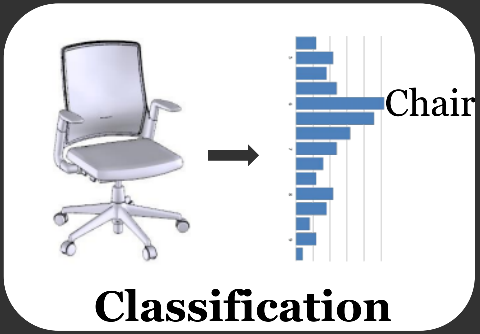
Volumetric Gaussian distance field

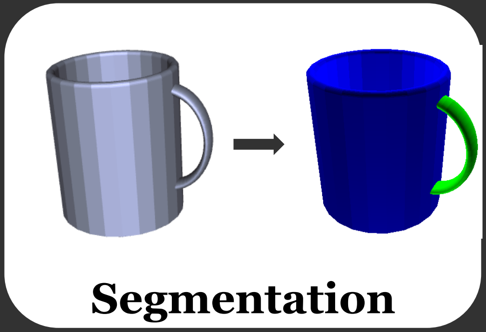
Shape Analysis

3D Shape Analysis



Robotics





CNN for 3D Shape Analysis

Goal

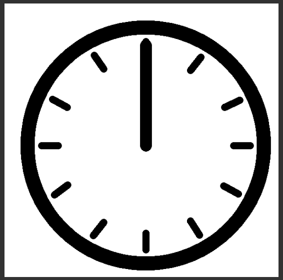
- General
- Efficient
- Effective

3D data

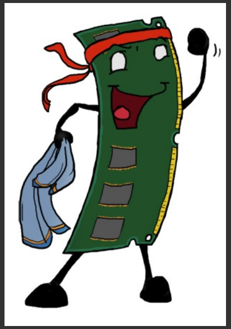
CNN
structure

Goal

- General
- Efficient
- Effective



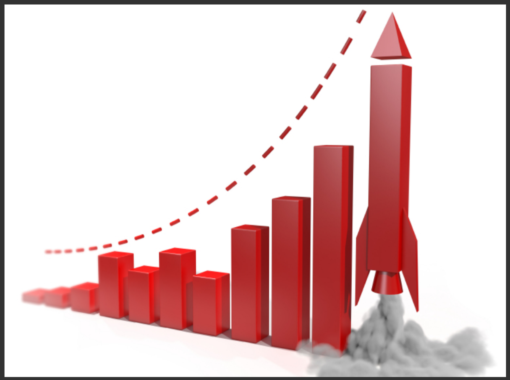
Time cost



Memory cost

Goal

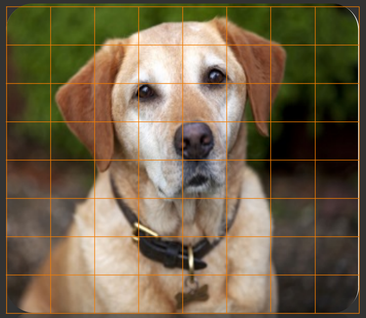
- General
- Efficient
- Effective



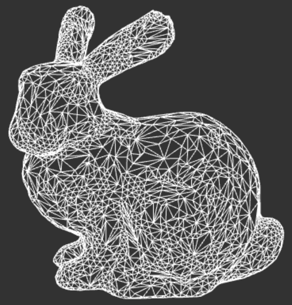
Performanc
e

Key Challenge

- A 3D shape representation for efficient CNN on GPU
  - 2D Regular grid
  - Irregular 3D shape

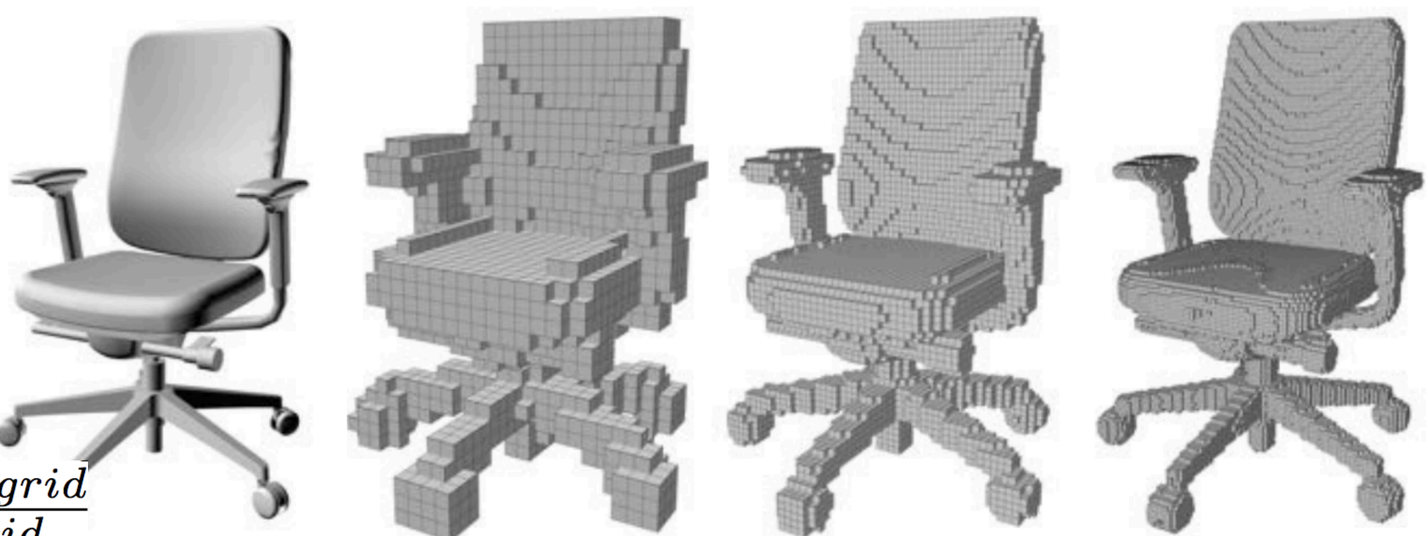


Image



Mesh

The sparsity characteristic of 3D data

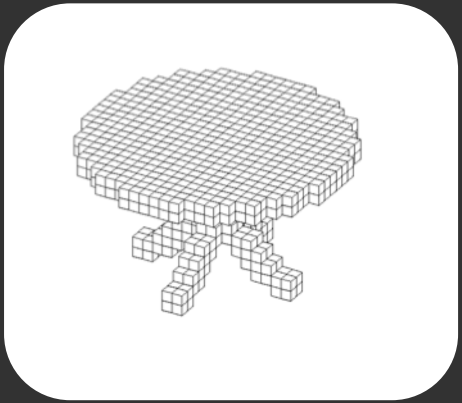


$$\frac{\#occupied\ grid}{\#total\ grid}$$

Occupancy:	10.41%	5.09%	2.41%
Resolution:	32	64	128

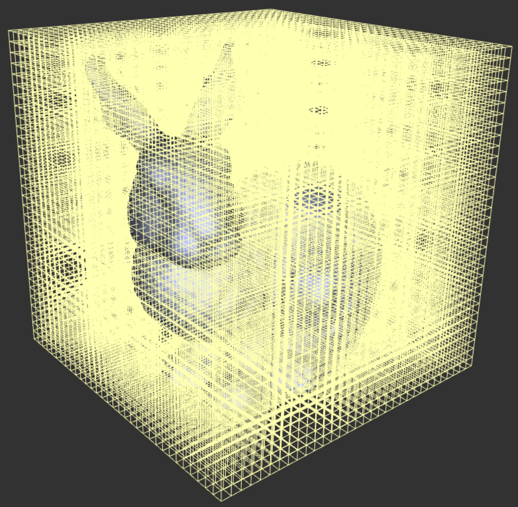
Full Voxel based Solutions

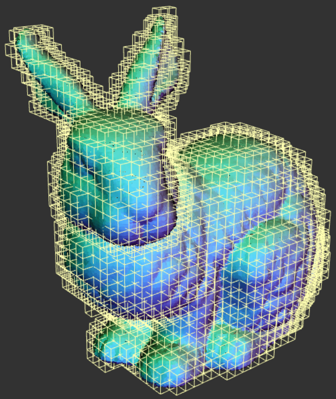
- Related work: [Wu et al. 2015], [Maturana and Scherer 2015], …
  - General: intuitive extension of images
  - Efficient: $O(N^3)$



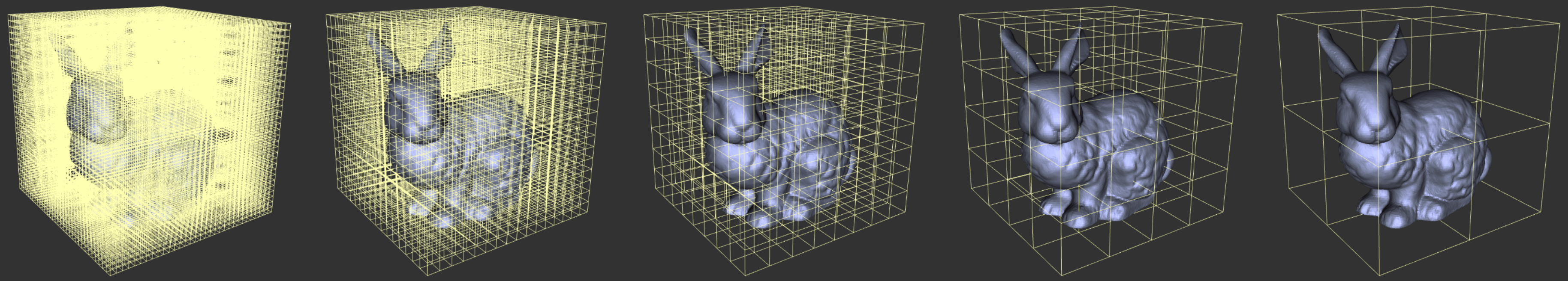
Key Idea

- Store the sparse surface signals
- Constrain the computation near the surface

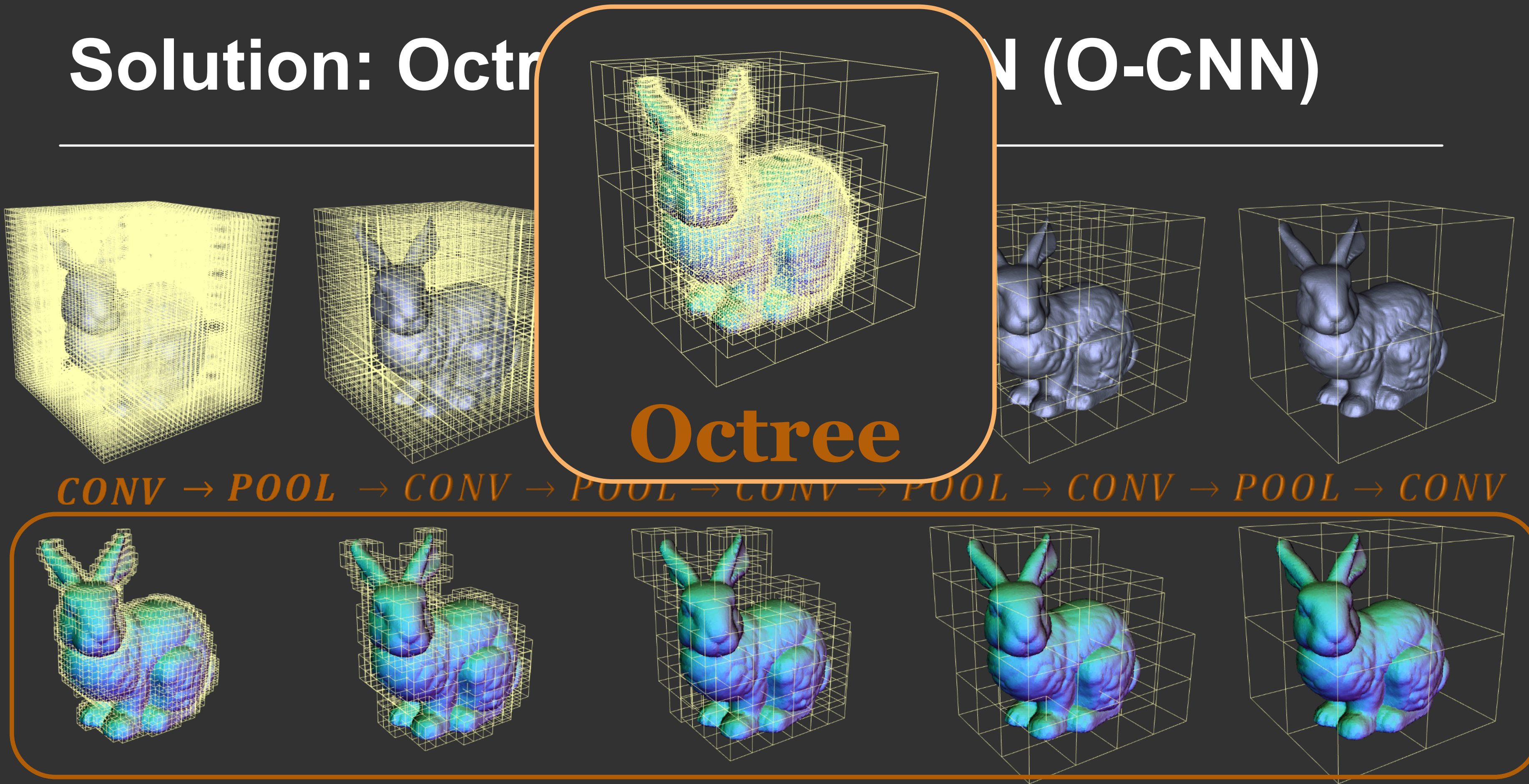




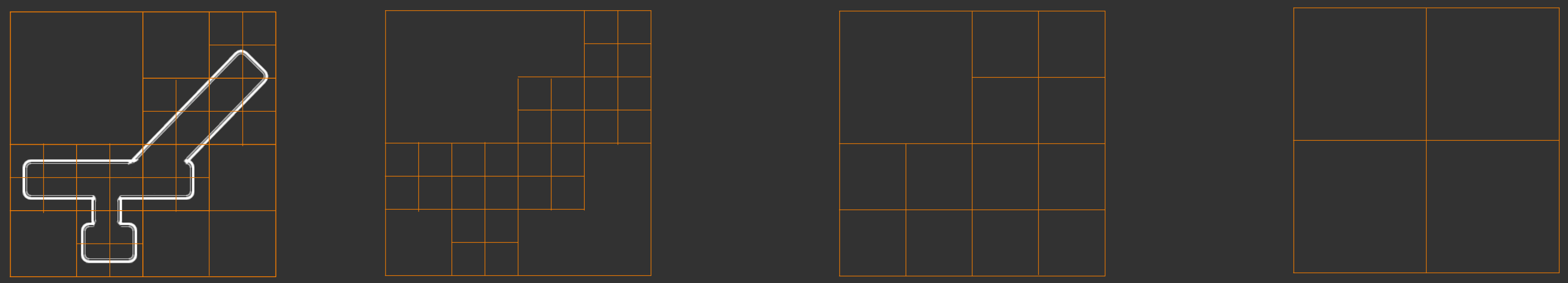
Solution: Octree based CNN (O-CNN)



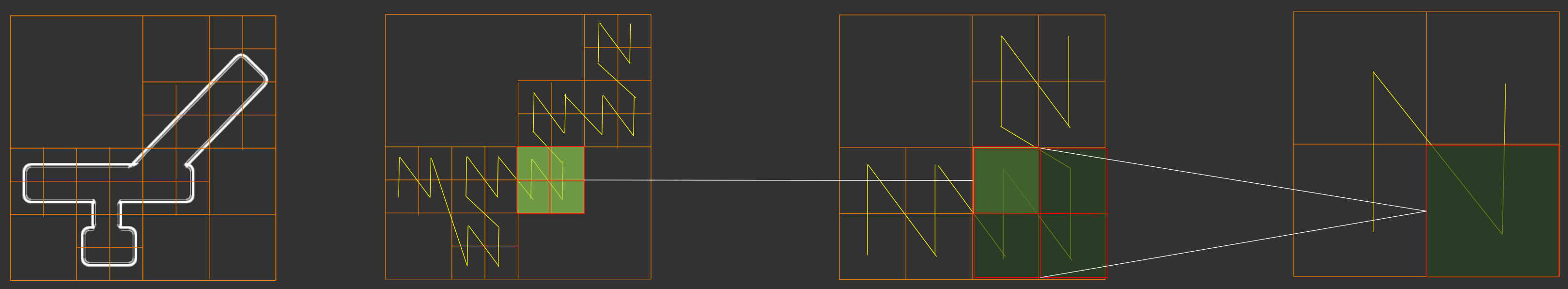
Solution: Octr N (O-CNN)



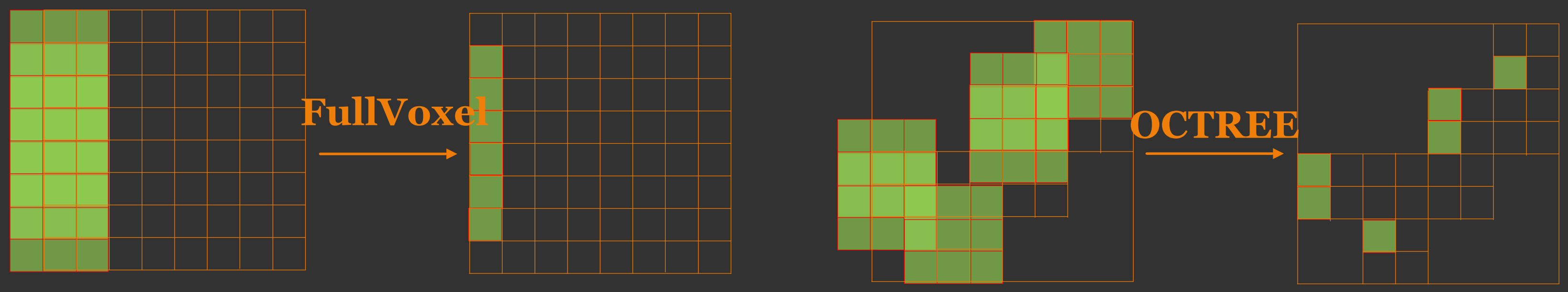
Octree Data Structure



Octree Data Structure

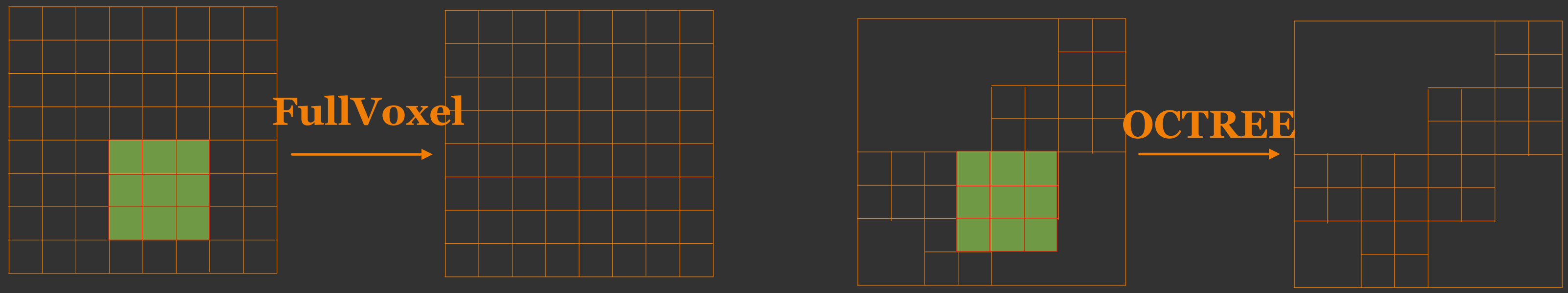


Convolution on Octree

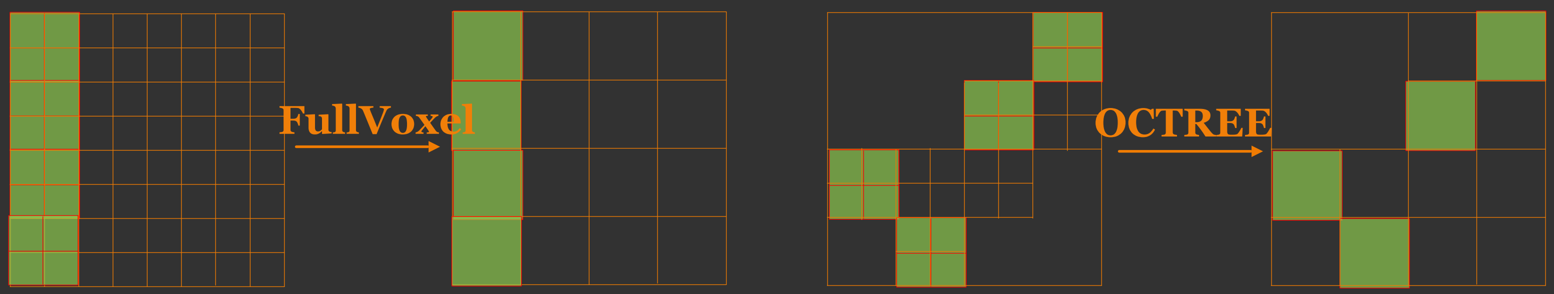


Convolution on Octree

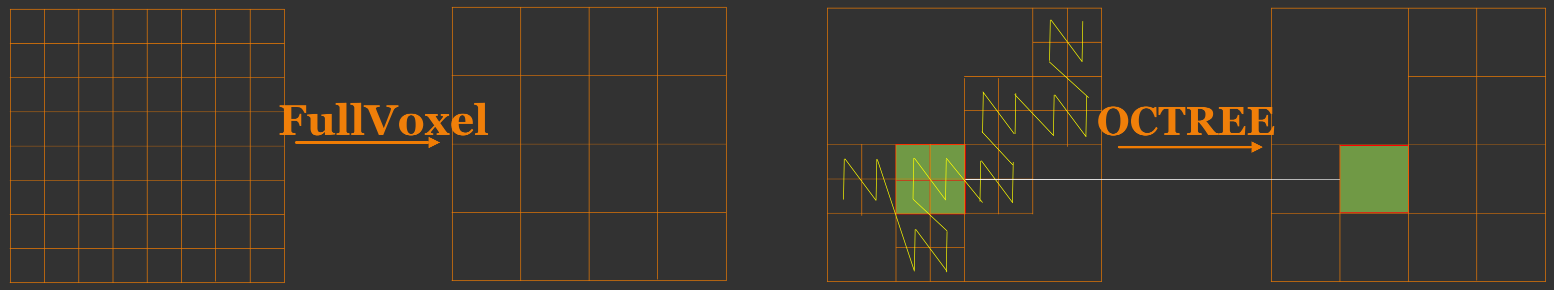
- Neighborhood searching: Hash table



Pooling on Octree



Pooling on Octree



Other CNN Operations on Octree

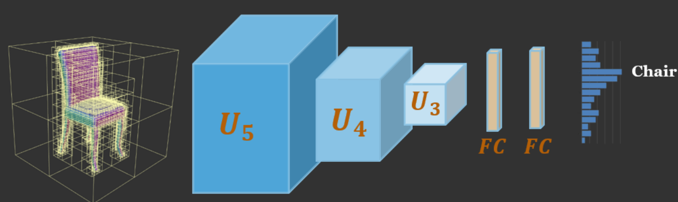
- Convolution with stride > 1

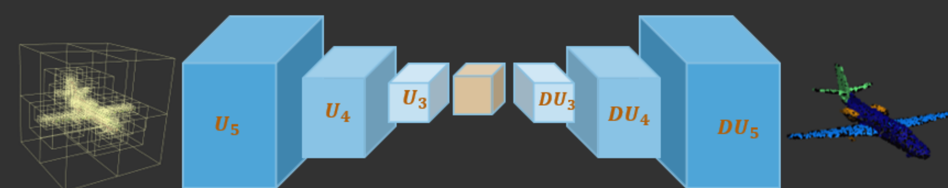
- Deconvolution and un-pooling
  - Inverse operations of convolution and pooling

- Support most CNN architectures for images
  - LeNet [Lecun et al. 1998], GoogLeNet [Szegedy et al. 2015], ResNet [He et al. 2016], DeconvNet [Noh et al . 2015], FCN [Long et al. 2015] …

O-CNN for Shape Analysis

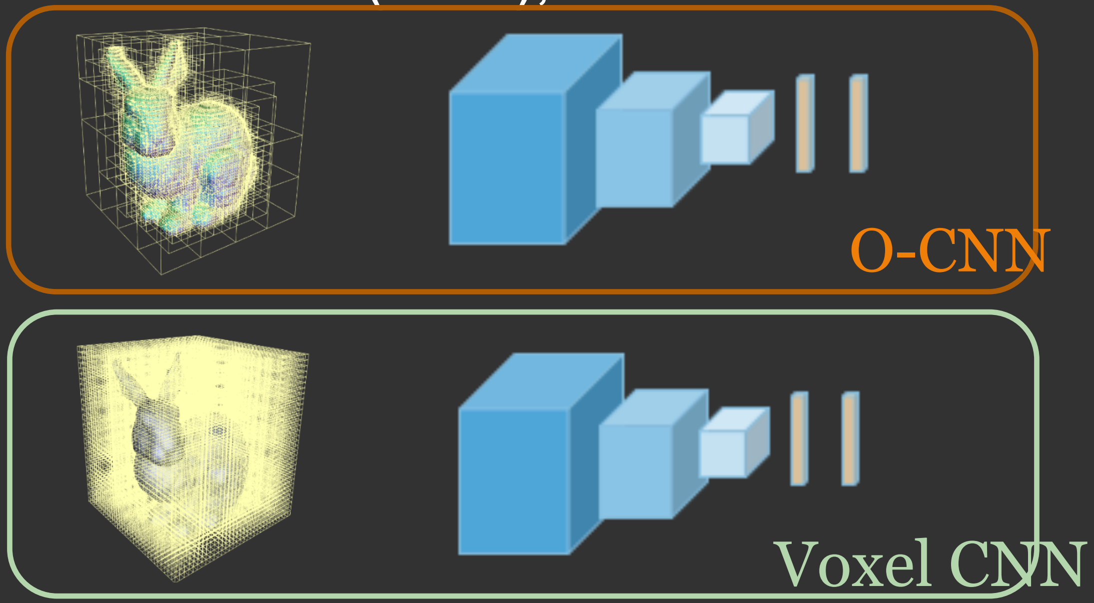
- Shape classification and retrieval
  - LeNet [Lecun et al. 1998]
- Shape segmentation
  - DeconvNet [Noh et al . 2015] + DenseCRF [Krähenbühl and Koltun2011]



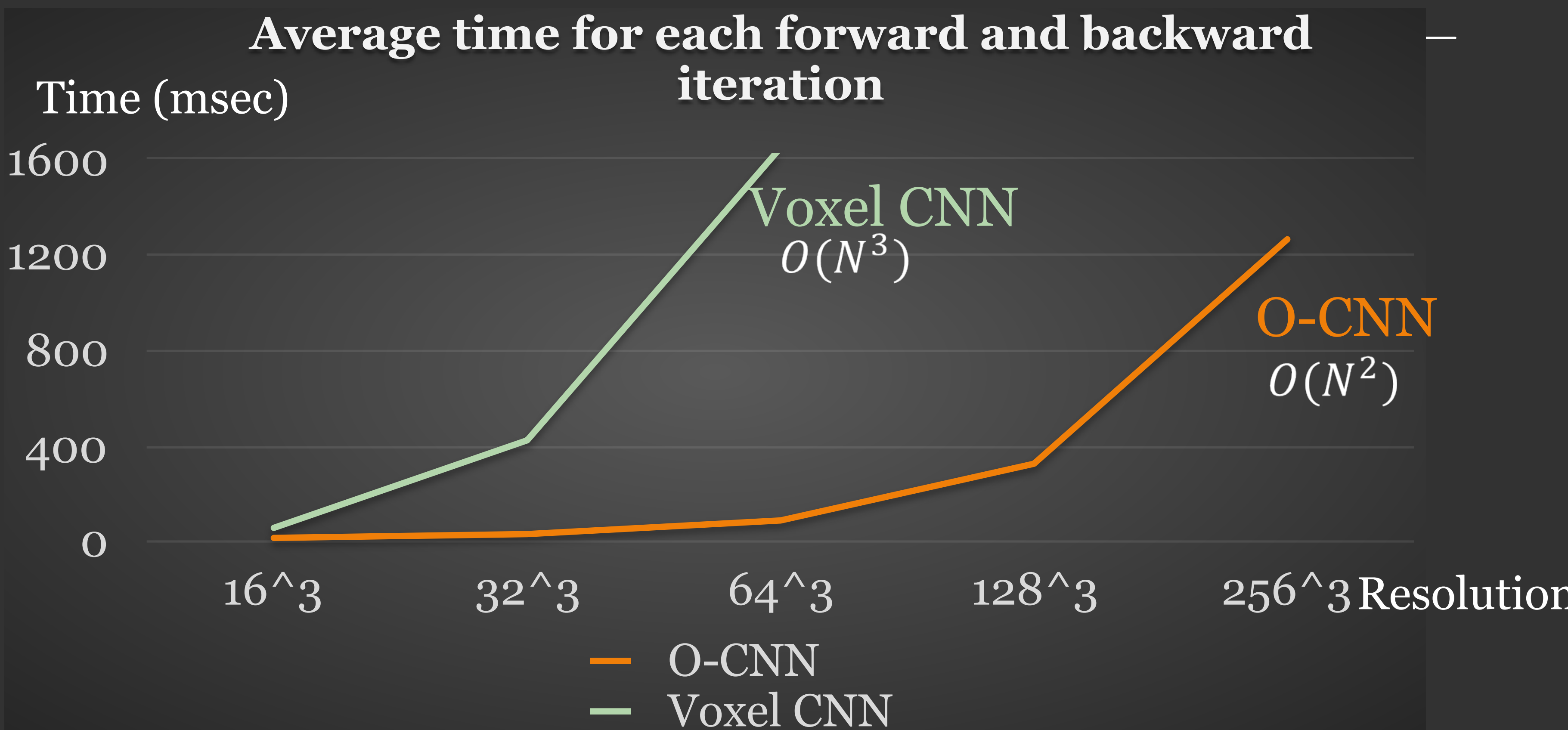


Efficiency of O-CNN

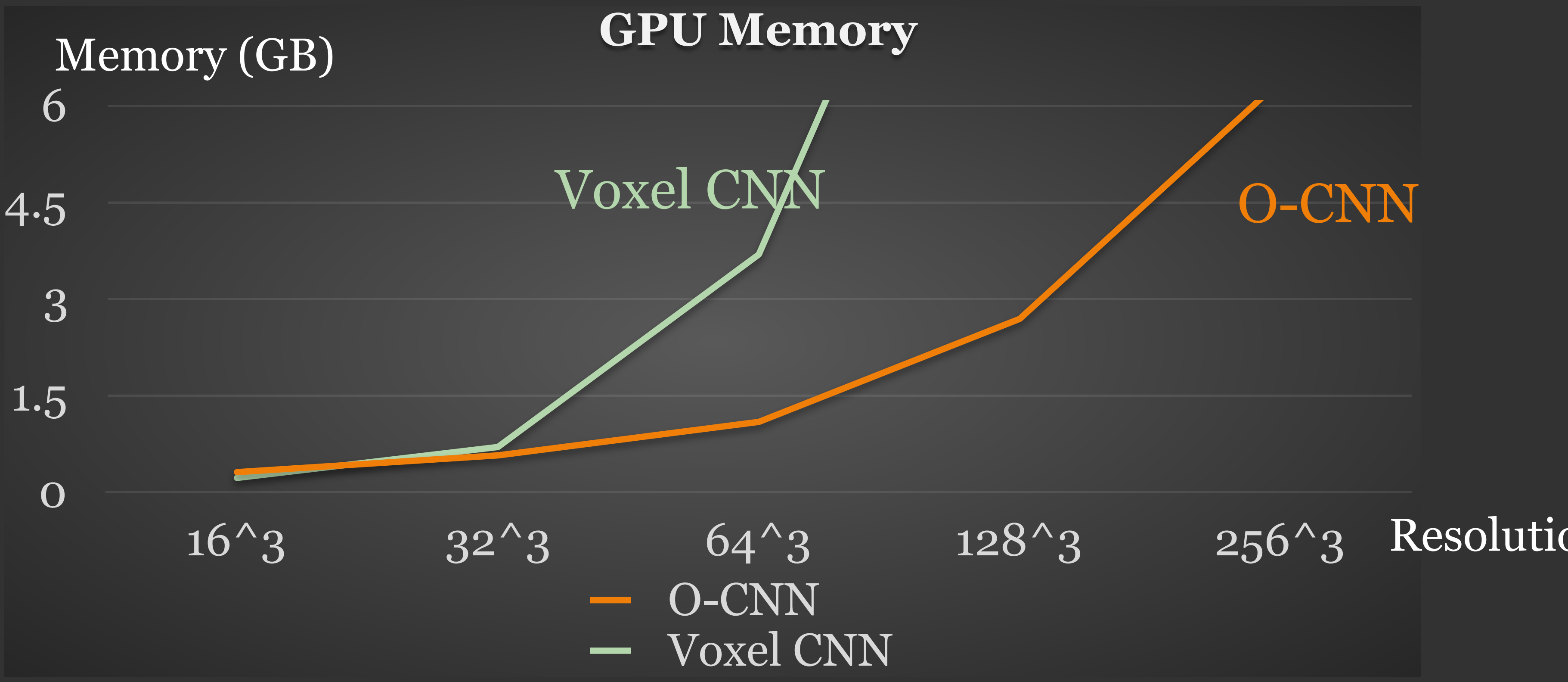
- O-CNN vs. full voxel CNN
  - Geforce 1080 GPU (8GB); Batch size 32



Computational Efficiency



Memory Efficiency



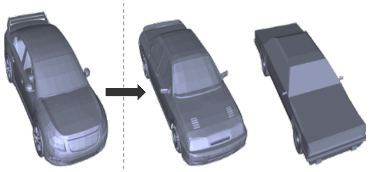
Results – Classification

- Task: recognize the shape category
- Dataset: Princeton ModelNet40,
  12311 3D models, 40 categories
- Evaluation metric: classification accuracy

Network	without voting
VoxNet (32^3)	82.0%
Geometry image	83.9%
SubVolSup (32^3)	87.2%
FPNN (64^3)	87.5%
FPNN+normal(64^3)	88.4%
PointNet	89.2%
VRN (32^3)	89.0%
O-CNN(3) 8^3	85.5%
O-CNN(4)	88.3%
O-CNN(5)	89.6%
O-CNN(6)	**89.9%**
O-CNN(7)	89.5%
O-CNN(8) 256^3	89.6%

Results – Shape Retrieval

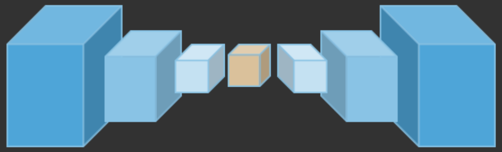
- Task: Given a query shape, retrieve similar shapes from the database
- Dataset: ShapeNet55 Core, 51190 3D models, 55 categories
- Evalution metric: precision, recall, mAP, F-score, and NDCG



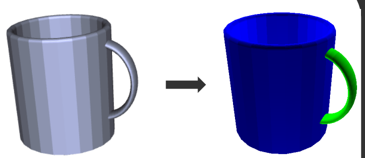
Retrieval

Method	P@N	R@N	F1@N	mAP	NDCG@N
Tatsuma_DB	0.427	0.689	0.472	0.728	0.875
Wang_CCMLT	0.718	0.350	0.391	0.823	0.886
Li_ViewAggr	0.508	**0.868**	0.582	0.829	0.904
Bai_GIFT	0.706	0.695	0.689	0.825	0.896
Su_MVCNN	0.770	0.770	0.764	0.873	0.899
O-CNN(5)	0.768	0.769	0.763	0.871	0.904
O-CNN(6)	**0.778**	0.782	**0.776**	**0.875**	**0.905**

Results – Segmentation



- Task: Segment a 3D shape into semantic parts
- Dataset: dataset from [Yi et al. 2016], 16881 models, 2~6 parts
- Evaluation metric: Intersection over Union



	mean	plane	bag	cap	car	chair	e.ph.	guitar	knife
# shapes		2690	76	55	898	3758	69	787	392
[Yi et al. 2016]	81.4	81.0	78.4	77.7	75.7	87.6	61.9	92.0	85.4
PointNet [Qi et al. 2017]	83.7	83.4	78.7	82.5	74.9	89.6	73.0	91.5	85.9
SpecCNN [Yi et al. 2017]	84.7	81.6	81.7	81.9	75.2	90.2	74.9	93.0	86.1
O-CNN(5)	85.2	84.2	86.9	84.6	74.1	90.8	81.4	91.3	87.0
O-CNN(6)	**85.9**	**85.5**	**87.1**	**84.7**	**77.0**	**91.1**	**85.1**	**91.9**	**87.4**

Conclusion

- Key idea
  - Store sparse surface signal
  - Constrain the computation near surface

- Octree based 3D CNNs
  - General, efficient, and effective

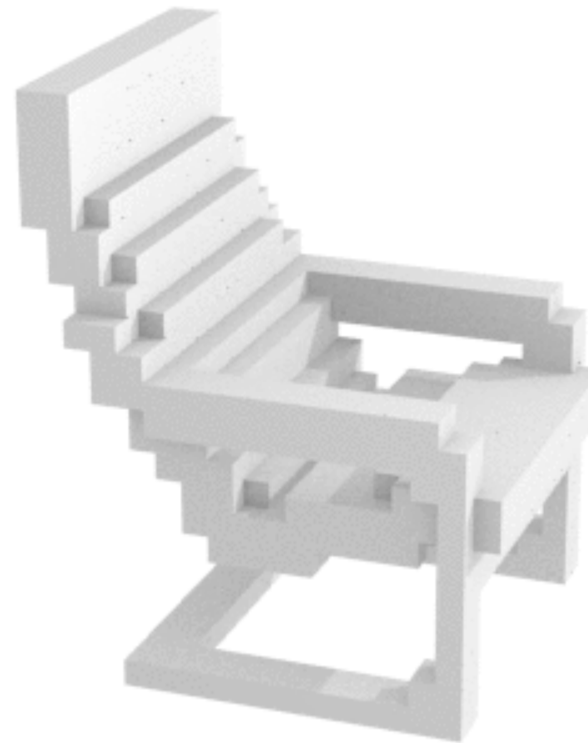
Code and data online

http://wang-ps.github.io/O-CNN

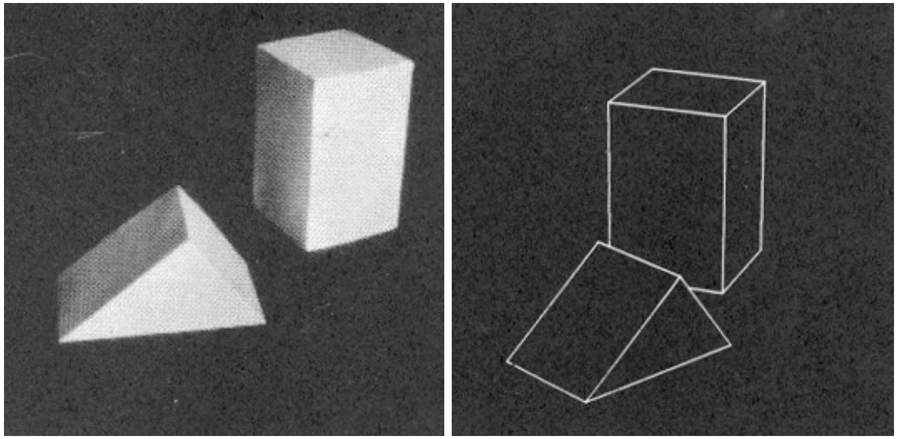
Shape Reconstruction

How do we learn to perceive 3D ?

How do we learn to perceive 3D ?



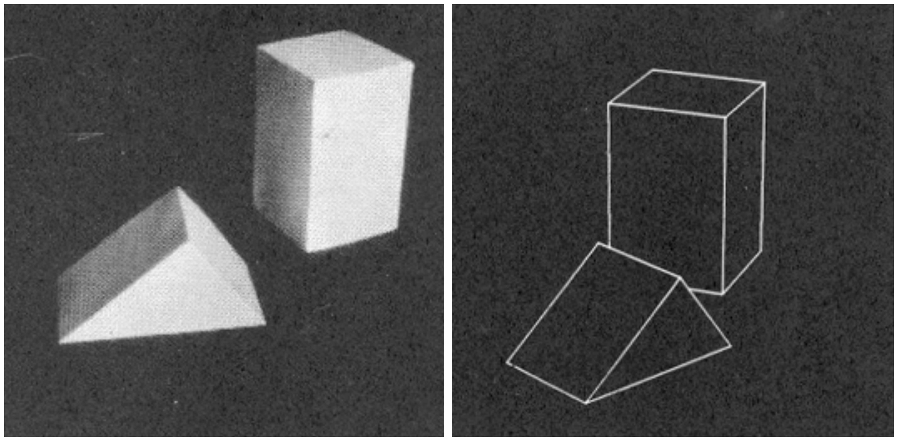
Single-view Reconstruction



Roberts. PhD Thesis, MIT. 1963

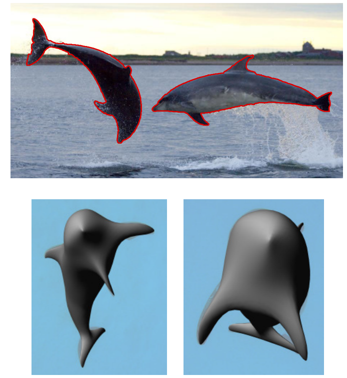
Unsupervised

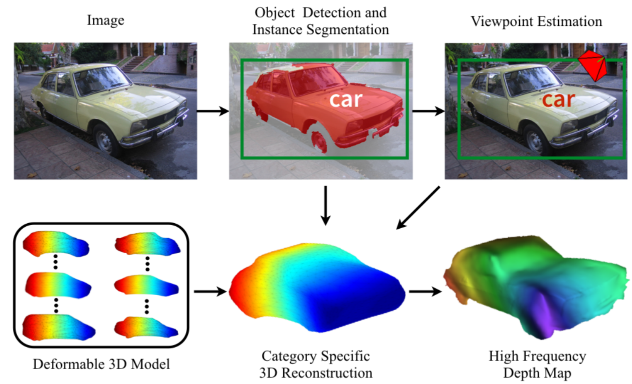
Single-view Reconstruction



Roberts. PhD Thesis, MIT. 1963

Unsupervised



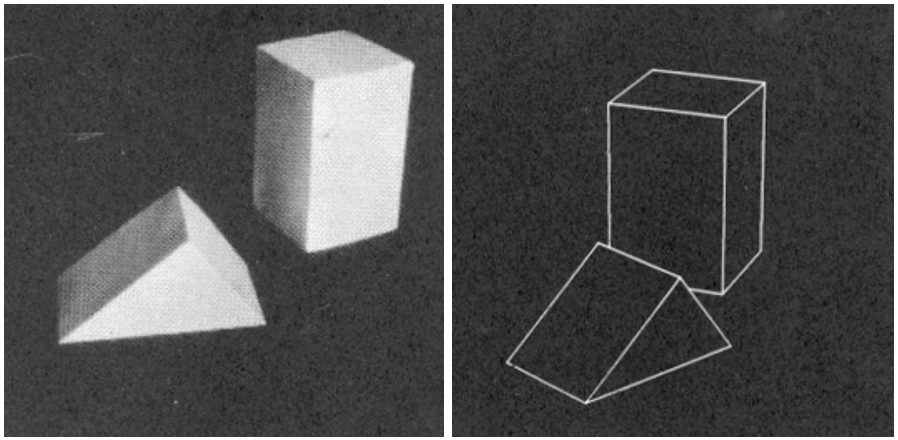


Cashman & Fitzgibbon, PAMI 2013

Kar et al., CVPR 2015

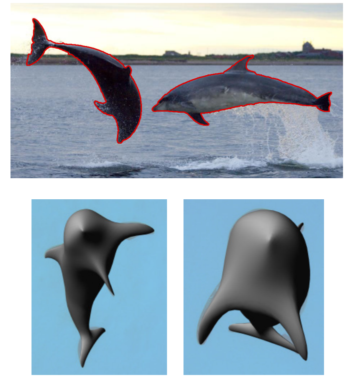
Supervision : Masks + Pose

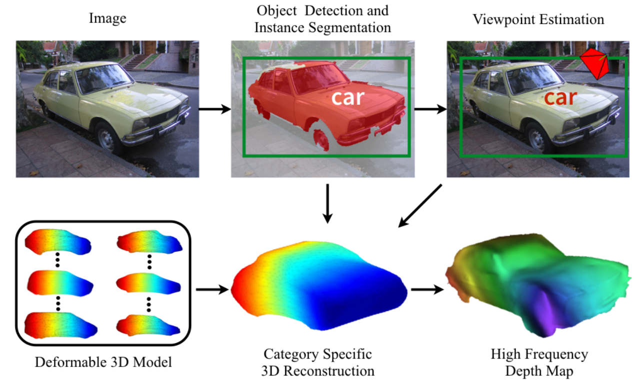
Single-view Reconstruction



Roberts. PhD Thesis, MIT. 1963

Unsupervised

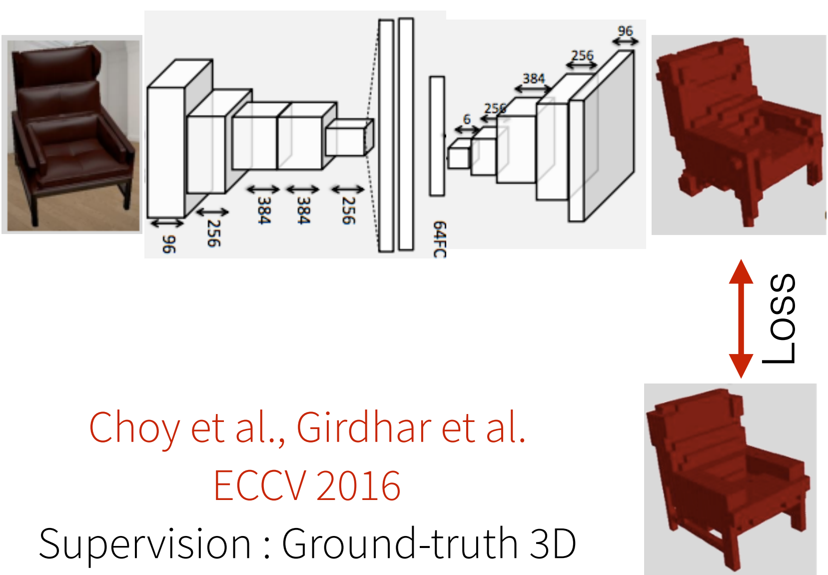




Cashman & Fitzgibbon, PAMI 2013

Kar et al., CVPR 2015

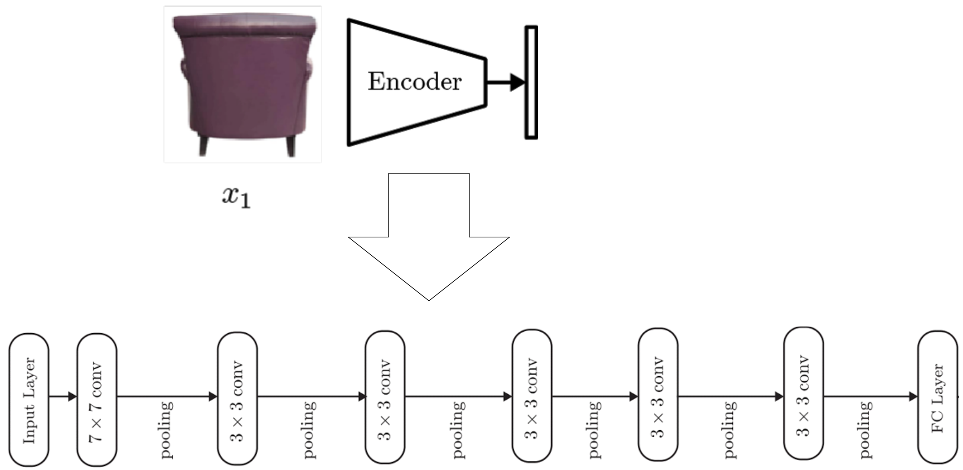
Supervision : Masks + Pose

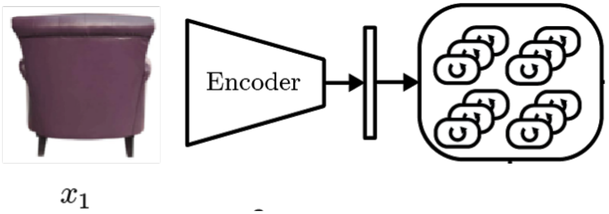


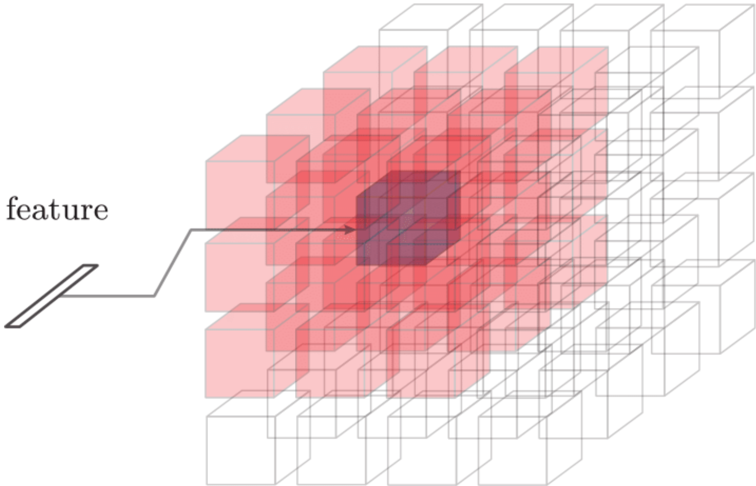
Choy et al., Girdhar et al.

ECCV 2016

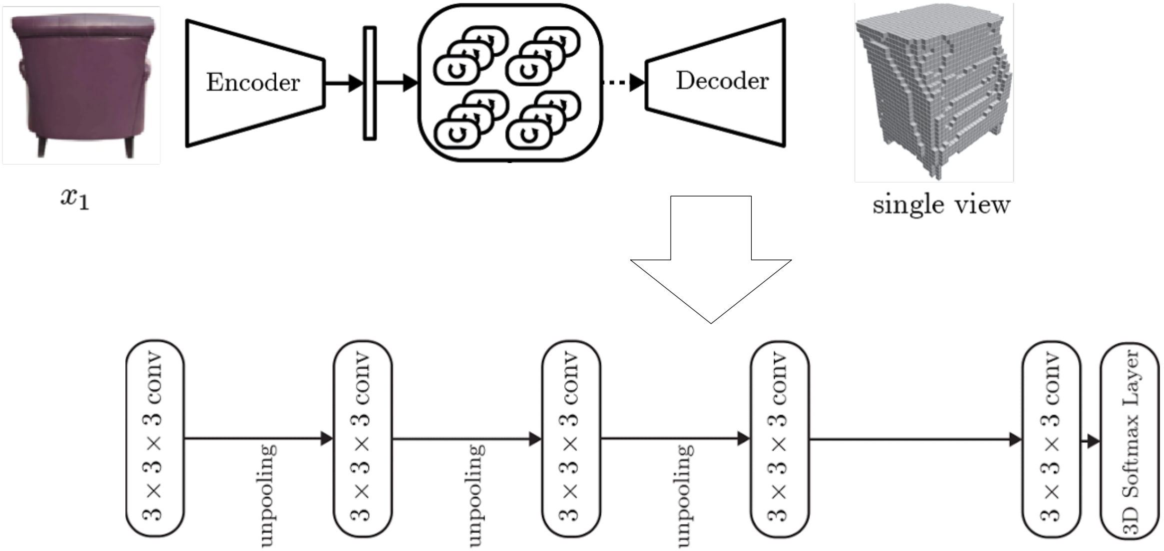
Supervision : Ground-truth 3D


Encoder
x_1
Input Layer
7 × 7 conv
pooling
3 × 3 conv
pooling
3 × 3 conv
pooling
3 × 3 conv
pooling
3 × 3 conv
pooling
3 × 3 conv
pooling
FC Layer

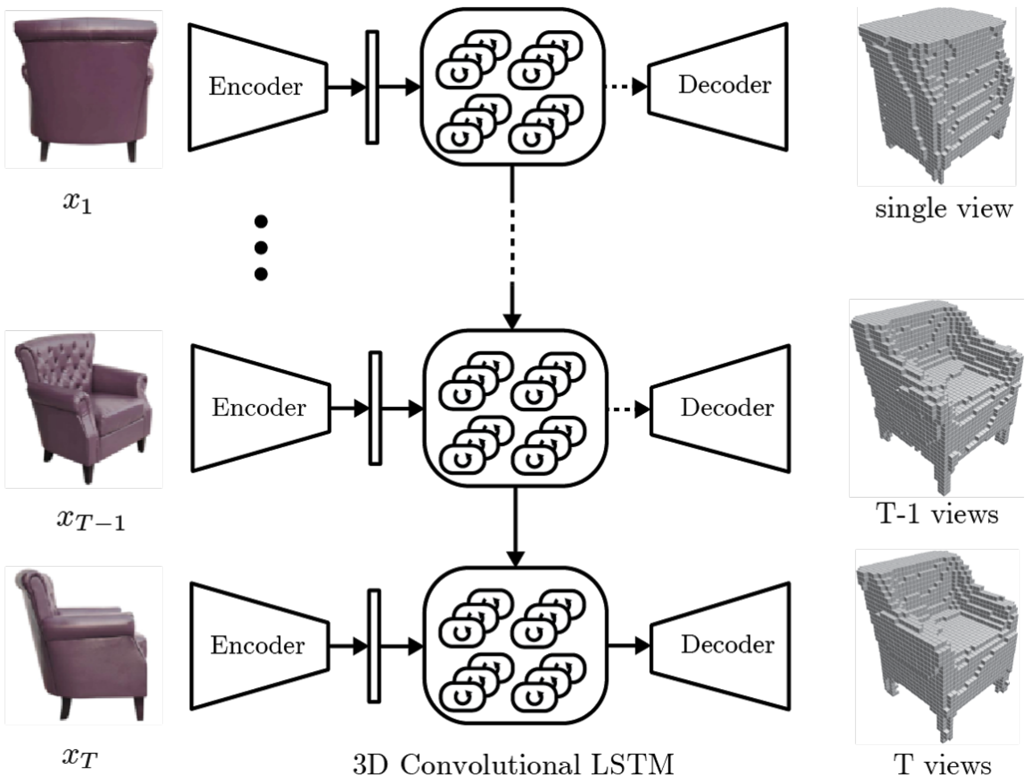




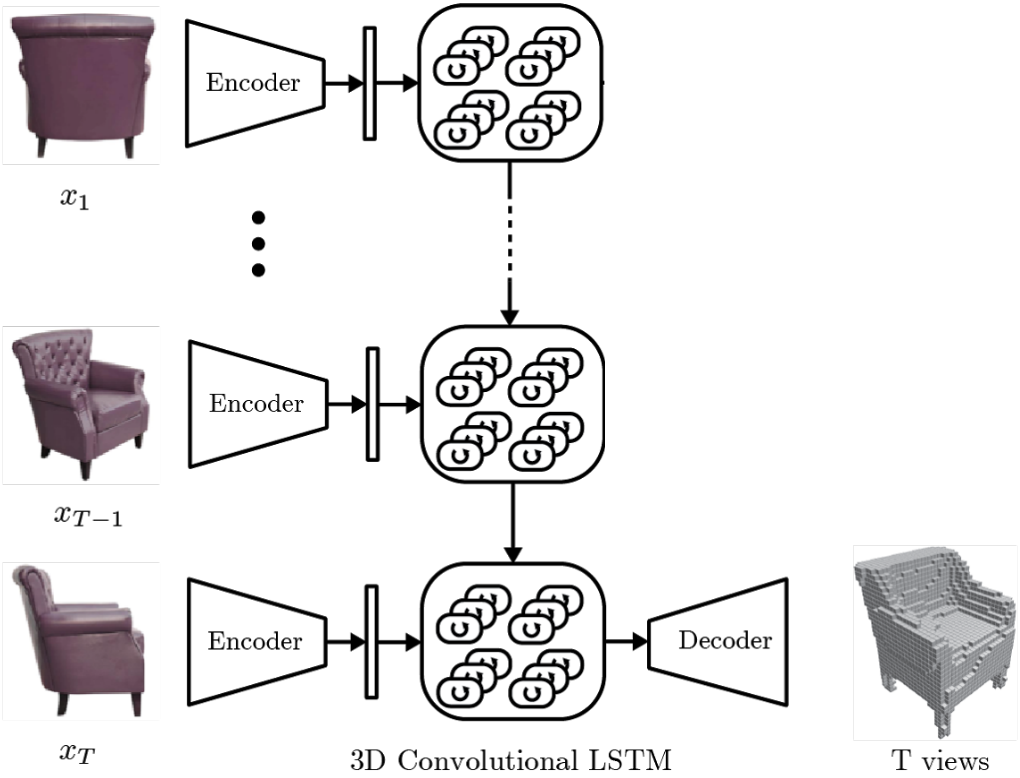
3D Convolutional LSTM



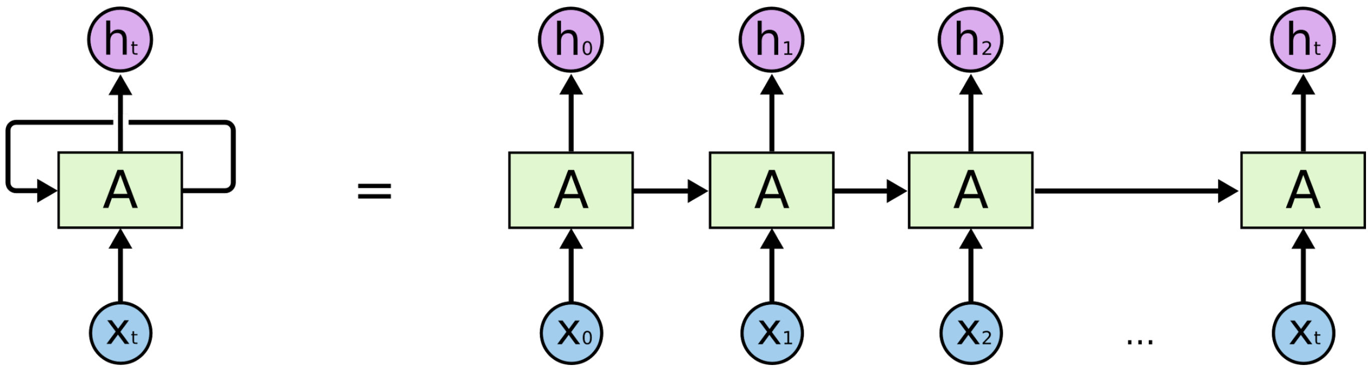
Encoder
Decoder
x_1
single view
$3 \times 3 \times 3$ conv
unpooling
$3 \times 3 \times 3$ conv
unpooling
$3 \times 3 \times 3$ conv
unpooling
$3 \times 3 \times 3$ conv
$3 \times 3 \times 3$ conv
3D Softmax Layer


Encoder
Decoder
x_1
single view
Encoder
Decoder
x_{T-1}
T-1 views
Encoder
Decoder
x_T
3D Convolutional LSTM
T views

It is possible to aggregate information from multiple views

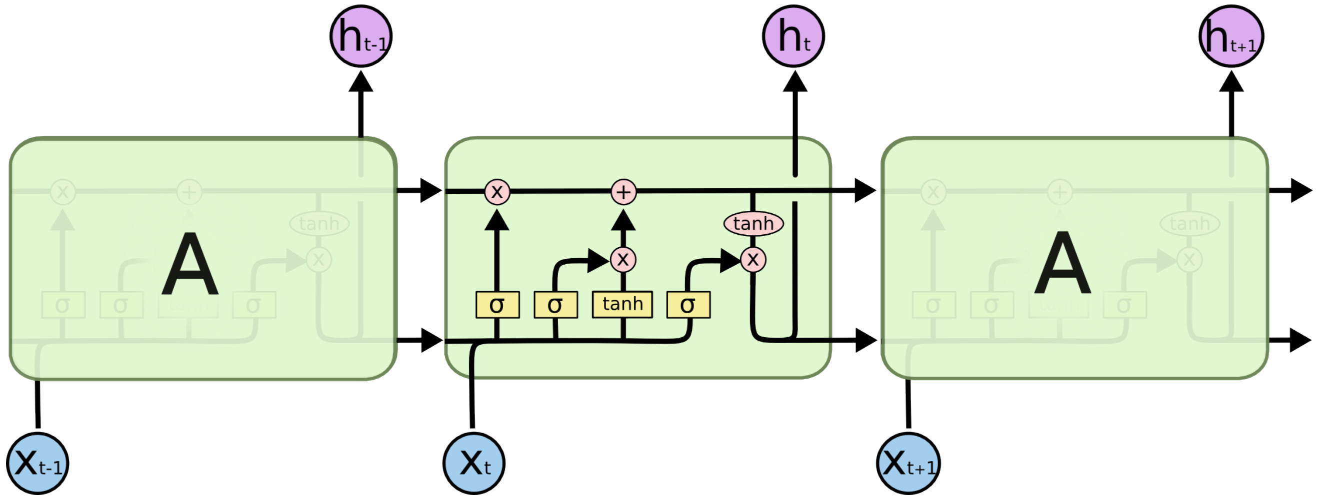


Recurrent Neural Network



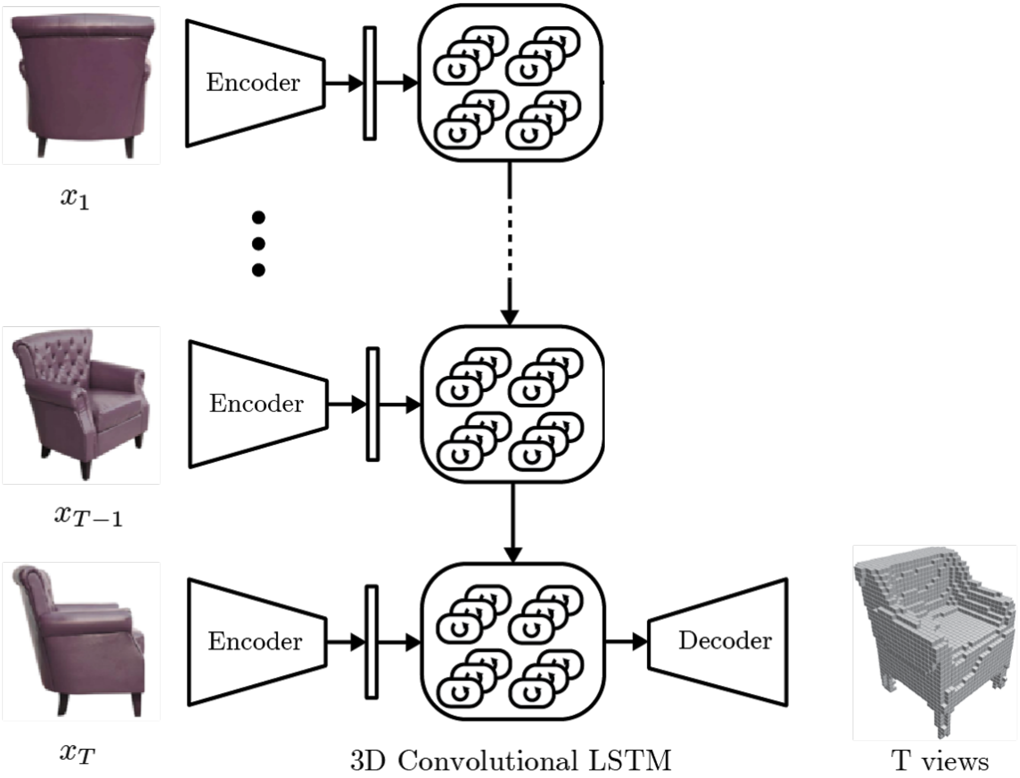
[Christopher Olah] Understanding LSTM Networks, http://colah.github.io/posts/2015-08-Understanding-LSTMs/

Long Short Term Memory



[Christopher Olah] Understanding LSTM Networks, http://colah.github.io/posts/2015-08-Understanding-LSTMs/

It is possible to aggregate information from multiple views

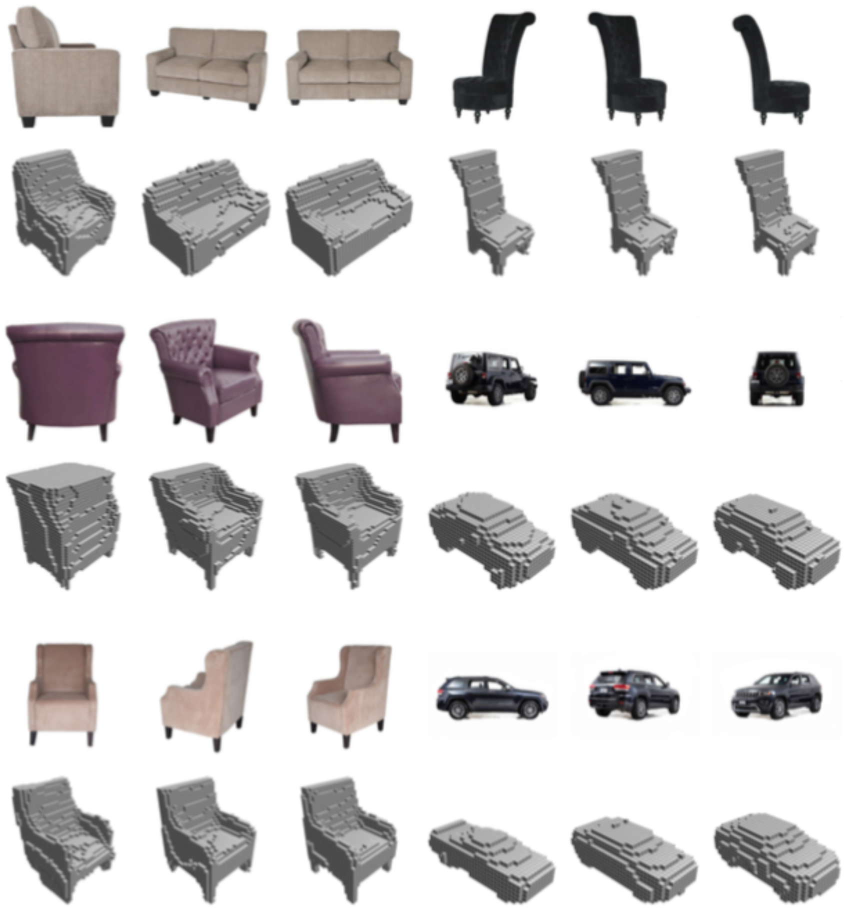


Training

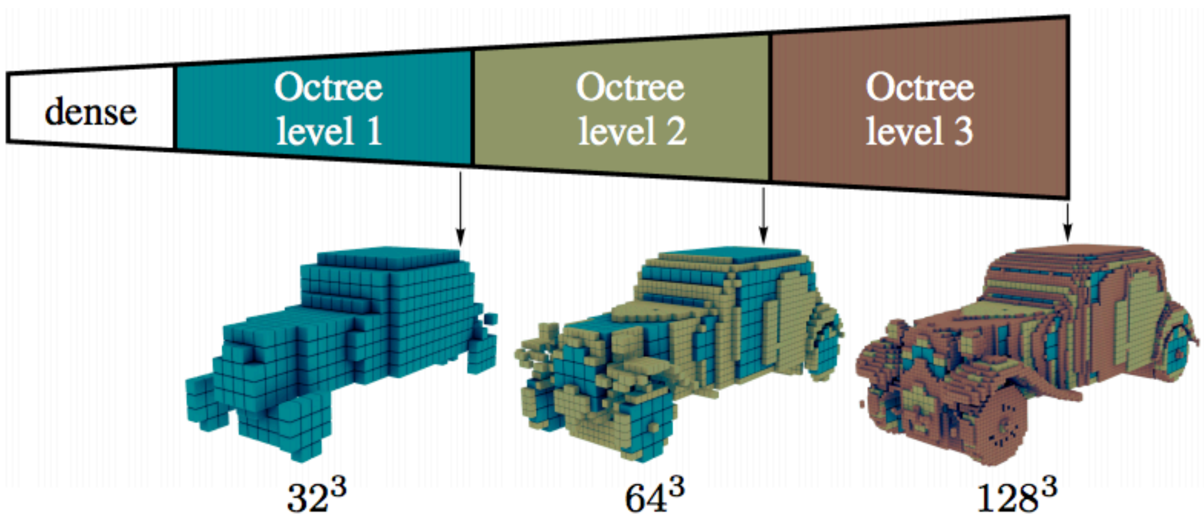
- ShapeNet
  - 50k CAD models
  - Render from arbitrary views
  - Random number of images w/ random order
  - Random background, translation

$$L(\mathcal{X}, y) = \sum_{i,j,k} y_{(i,j,k)} \log(p_{(i,j,k)}) + (1 - y_{(i,j,k)}) \log(1 - p_{(i,j,k)})$$

- Voxel-wise cross entropy loss

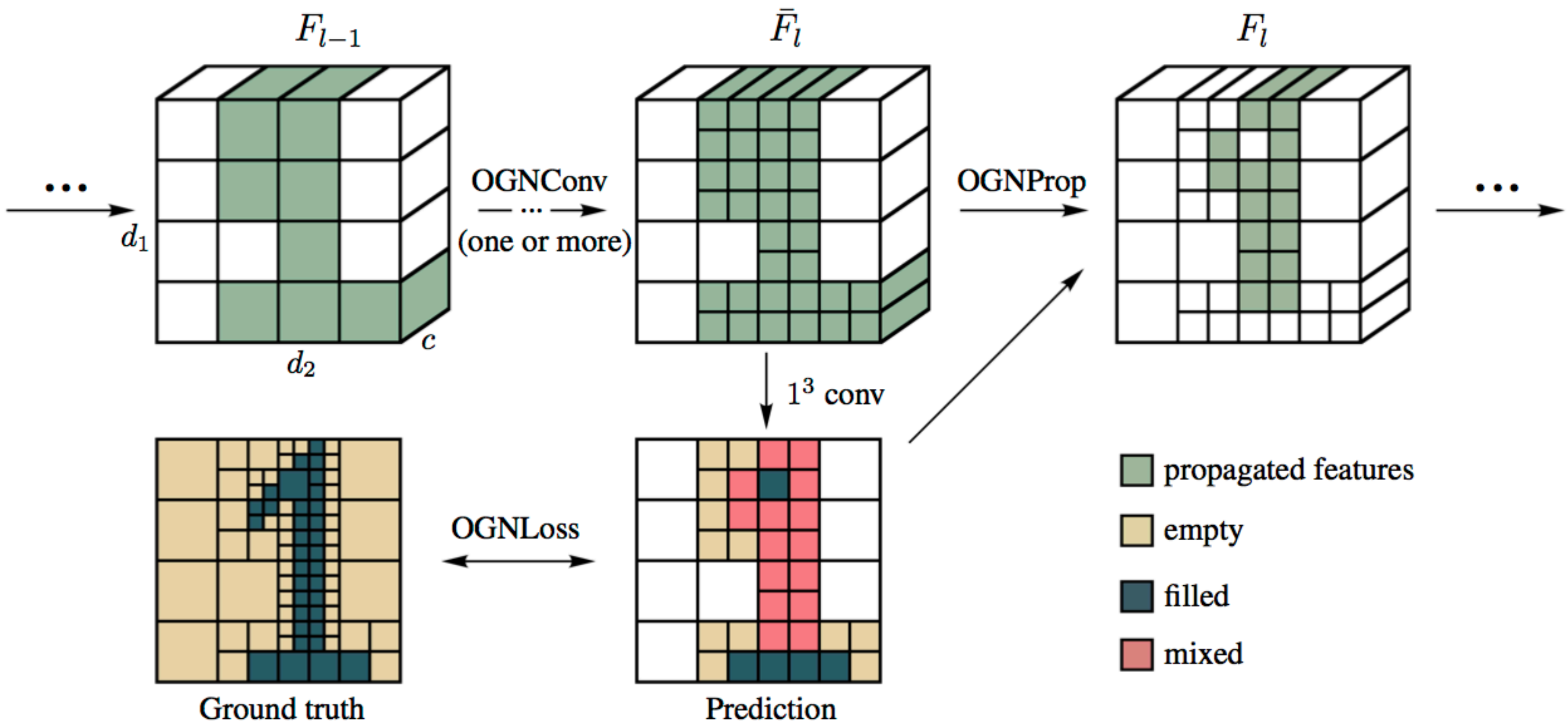


Towards higher spatial resolution

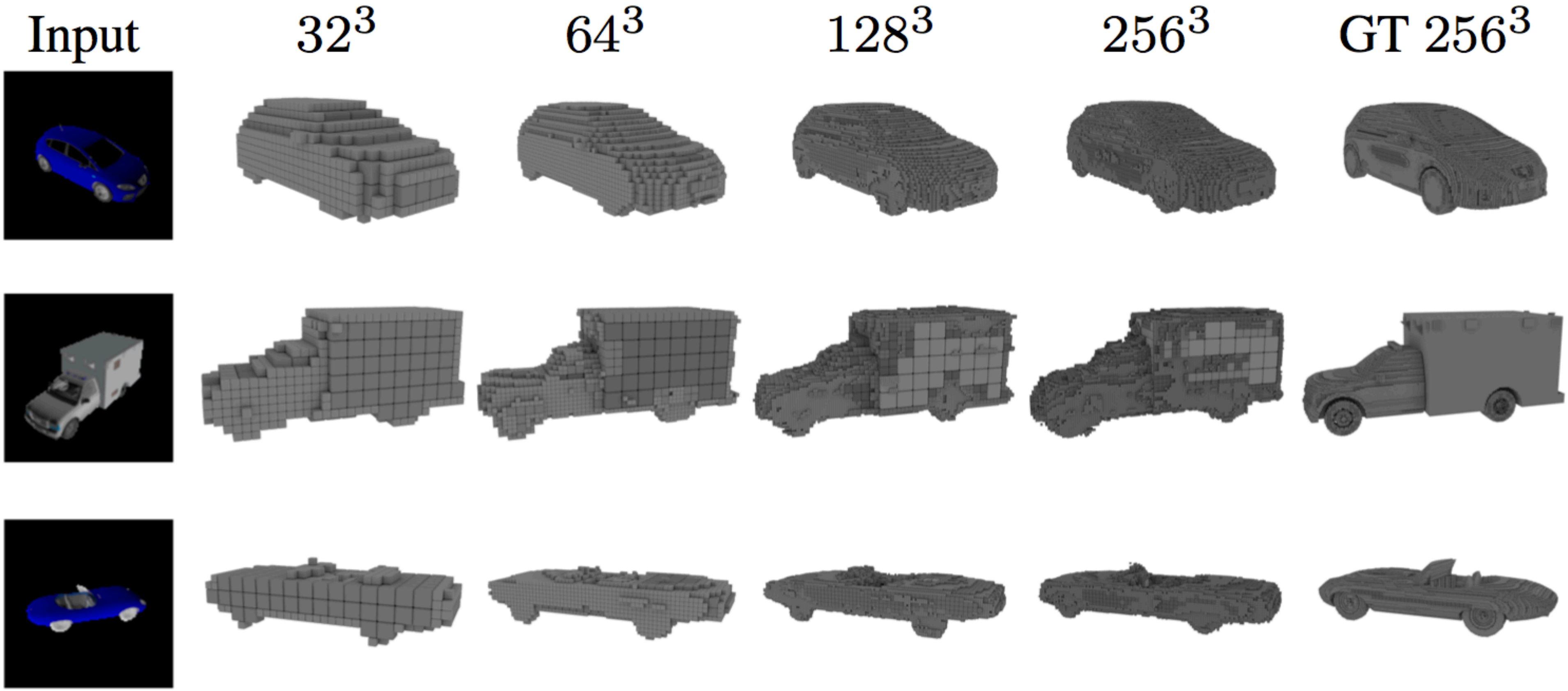


Maxim Tatarchenko, Alexey Dosovitskiy, Thomas Brox
"Octree Generating Networks: Efficient Convolutional Architectures for High-resolution 3D Outputs"
arxiv (March, 2017)

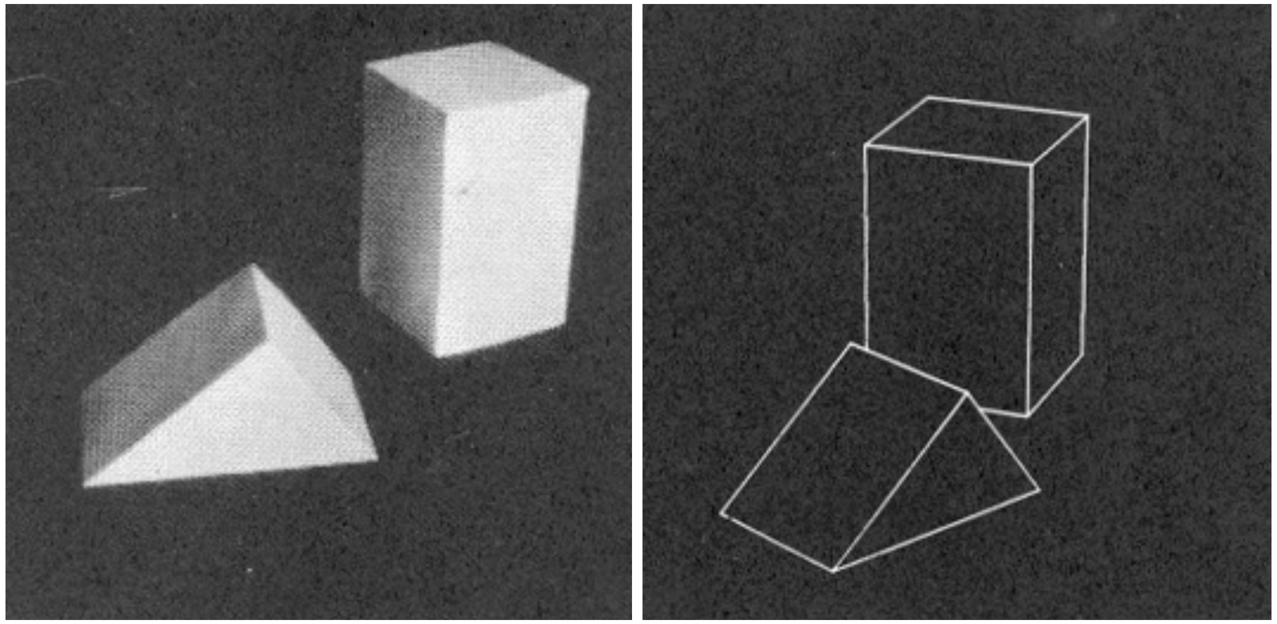
Progressive voxel refinement



Results

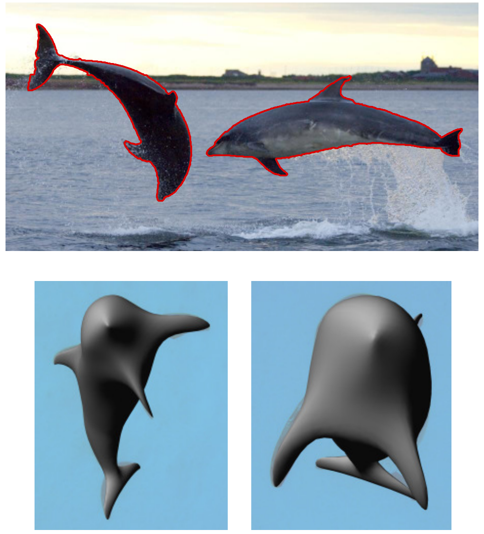


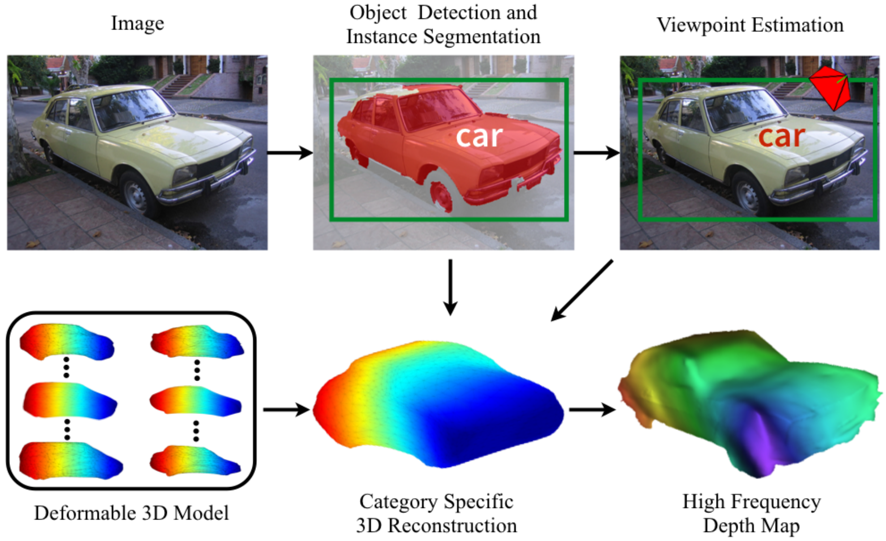
Single-view Reconstruction



Roberts. PhD Thesis, MIT. 1963

Unsupervised

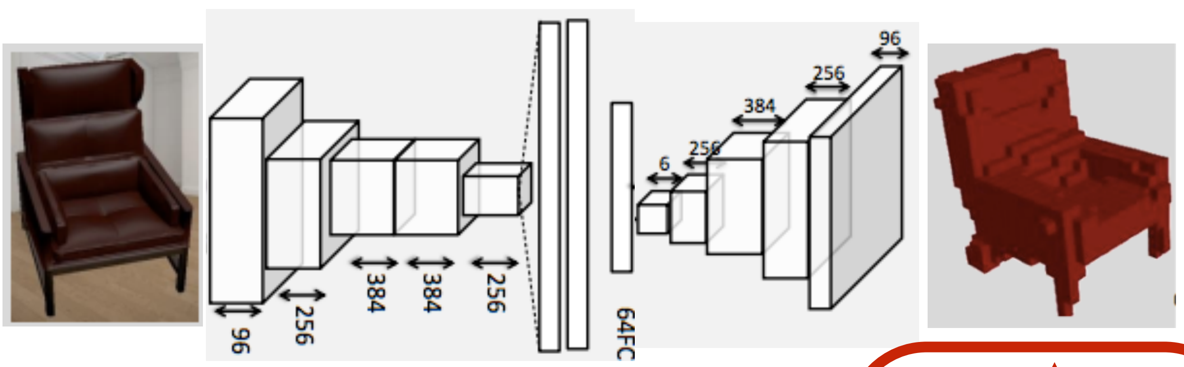


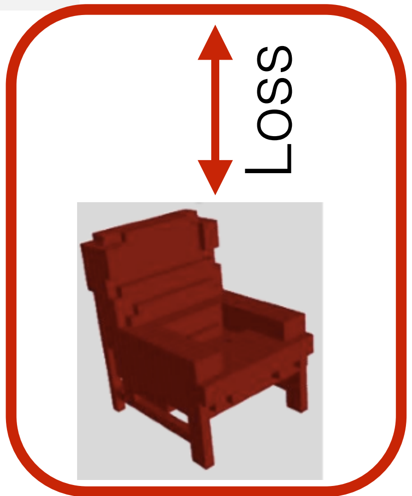


Cashman & Fitzgibbon, PAMI 2013

Kar et al., CVPR 2015

Supervision : Masks + Pose





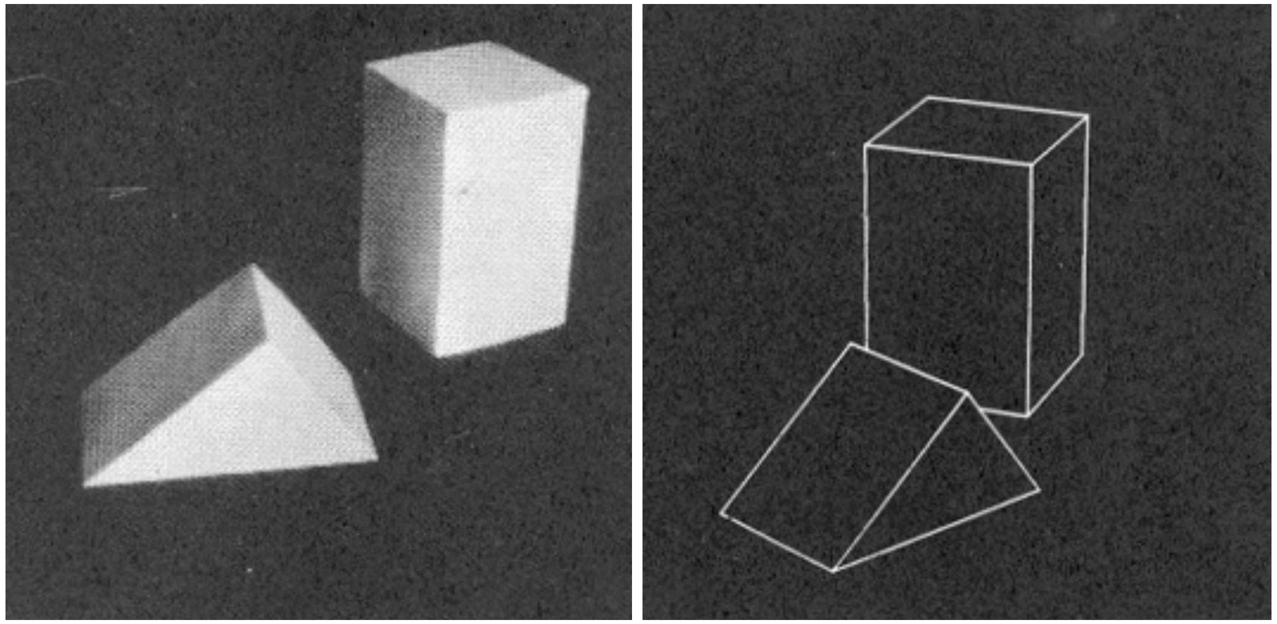
Choy et al., Girdhar et al.

ECCV 2016

Supervision : Ground-truth 3D

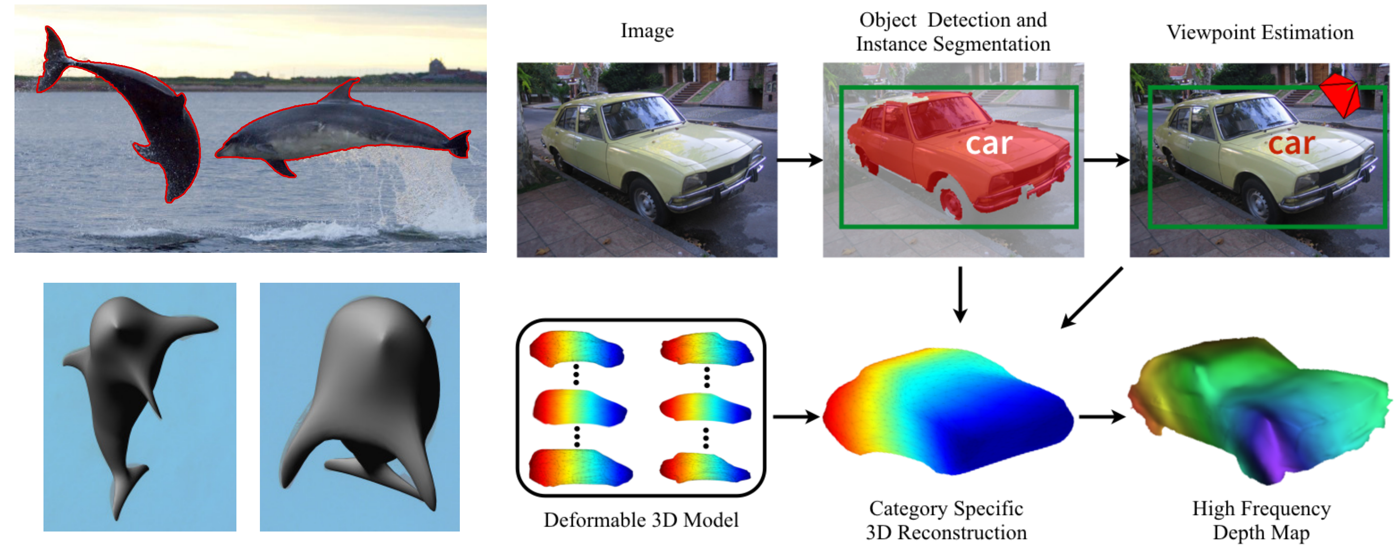
But we don’t have ground-truth 3D !

Single-view Reconstruction



Roberts. PhD Thesis, MIT. 1963

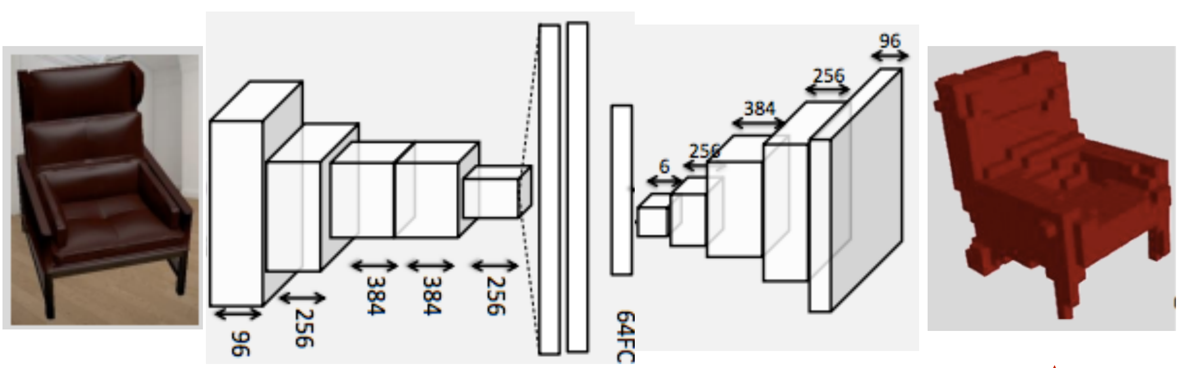
Unsupervised

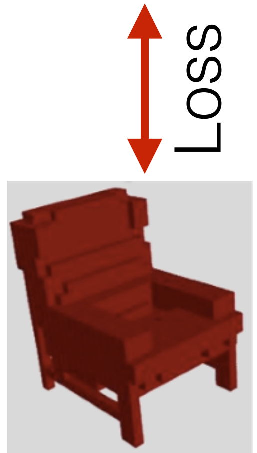


Cashman & Fitzgibbon, PAMI 2013

Kar et al., CVPR 2015

Supervision : Masks + Pose



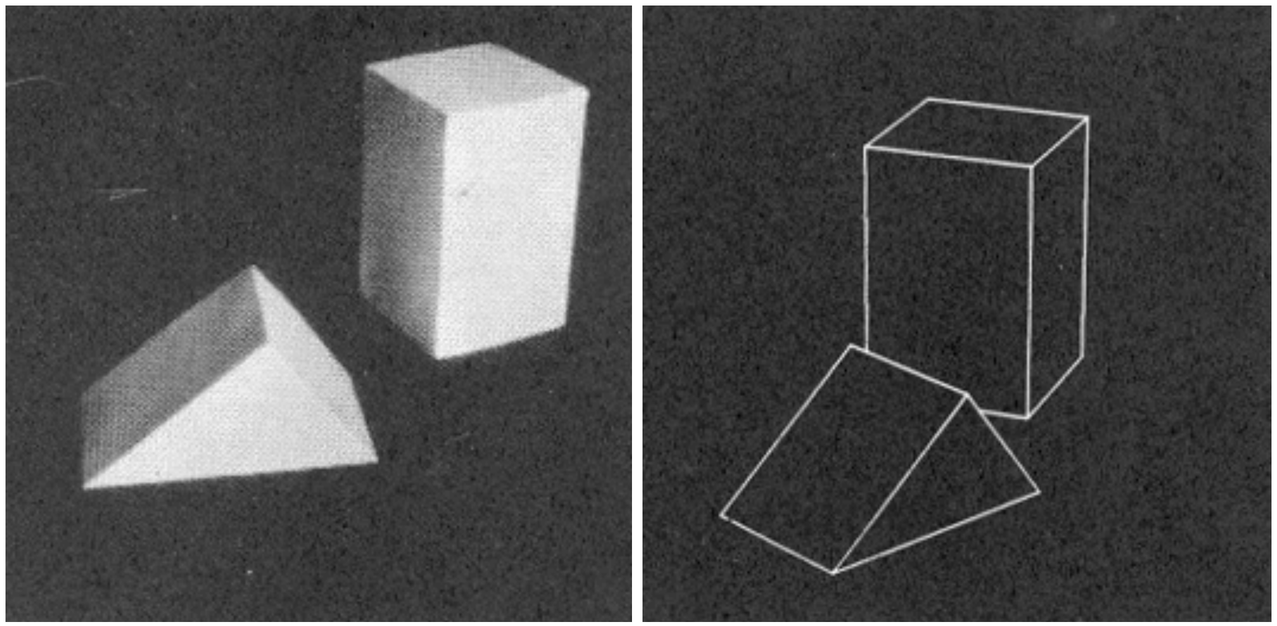


Choy et al., Girdhar et al.

ECCV 2016

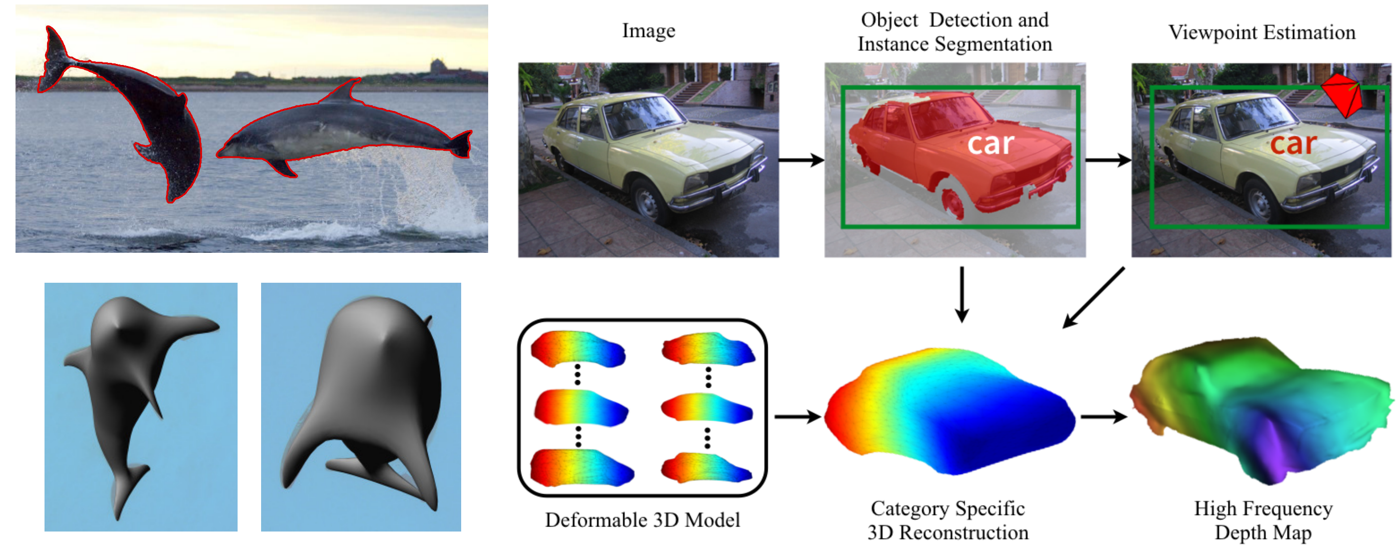
Supervision : Ground-truth 3D

Single-view Reconstruction



Roberts. PhD Thesis, MIT. 1963

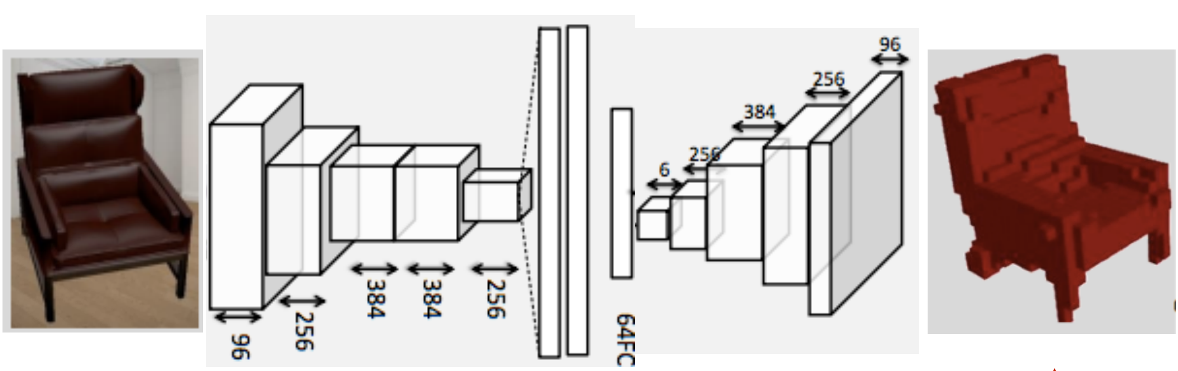
Unsupervised

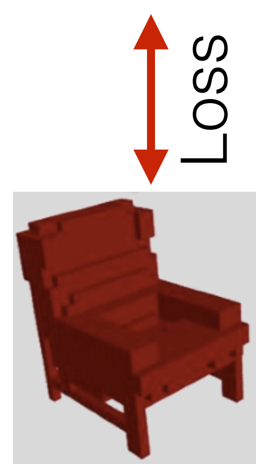


Cashman & Fitzgibbon, PAMI 2013

Kar et al., CVPR 2015

Supervision : Masks + Pose

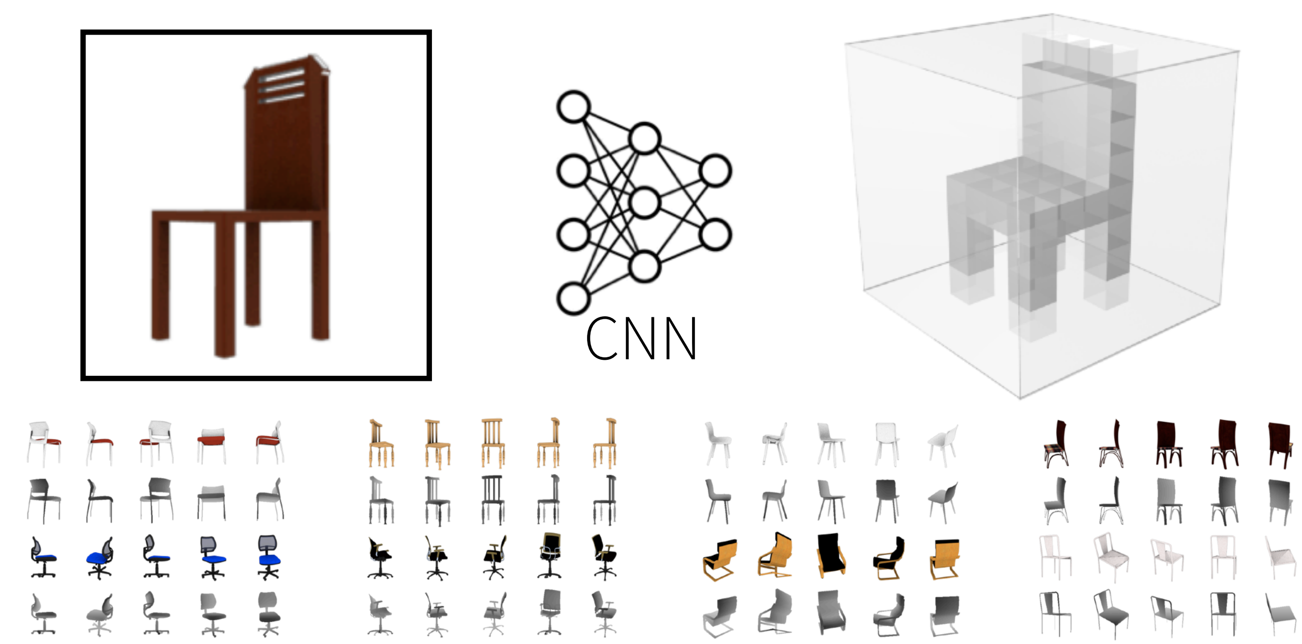




Choy et al., Girdhar et al.

ECCV 2016

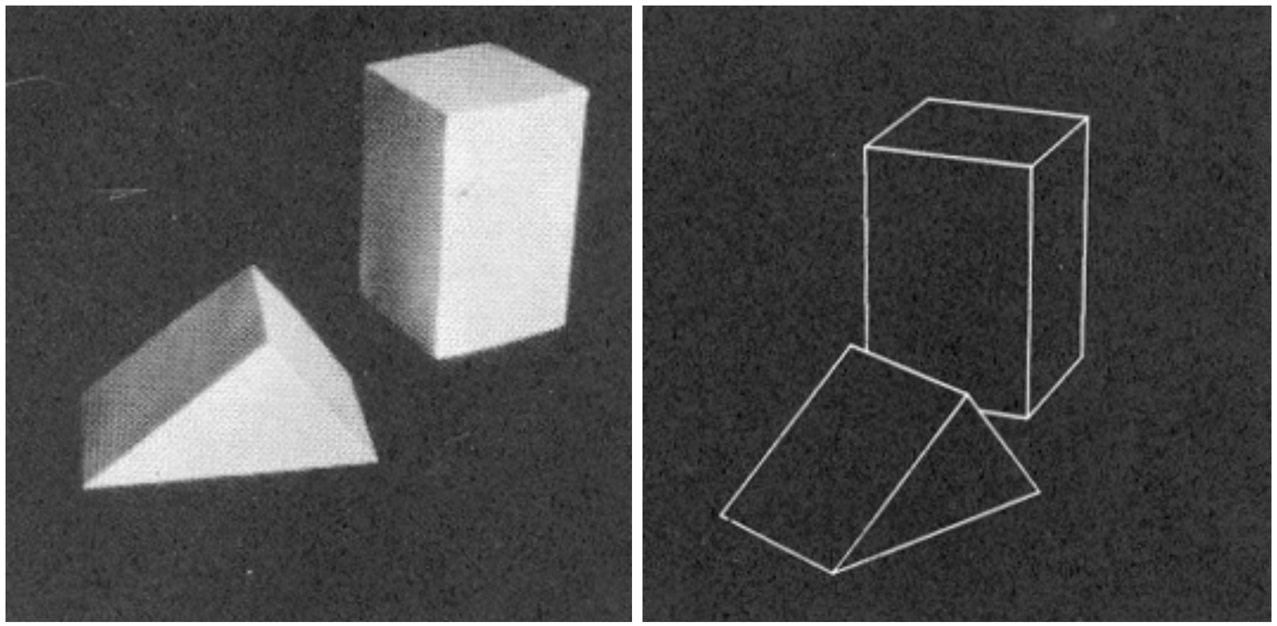
Supervision : Ground-truth 3D



Multi-view Supervision for Single-view Reconstruction via Differentiable Ray Consistency
Tulsiani et al.

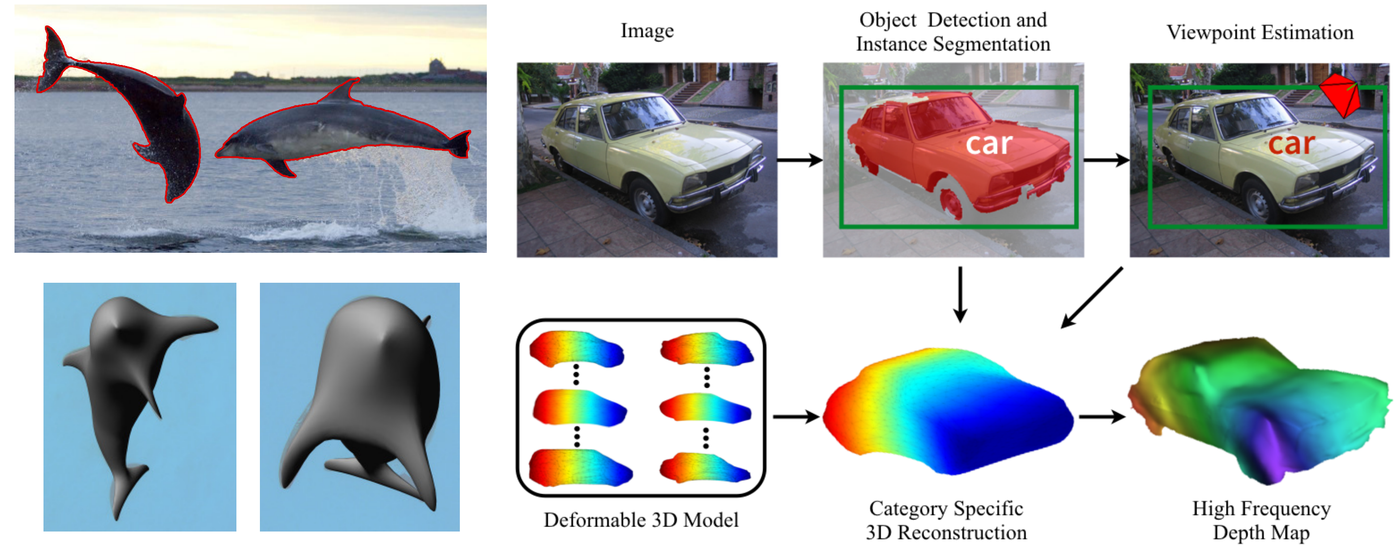
Supervision : Multi-view

Single-view Reconstruction



Roberts. PhD Thesis, MIT. 1963

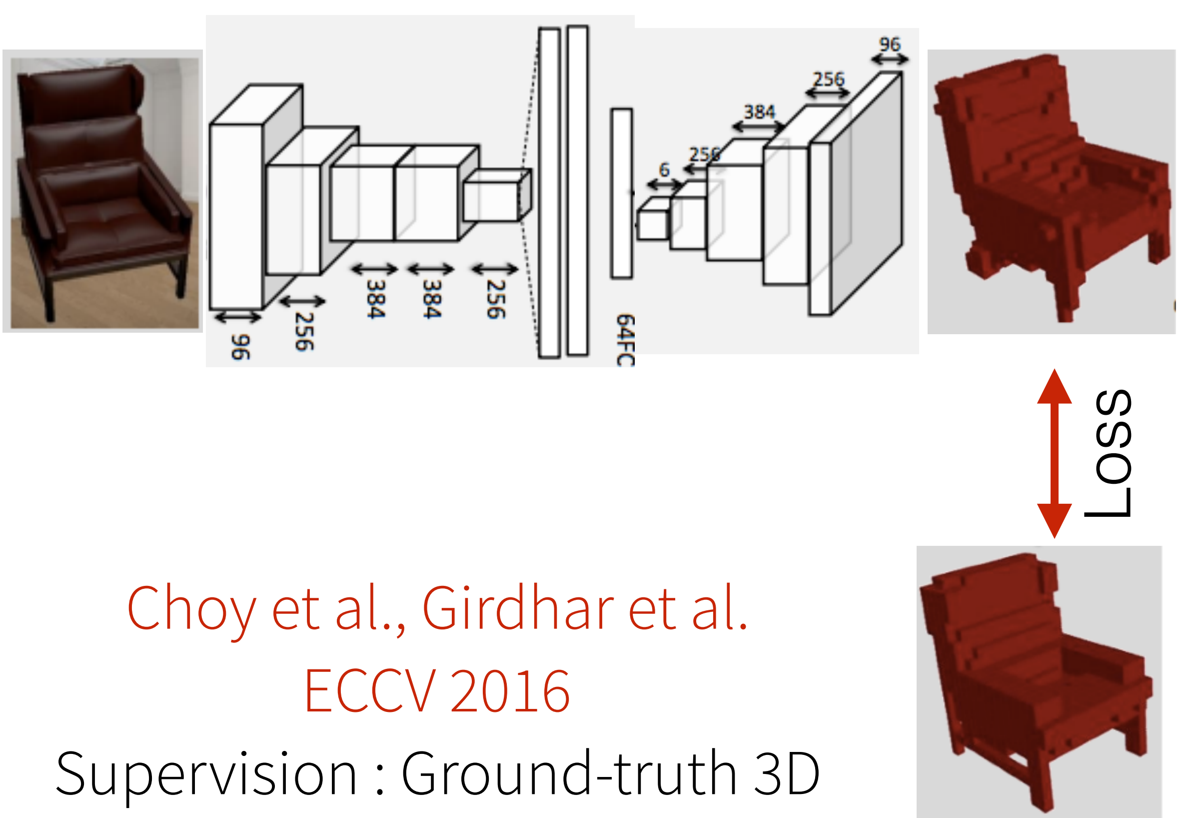
Unsupervised



Cashman & Fitzgibbon, PAMI 2013

Kar et al., CVPR 2015

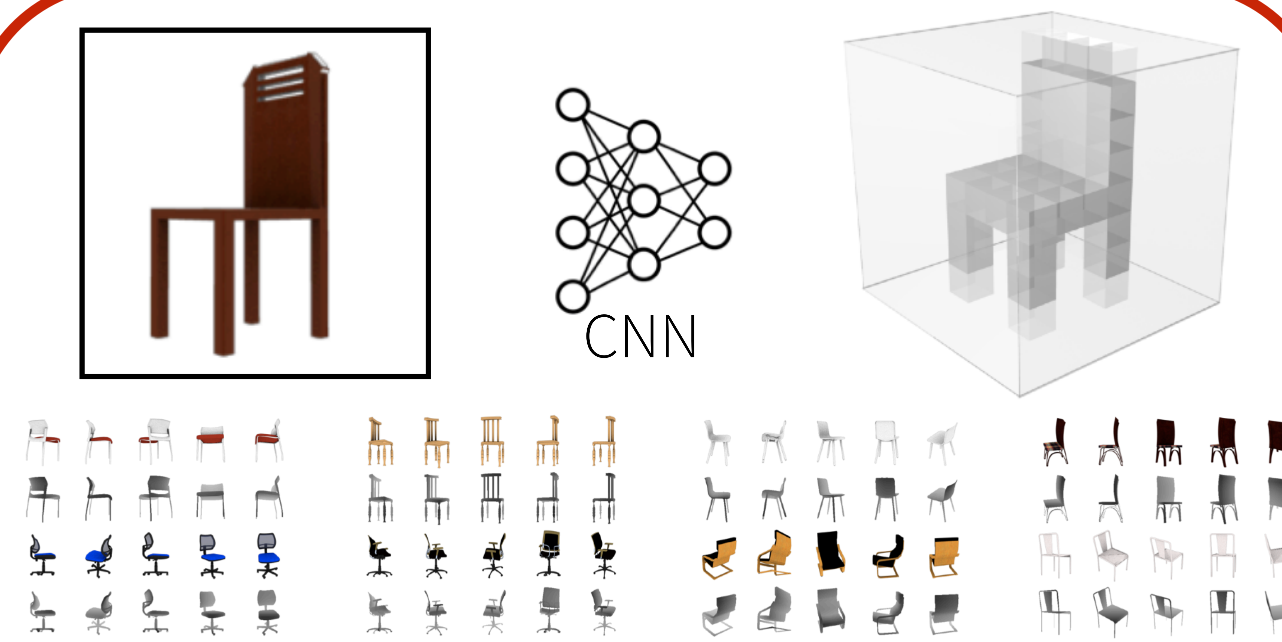
Supervision : Masks + Pose



Choy et al., Girdhar et al.

ECCV 2016

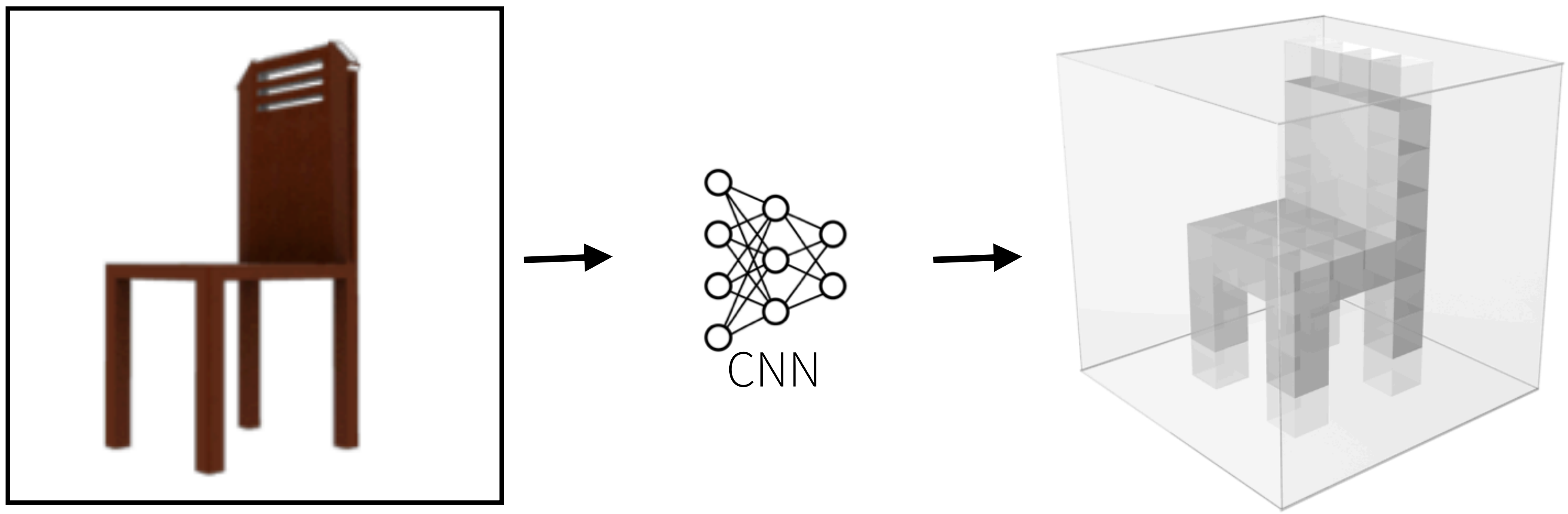
Supervision : Ground-truth 3D



Multi-view Supervision for Single-view Reconstruction via Differentiable Ray Consistency
Tulsiani et al.

Supervision : Multi-view

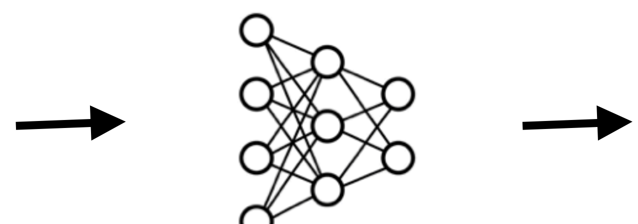
How to use Multi-view Supervision ?

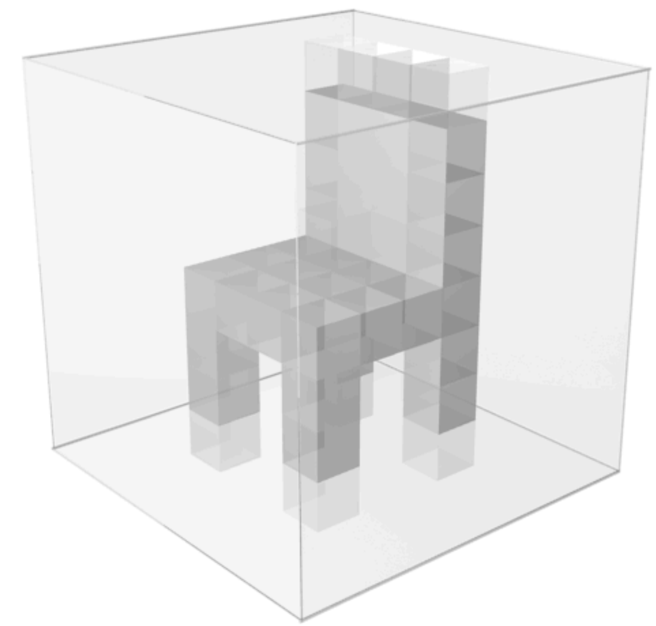


Input Image

How to use Multi-view Supervision ?

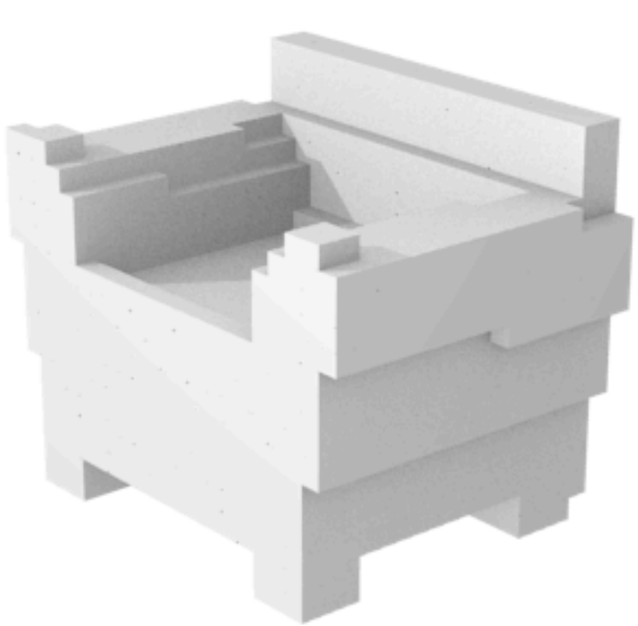
Input Image

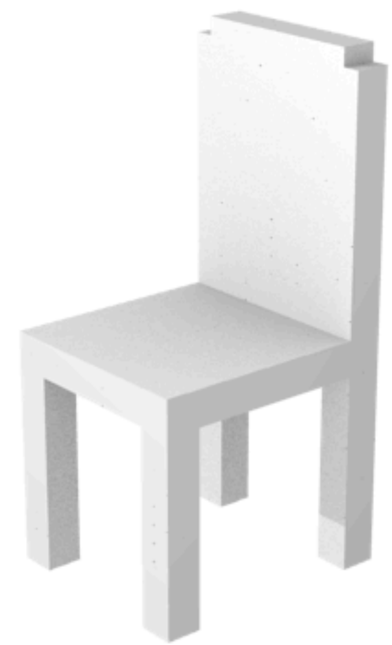


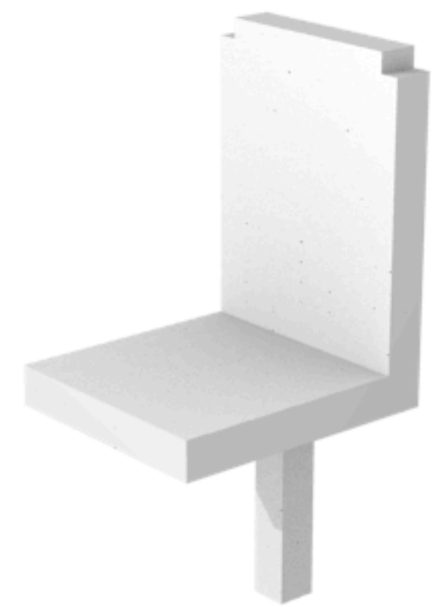


Observation **O**
from camera **C**

How to use Multi-view Supervision ?

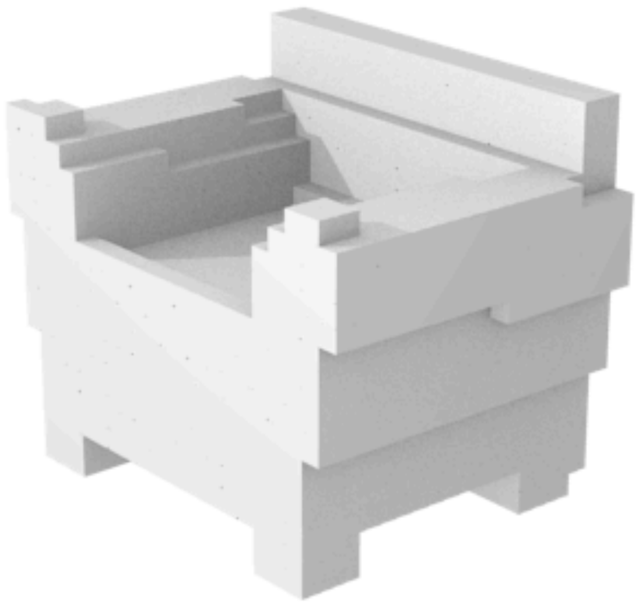




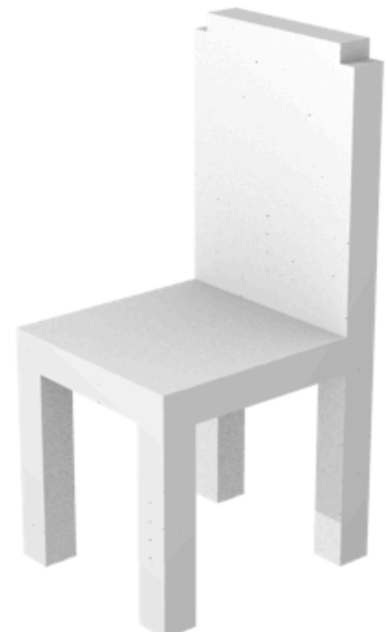


Observation **O**
from camera **C**

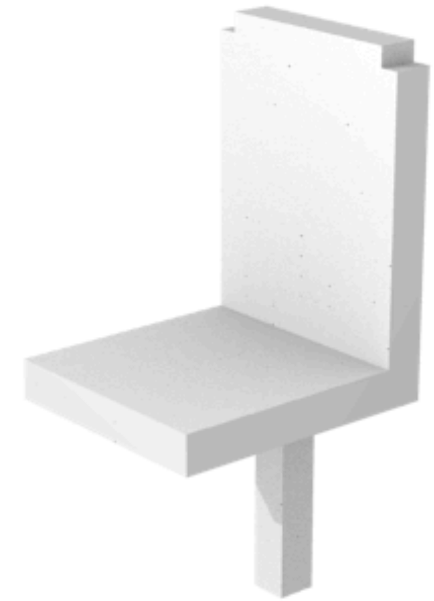
How to use Multi-view Supervision ?



Geometrically Inconsistent



Geometrically Consistent

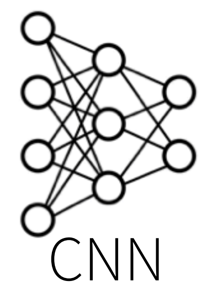


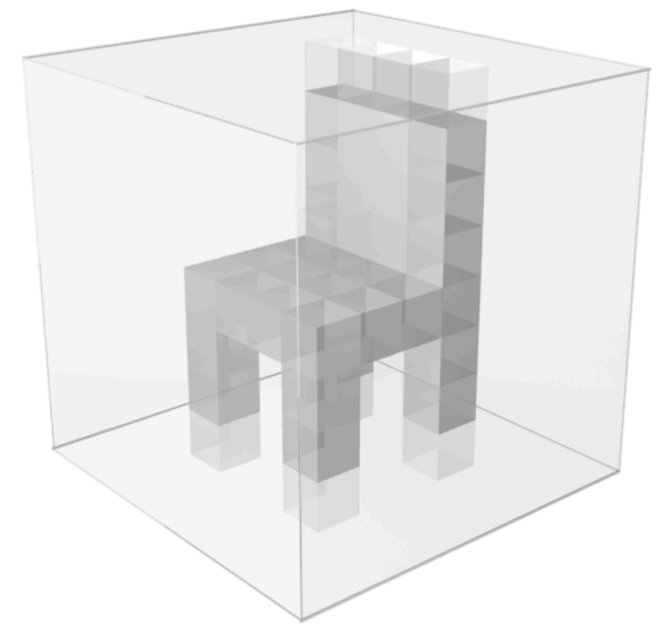
Geometrically Inconsistent

Observation **O** from camera **C**

Learning via Geometric Consistency

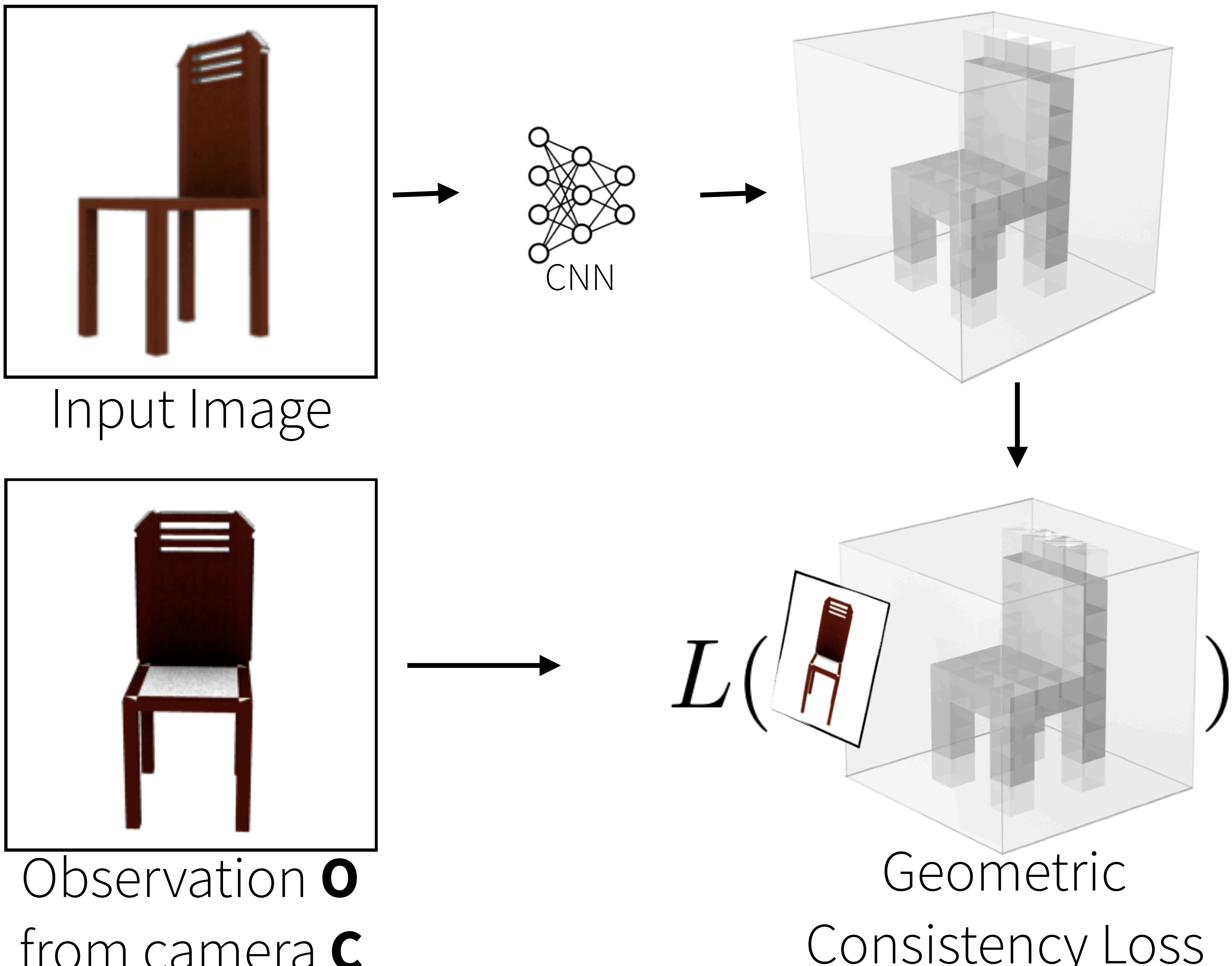
Input Image



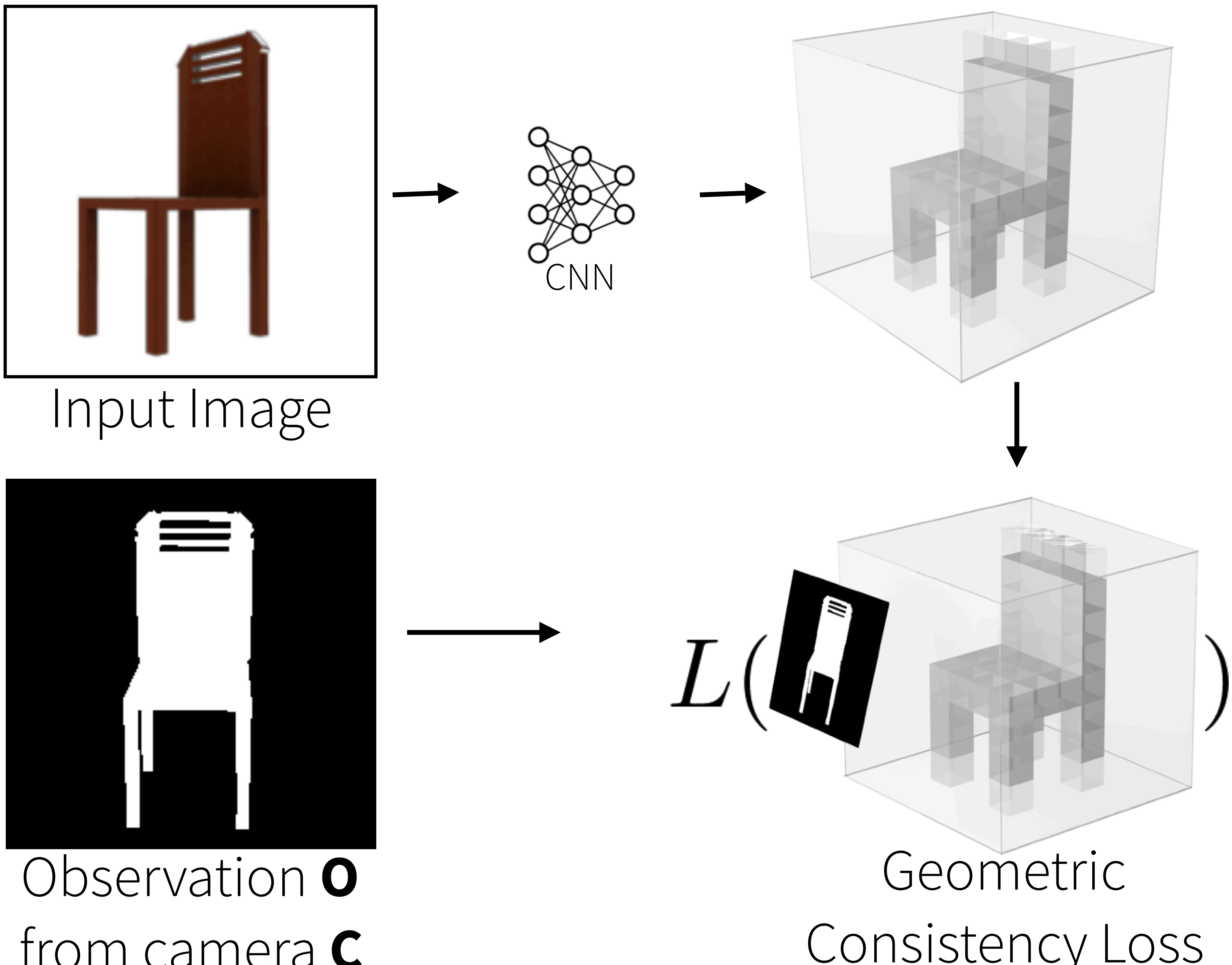


Observation **O**
from camera **C**

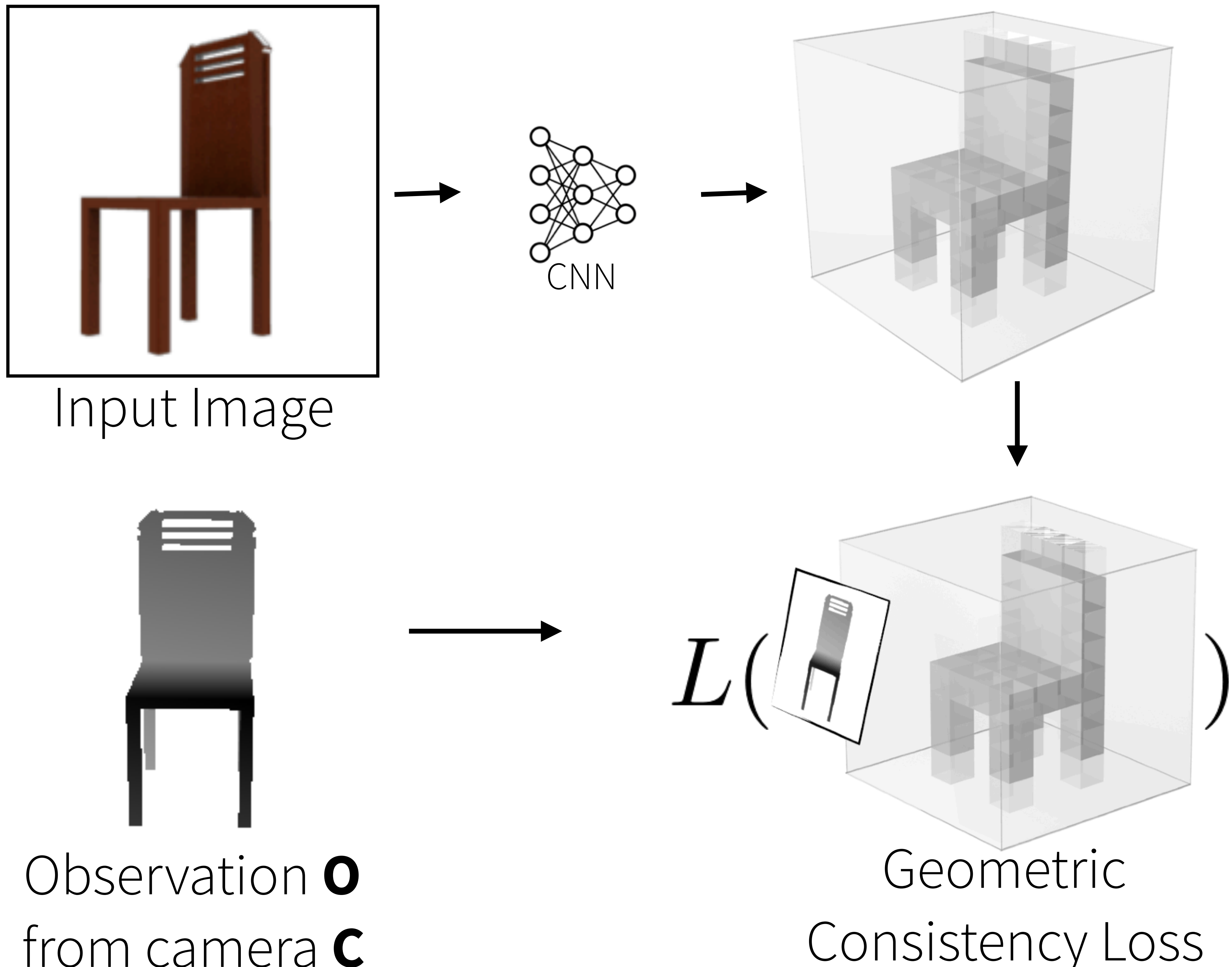
Learning via Geometric Consistency



Learning via Geometric Consistency

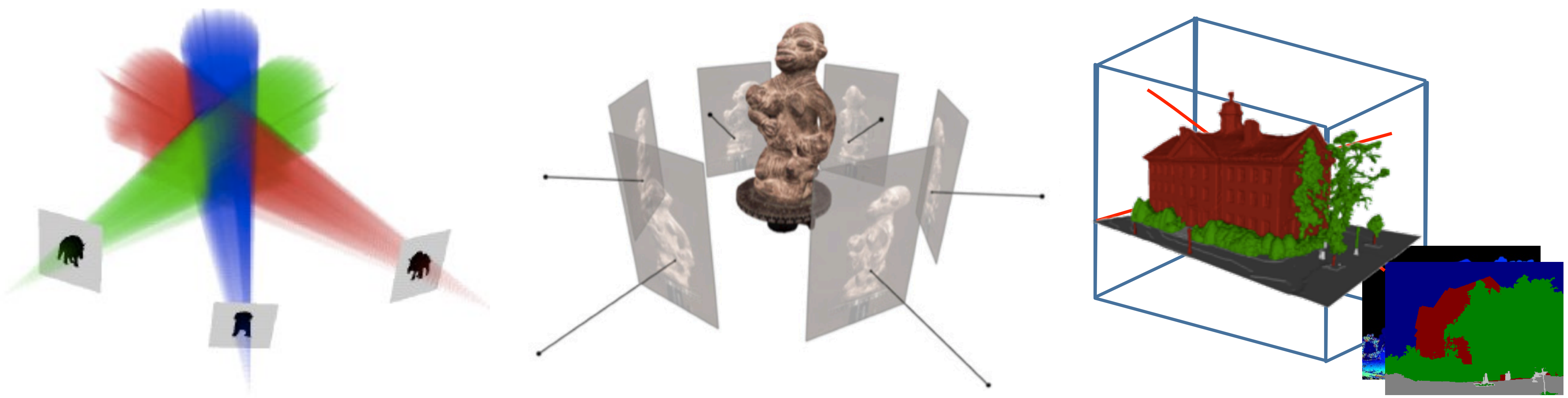


Learning via Geometric Consistency



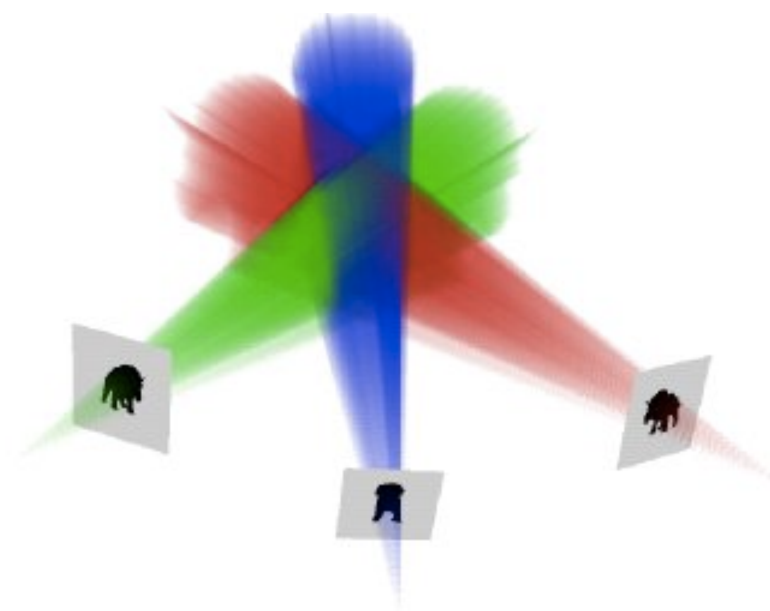
3D from Geometric Consistency

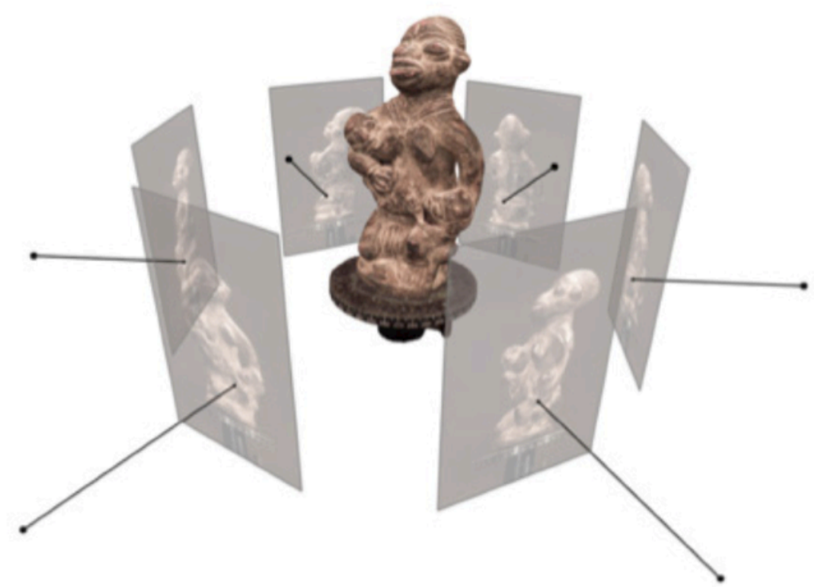
3D from Geometric Consistency

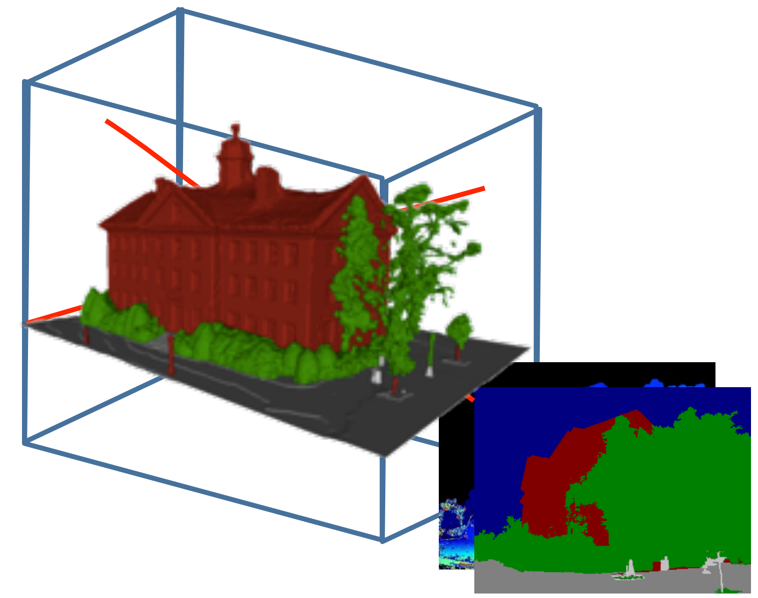


Space Carving, Multi-view Stereo, Multi-view Reconstruction

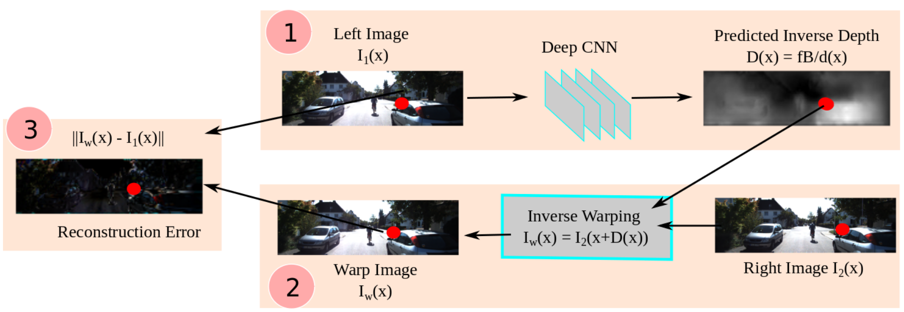
3D from Geometric Consistency

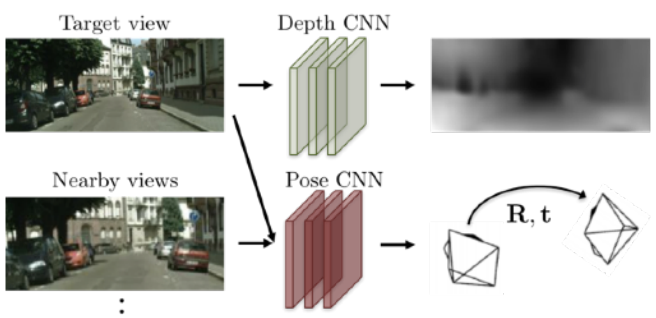


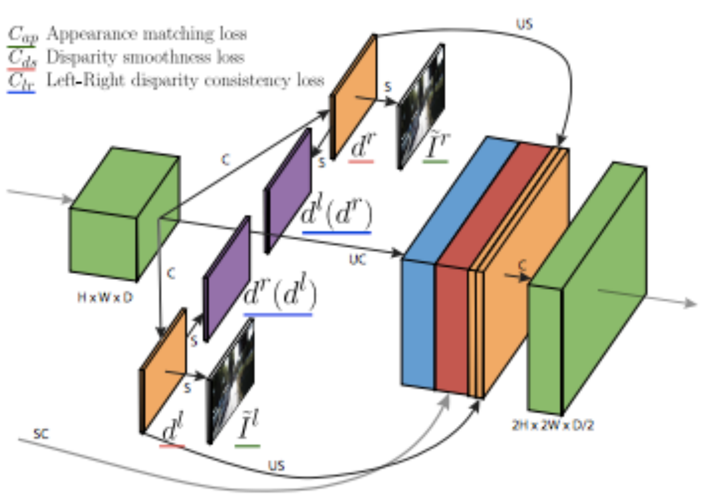




Space Carving, Multi-view Stereo, Multi-view Reconstruction



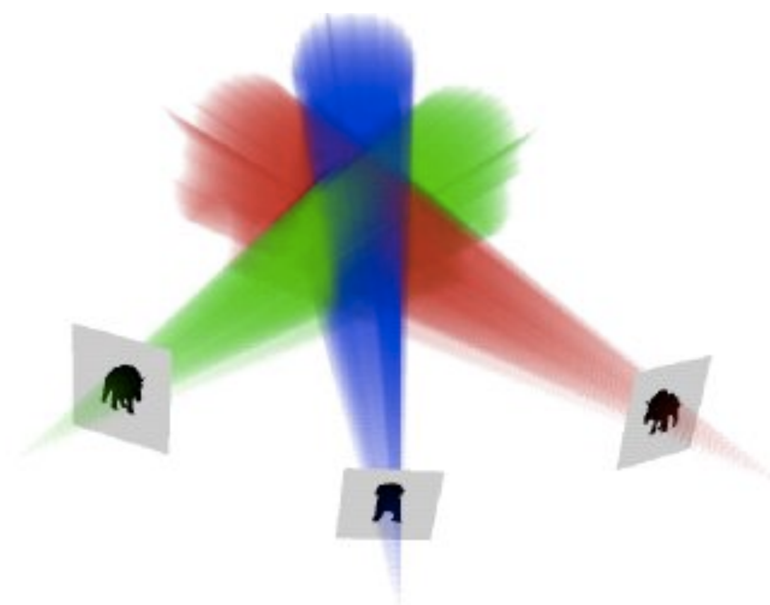


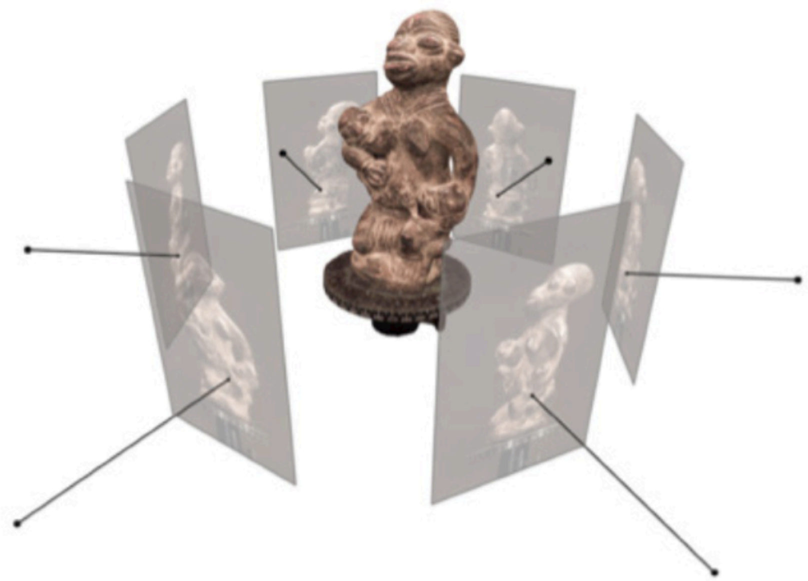


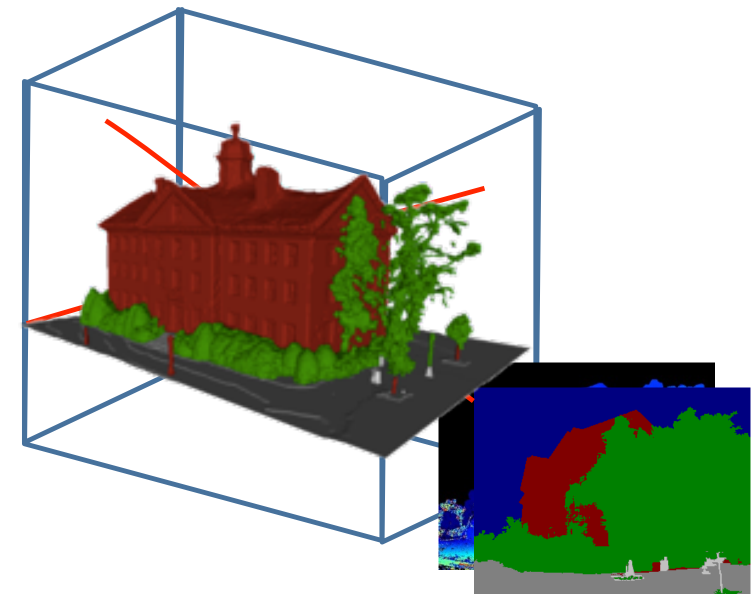
Garg et. al. ECCV 16

Godard et. al., Zhou et. al., CVPR 17

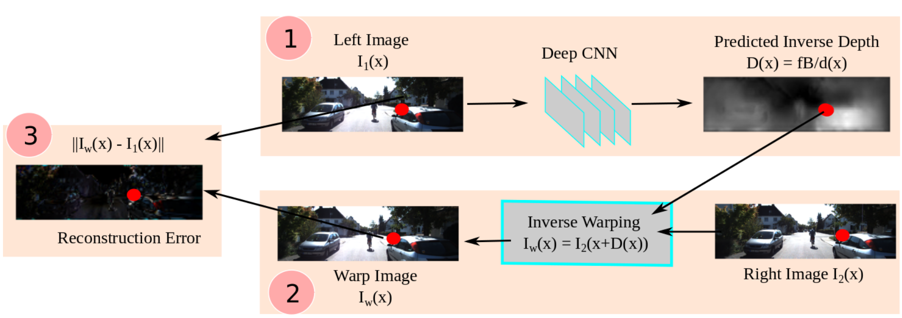
3D from Geometric Consistency

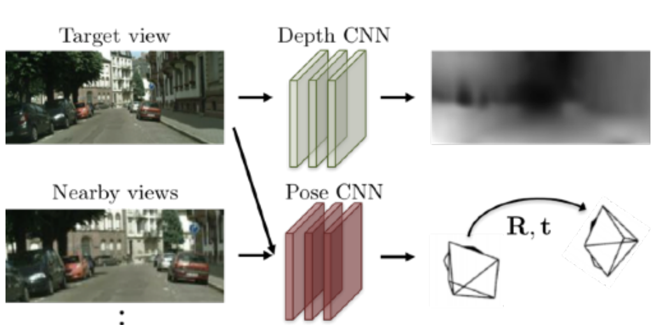


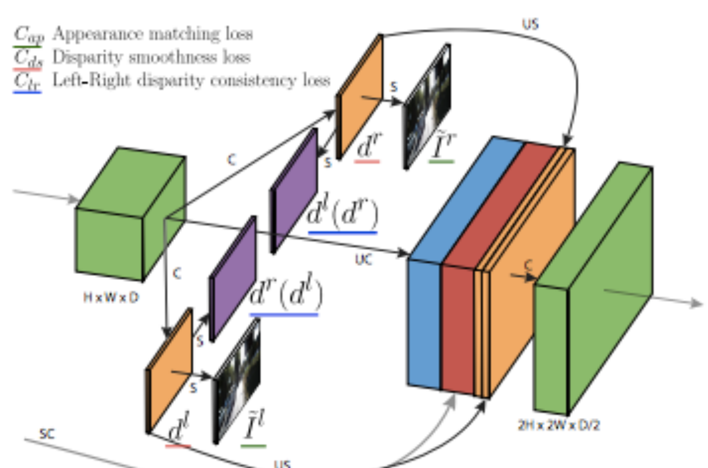


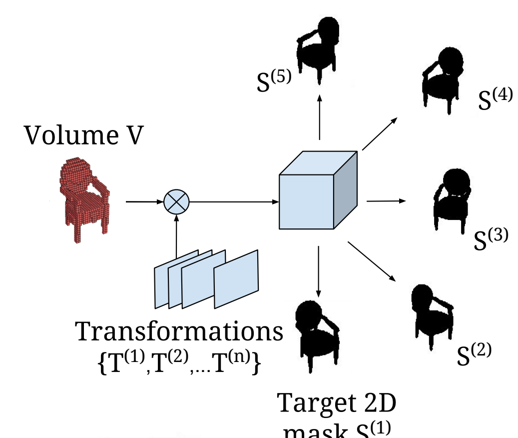


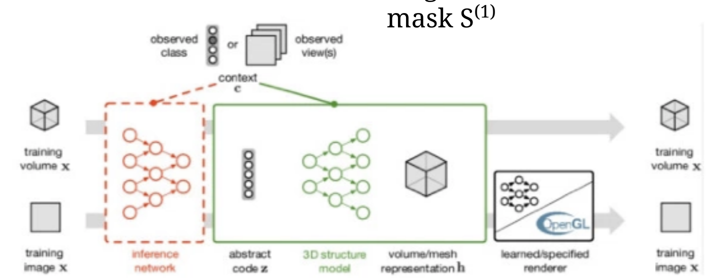
Space Carving, Multi-view Stereo, Multi-view Reconstruction











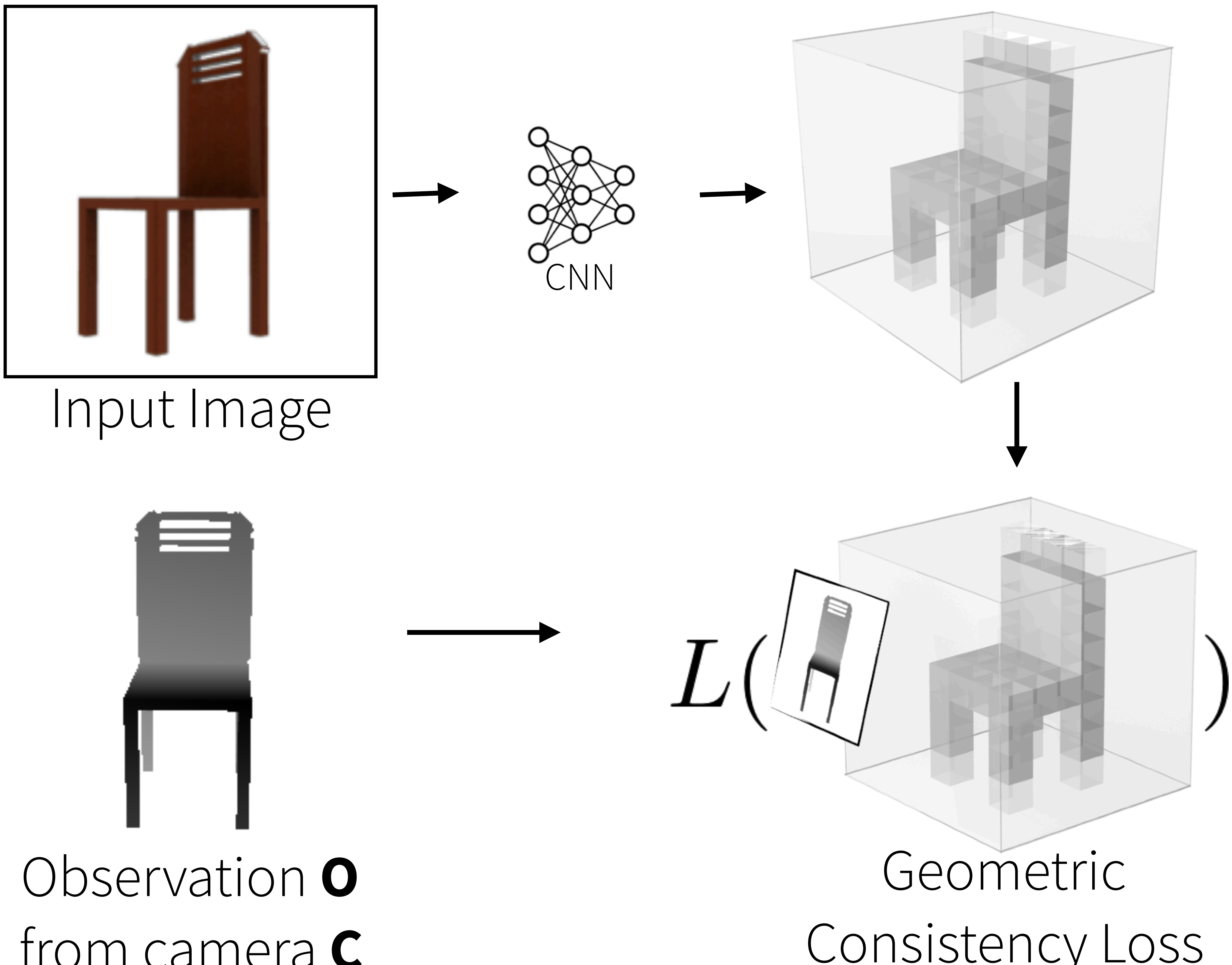
Garg et. al. ECCV 16

Godard et. al., Zhou et. al., CVPR 17

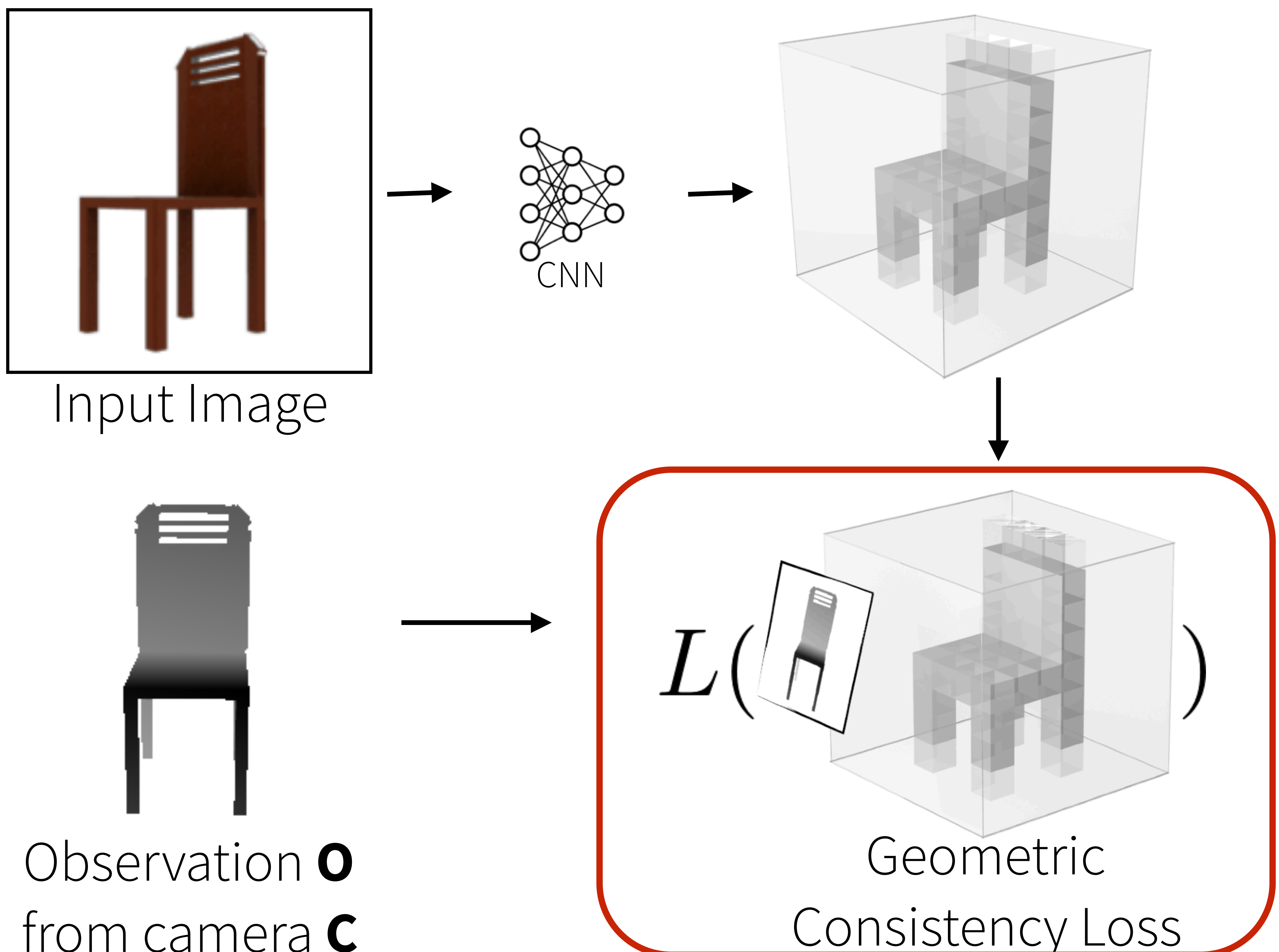
Yan et. al., Rezende et. al.

NIPS 16

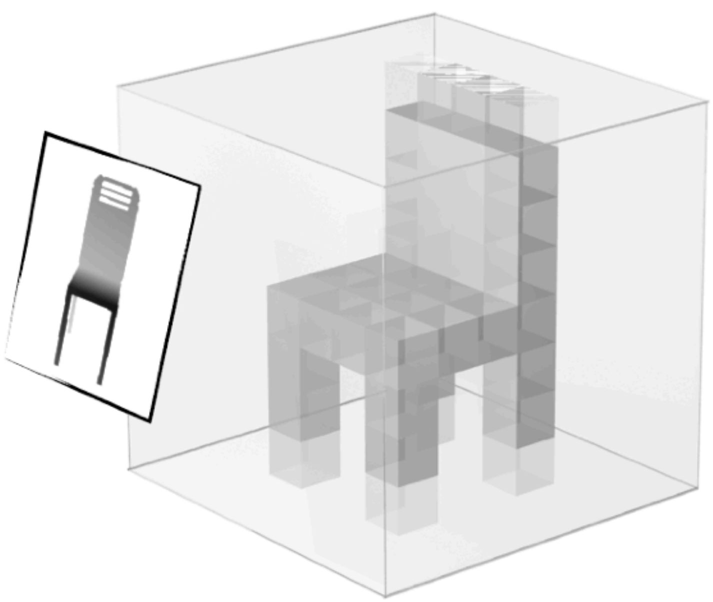
Learning via Geometric Consistency



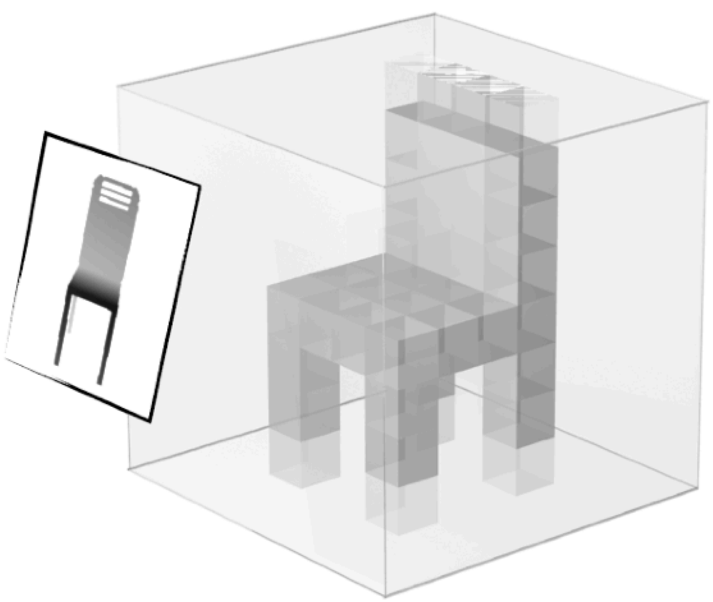
Learning via Geometric Consistency



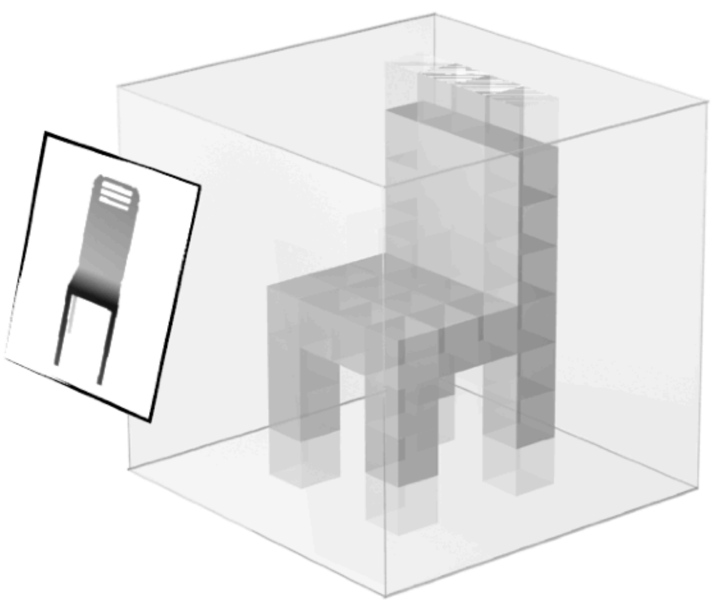
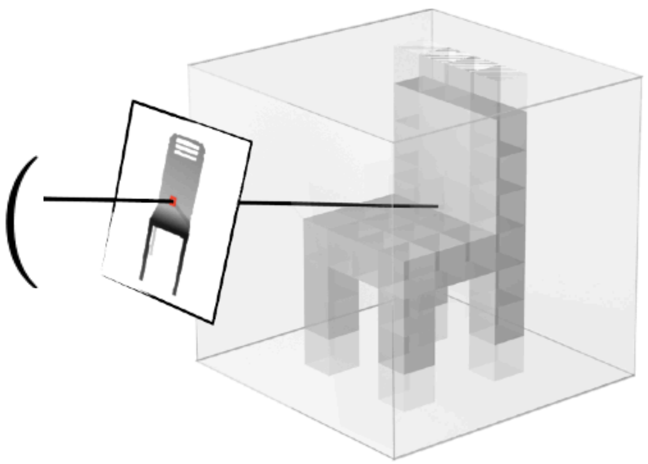
View Consistency as Ray Consistency

$L($  $)$

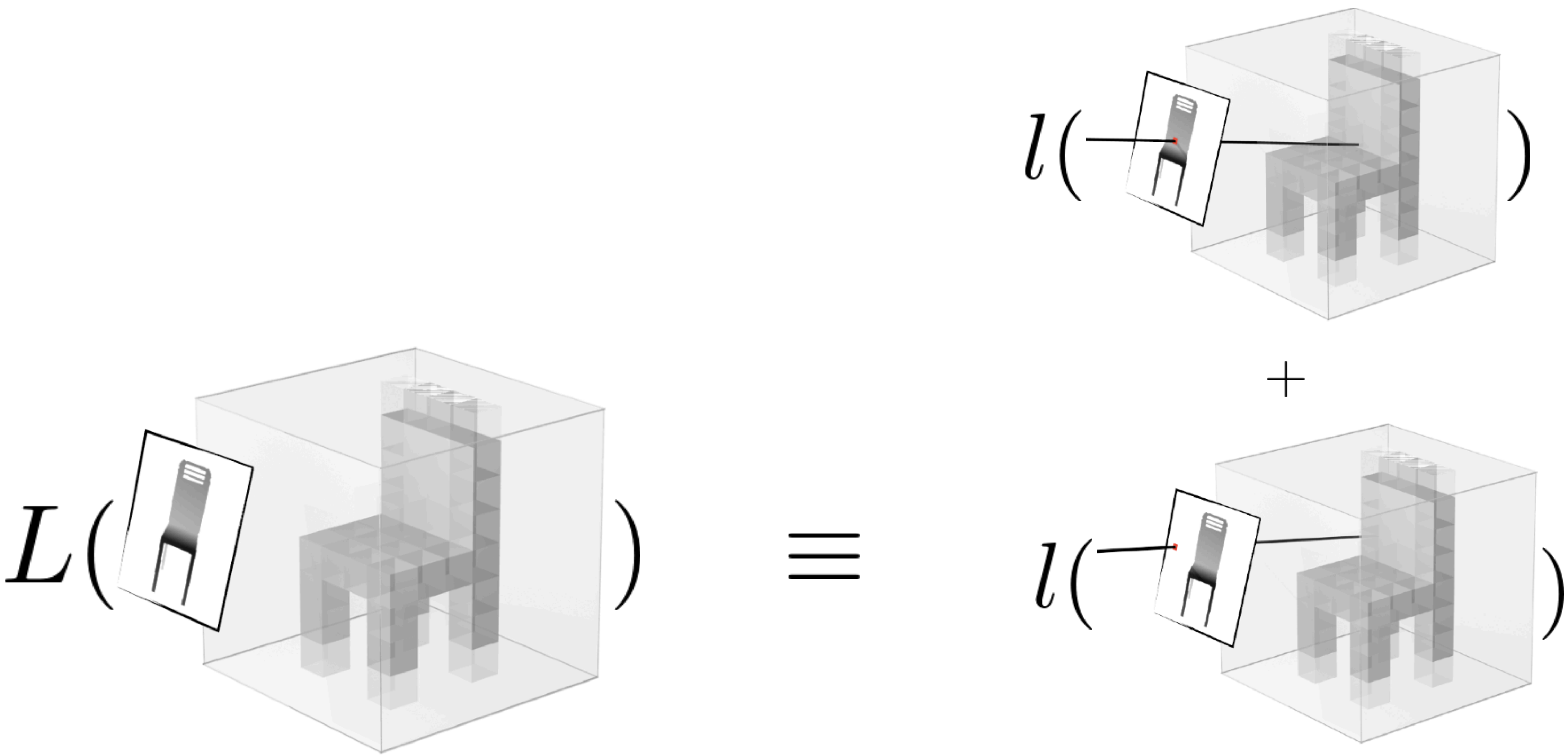
View Consistency as Ray Consistency

$L($  $) \quad \equiv$

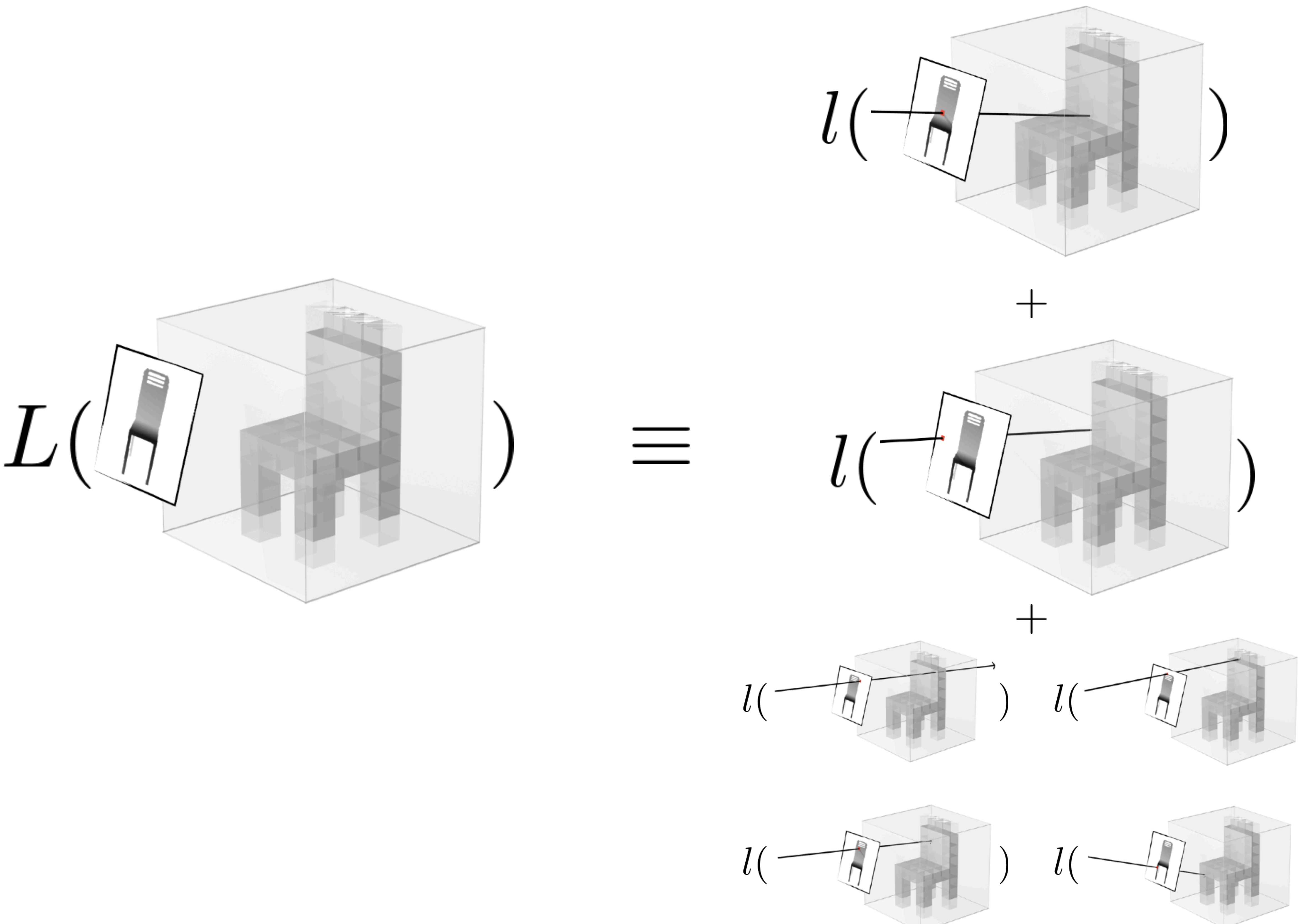
View Consistency as Ray Consistency

$L($  $)$ $\equiv$ $l($  $)$

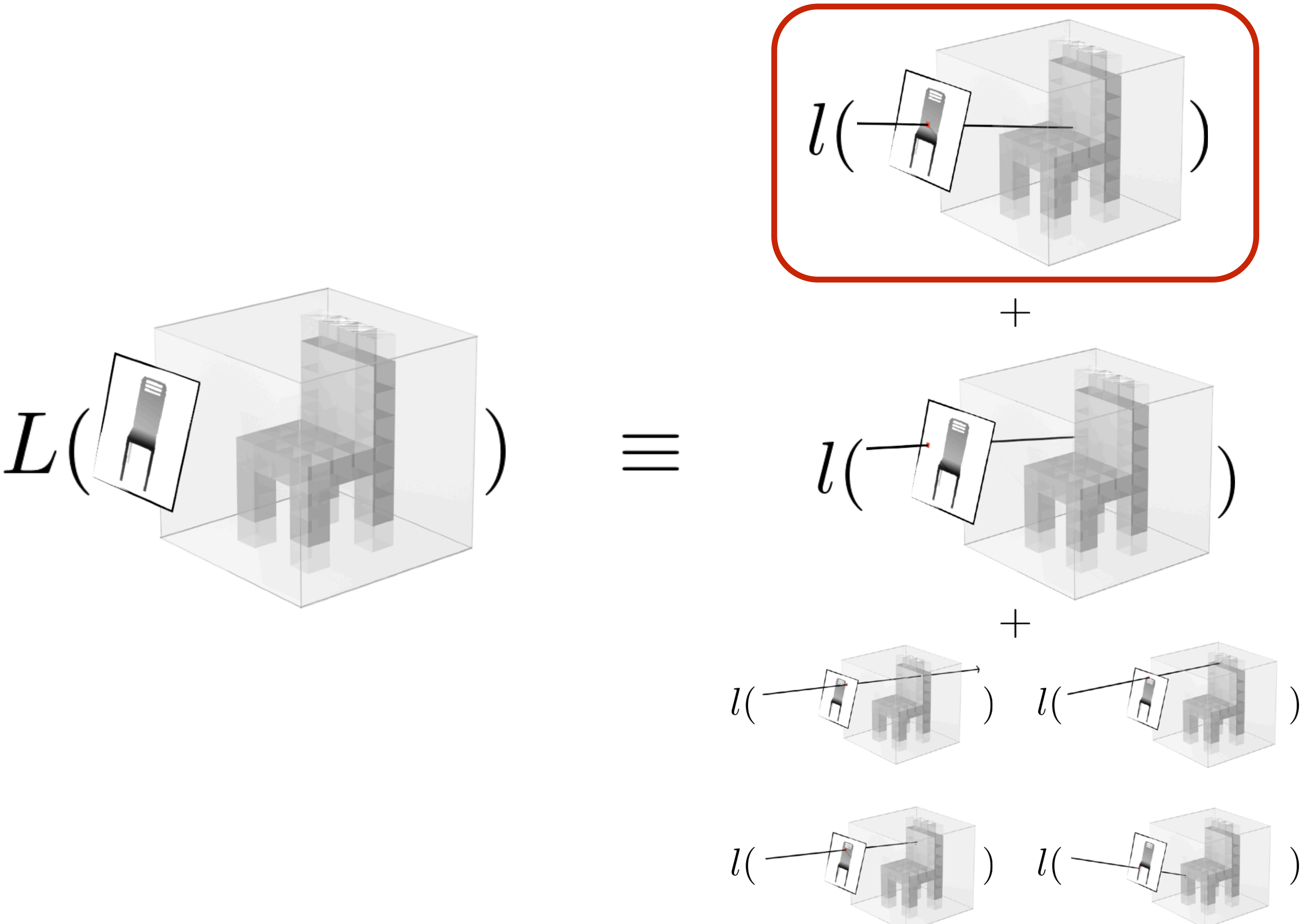
View Consistency as Ray Consistency



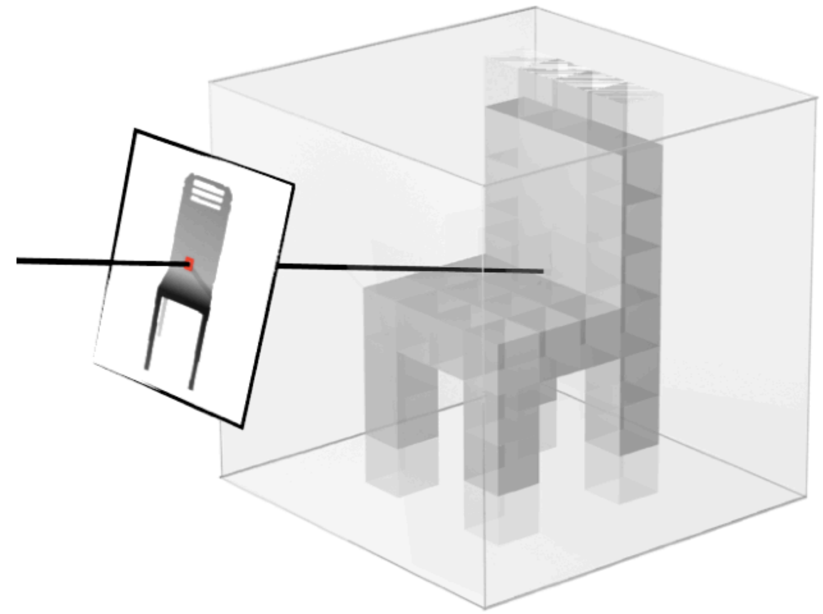
View Consistency as Ray Consistency



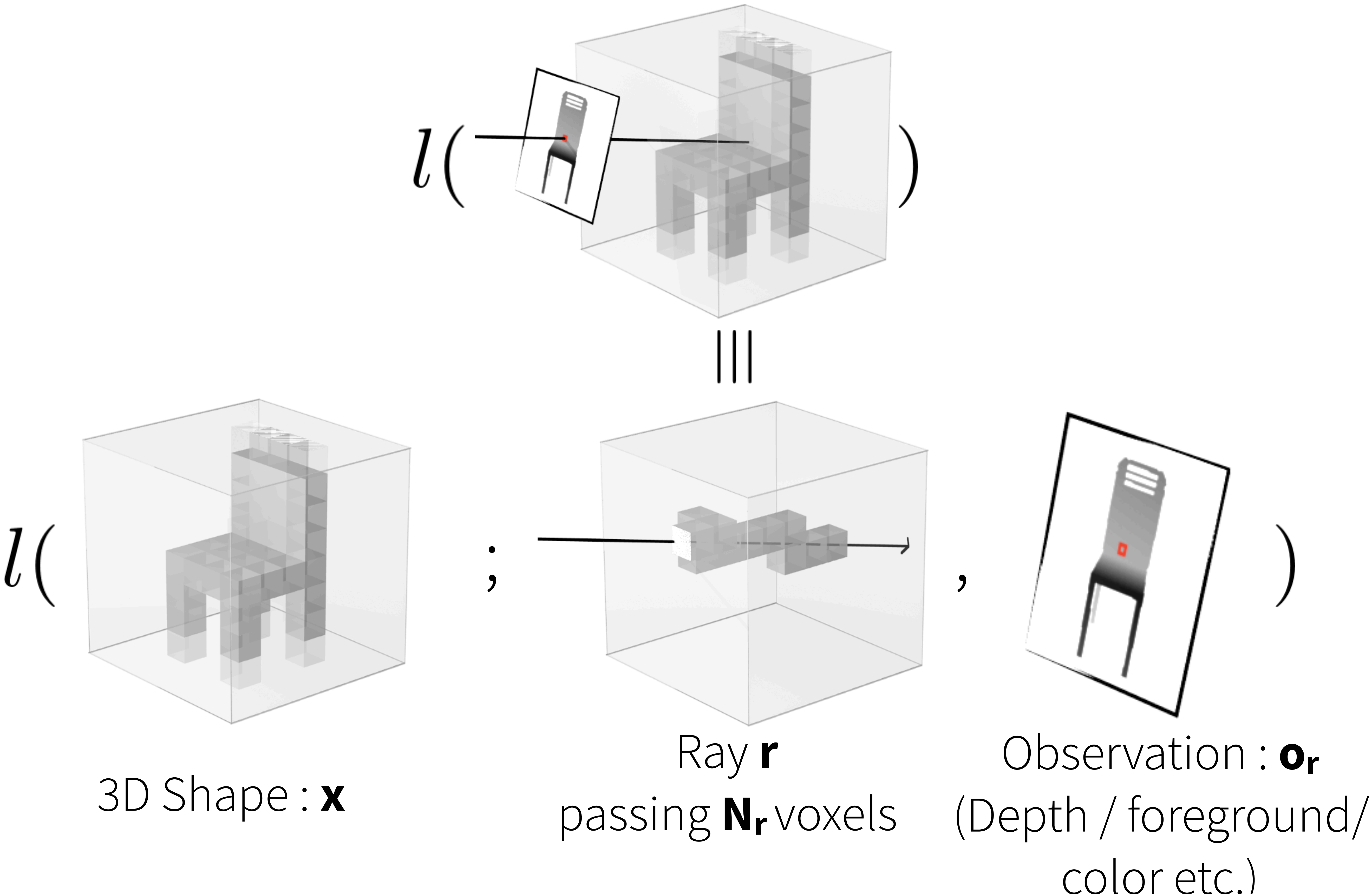
View Consistency as Ray Consistency



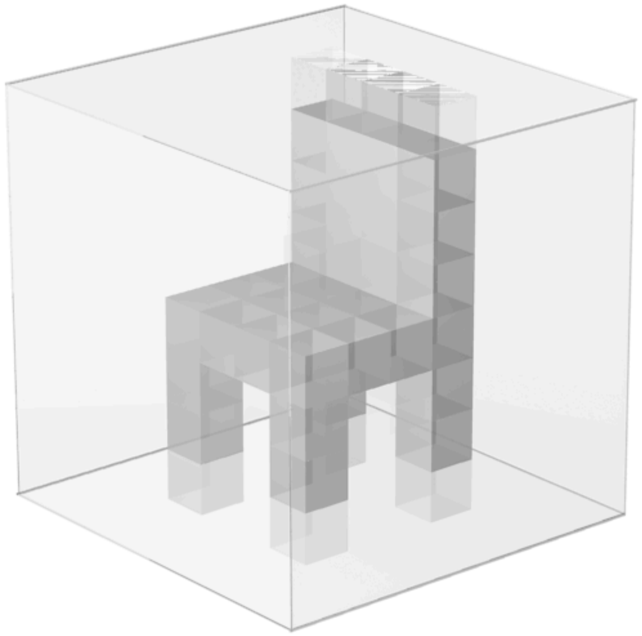
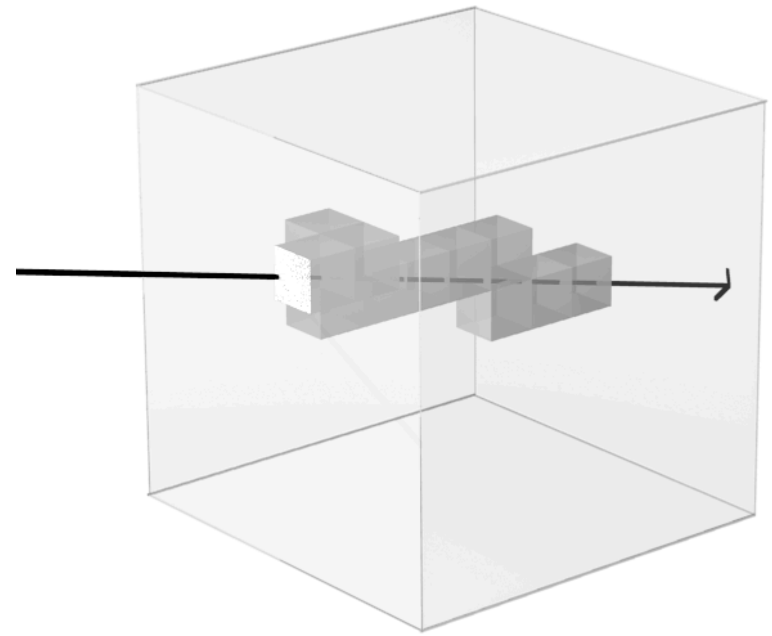
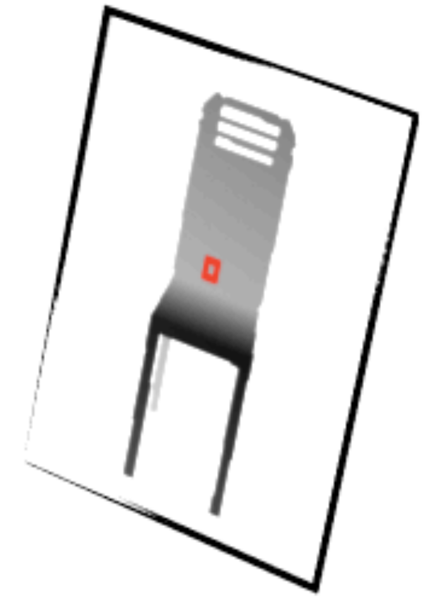
Differentiable Ray Consistency

$l($  $)$

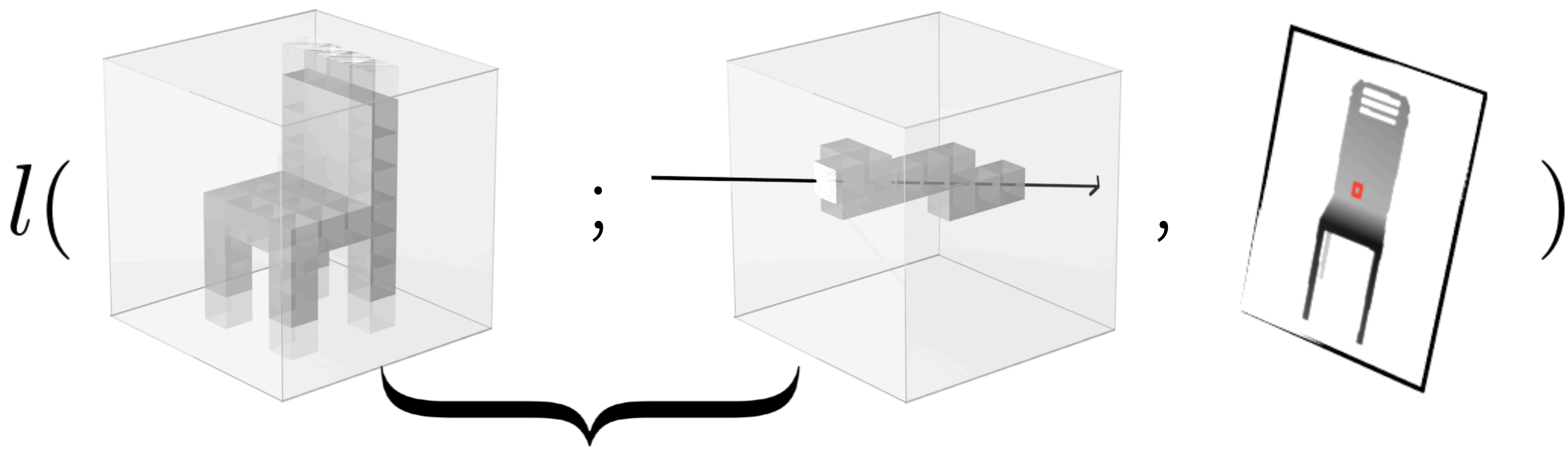
Differentiable Ray Consistency



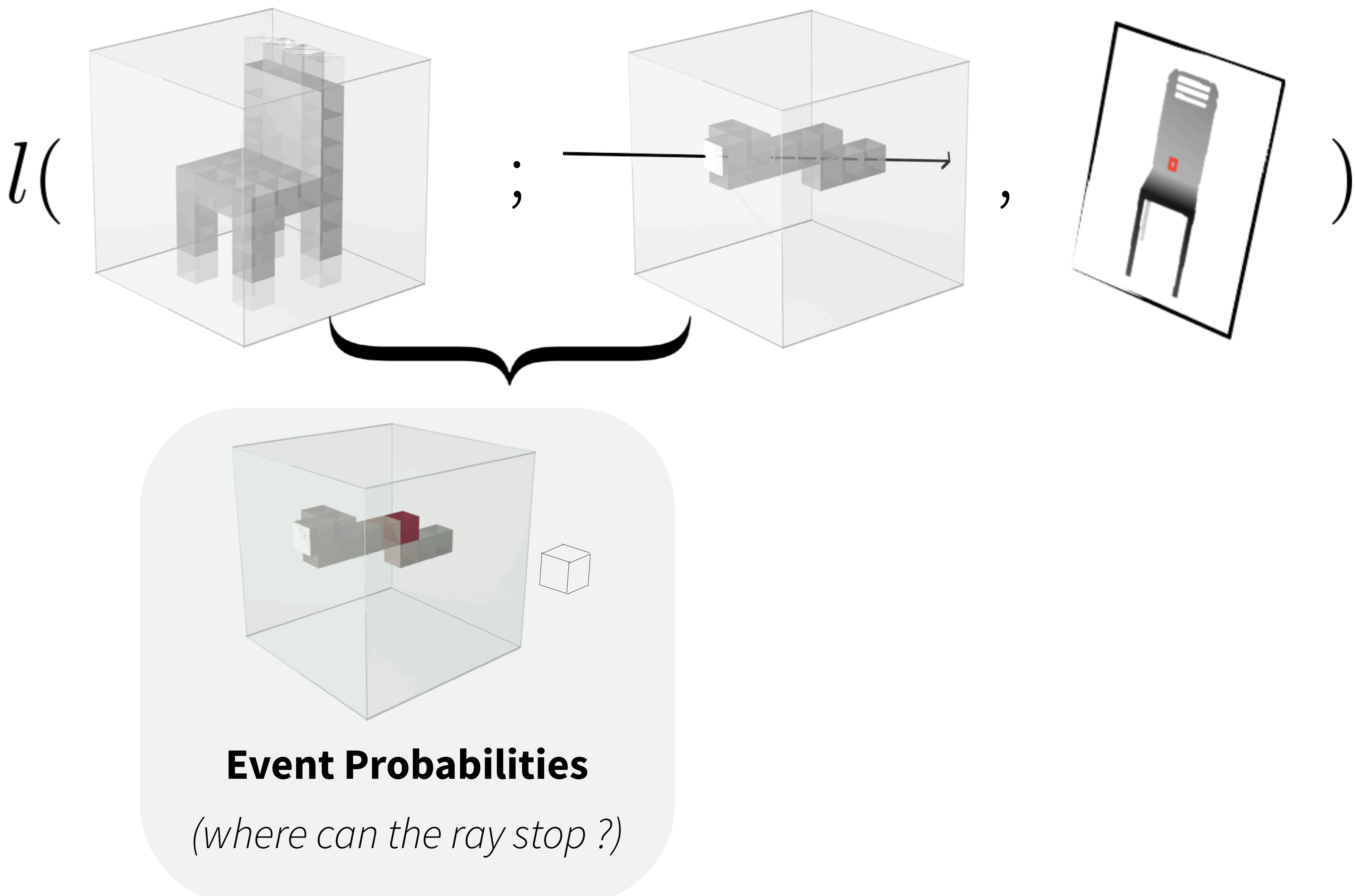
Differentiable Ray Consistency

$l($  ;  ,  $)$

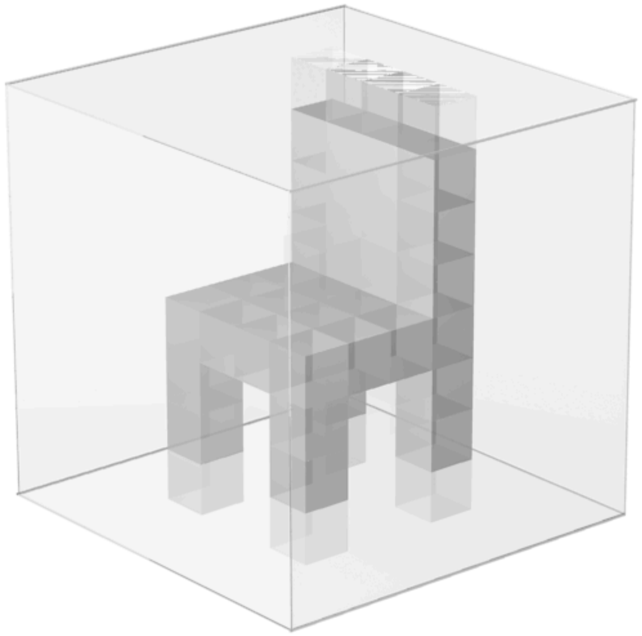
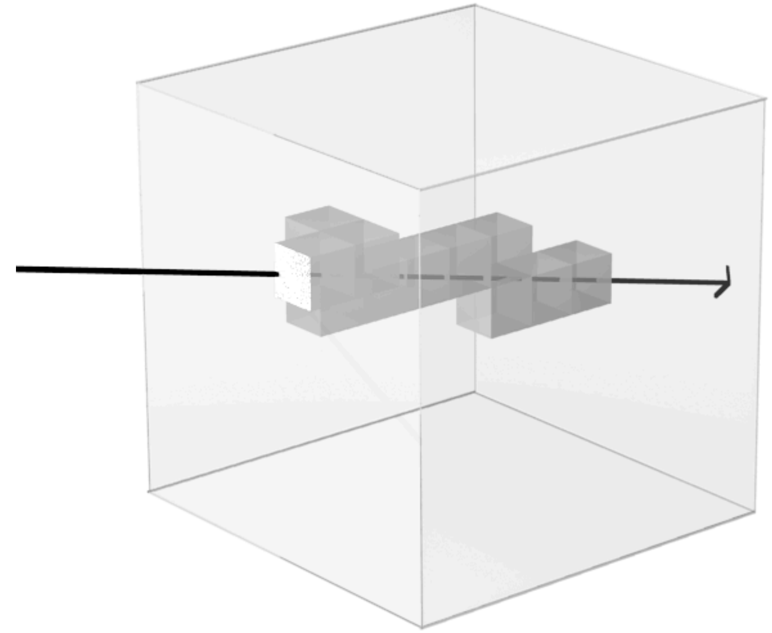
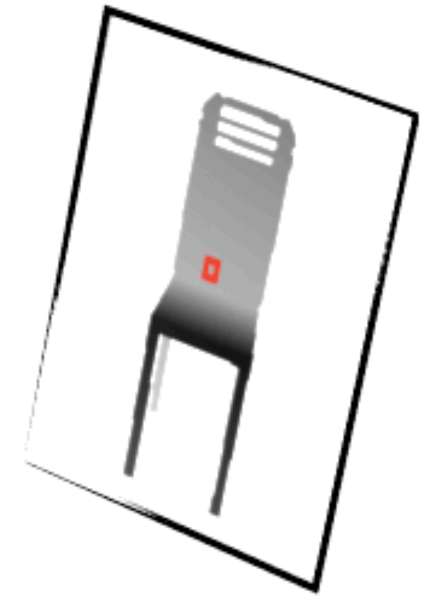
Differentiable Ray Consistency

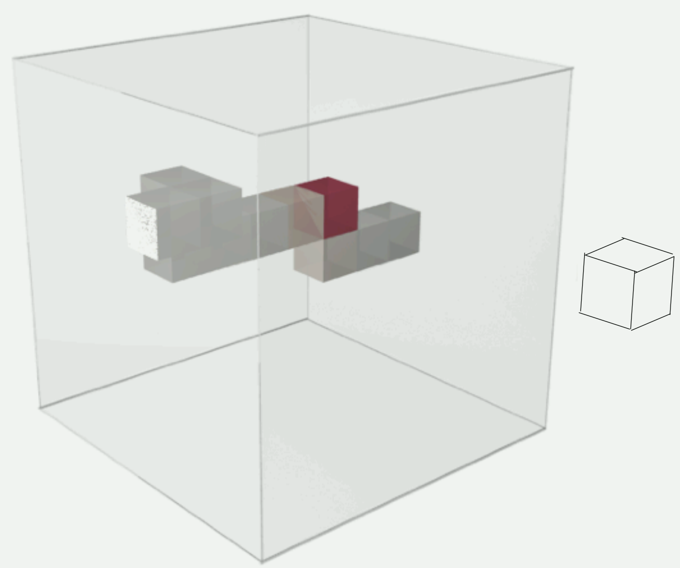


Differentiable Ray Consistency



Differentiable Ray Consistency

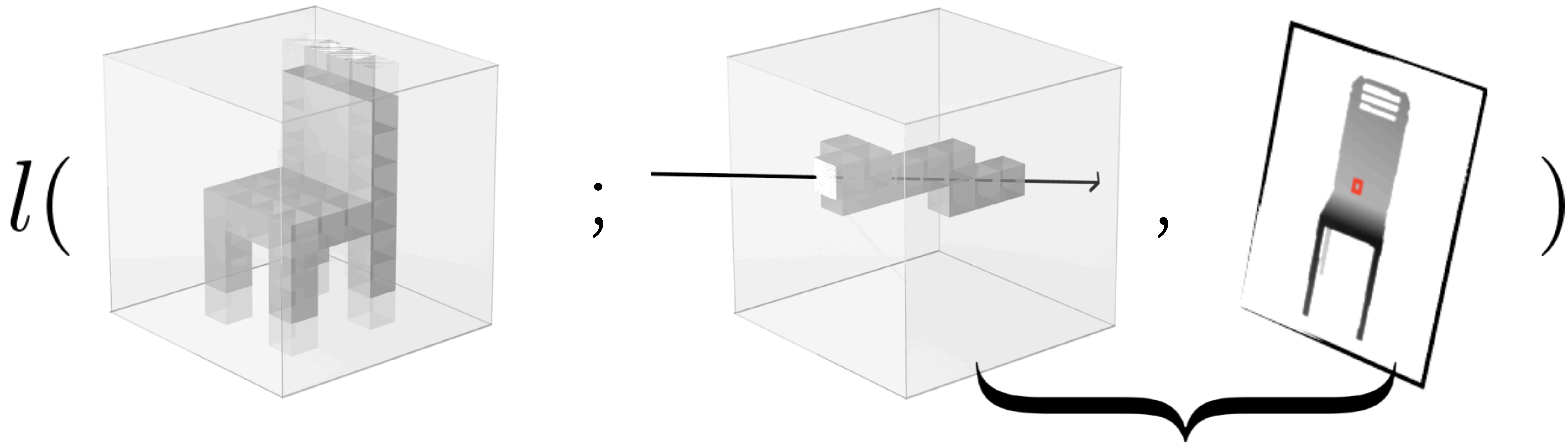
$l($  $;$  $,$  $)$

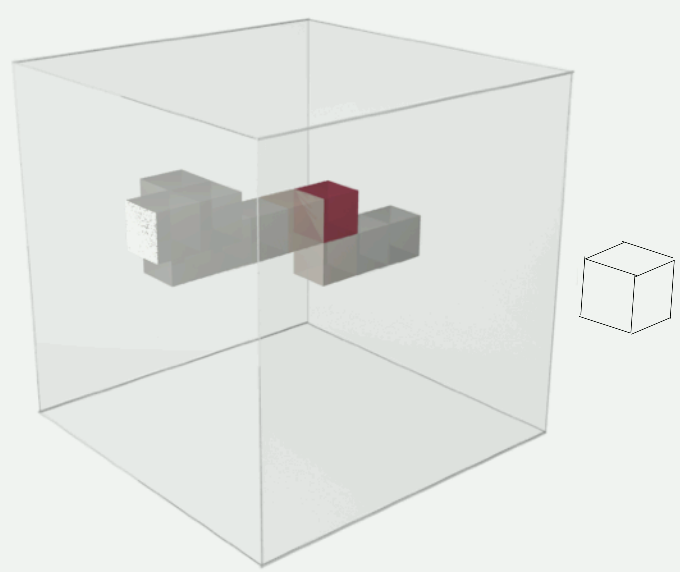


Event Probabilities

(where can the ray stop ?)

Differentiable Ray Consistency

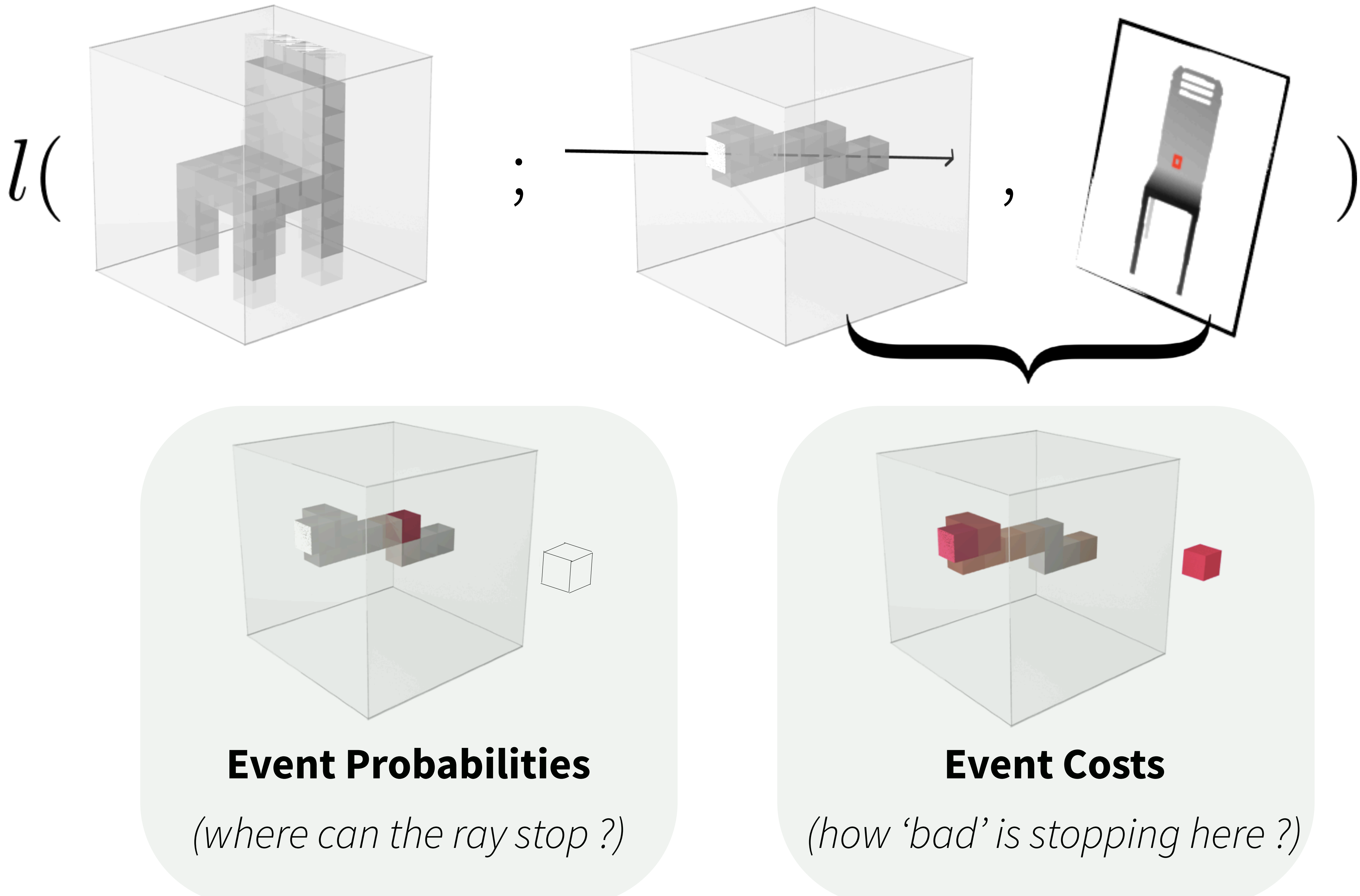




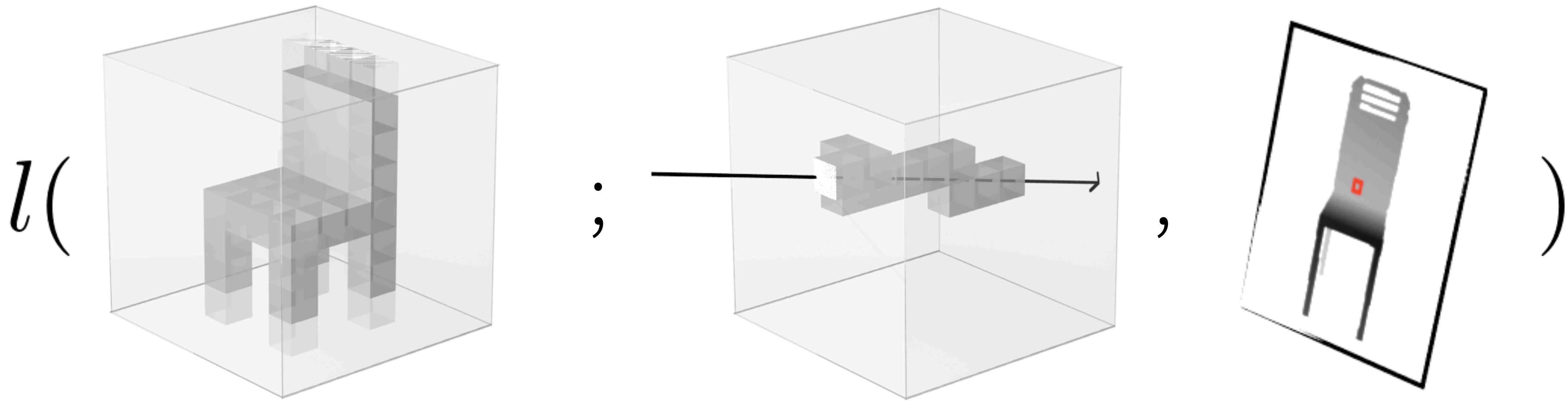
Event Probabilities

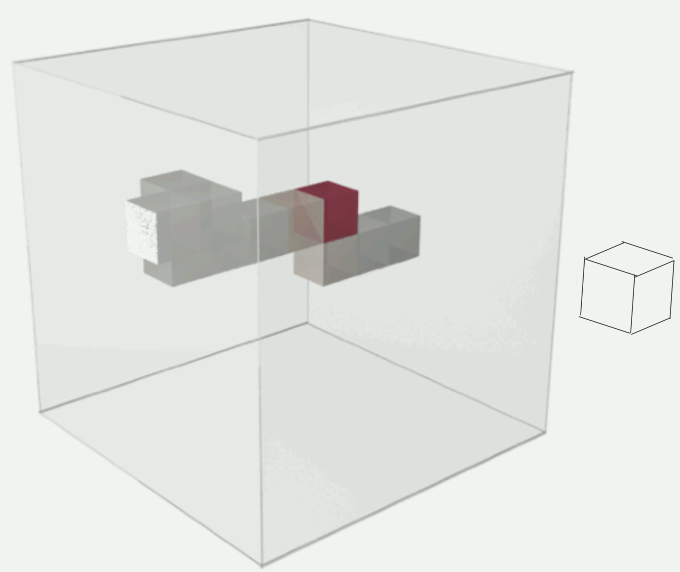
(where can the ray stop ?)

Differentiable Ray Consistency



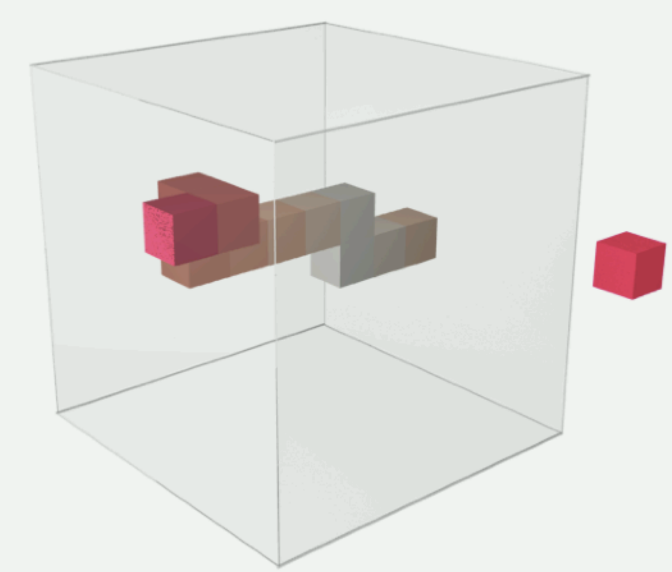
Differentiable Ray Consistency





Event Probabilities

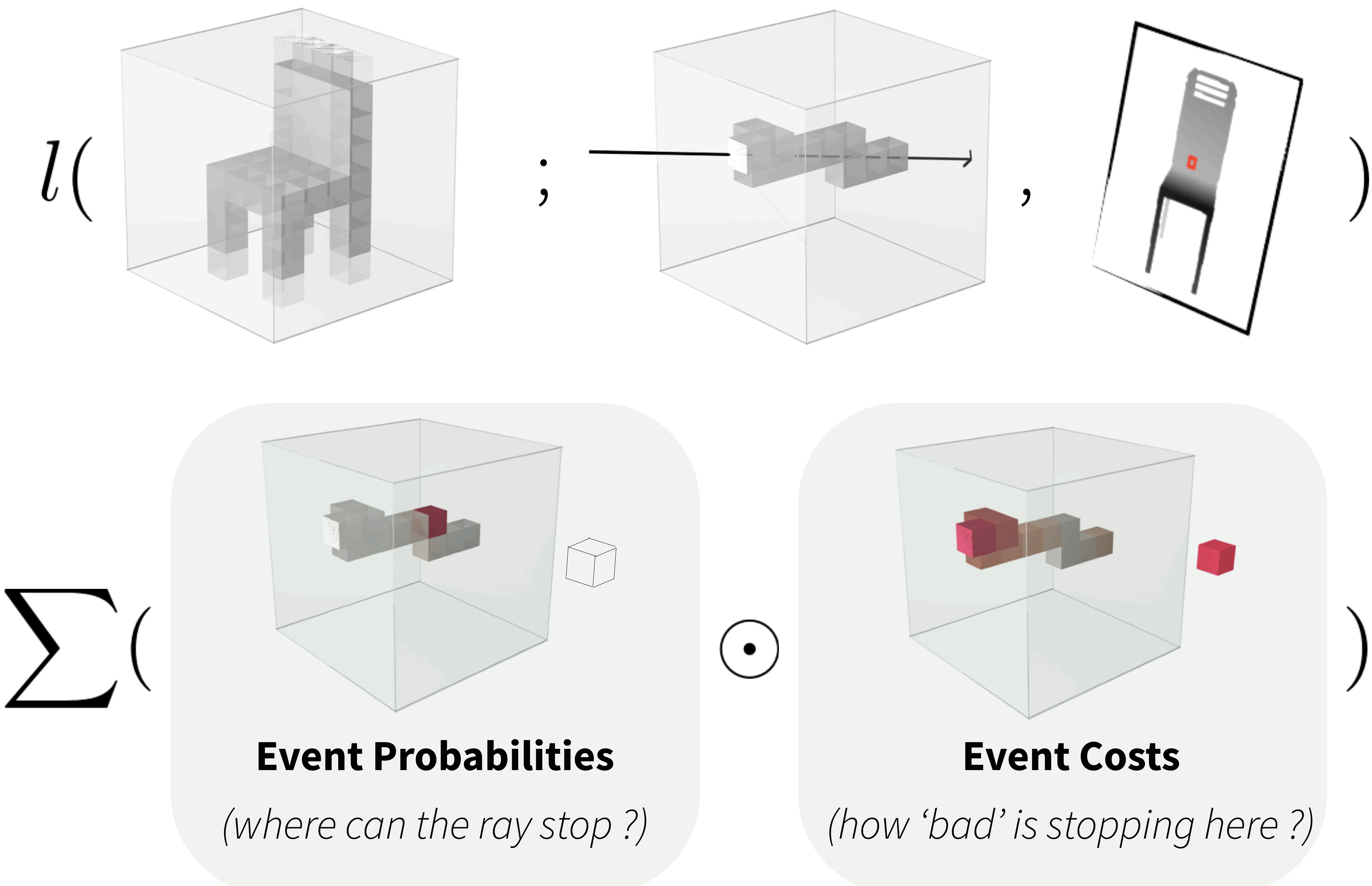
(where can the ray stop ?)



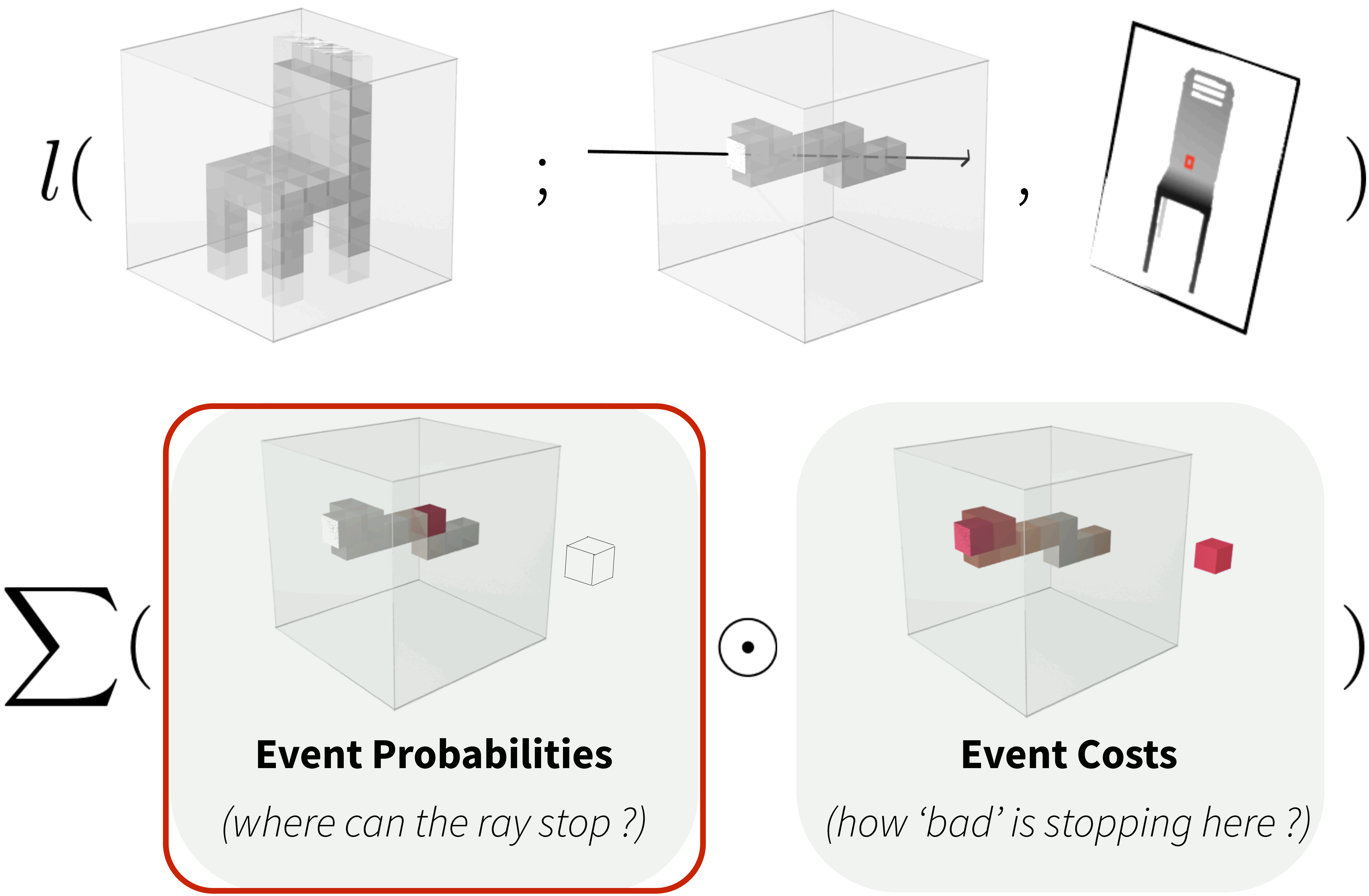
Event Costs

(how 'bad' is stopping here ?)

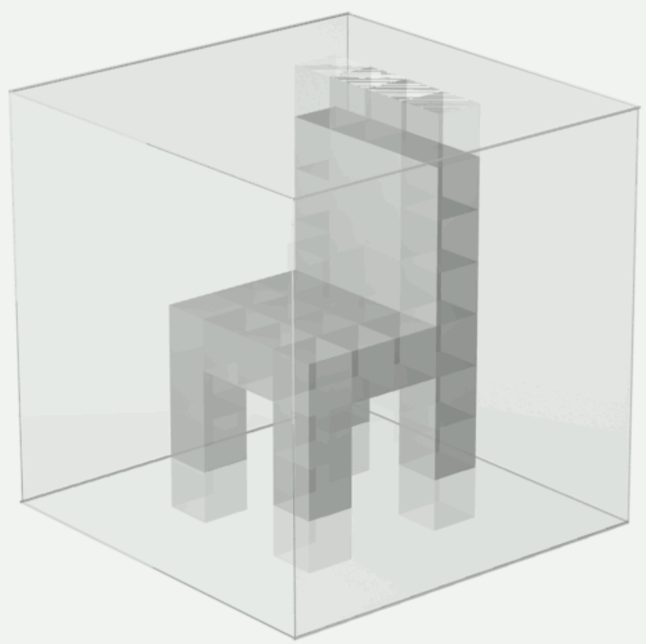
Differentiable Ray Consistency



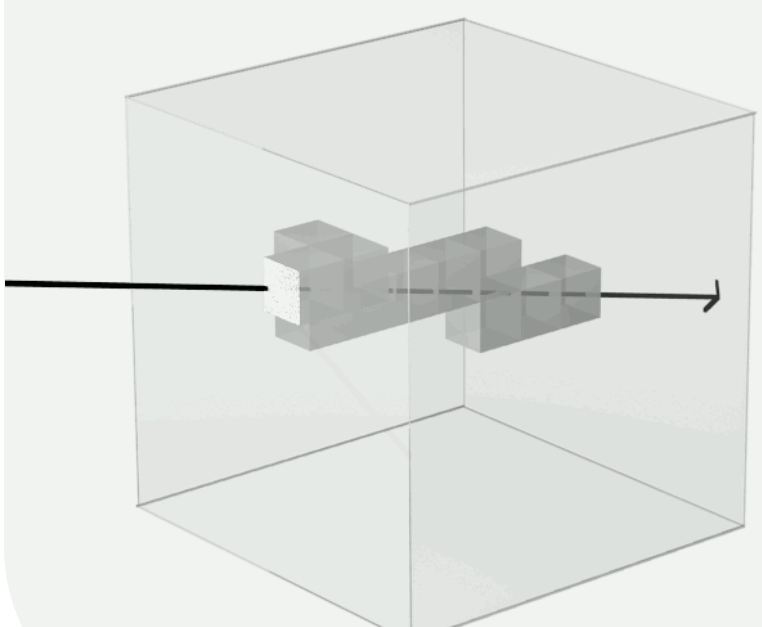
Differentiable Ray Consistency



Probabilistic Ray Tracing



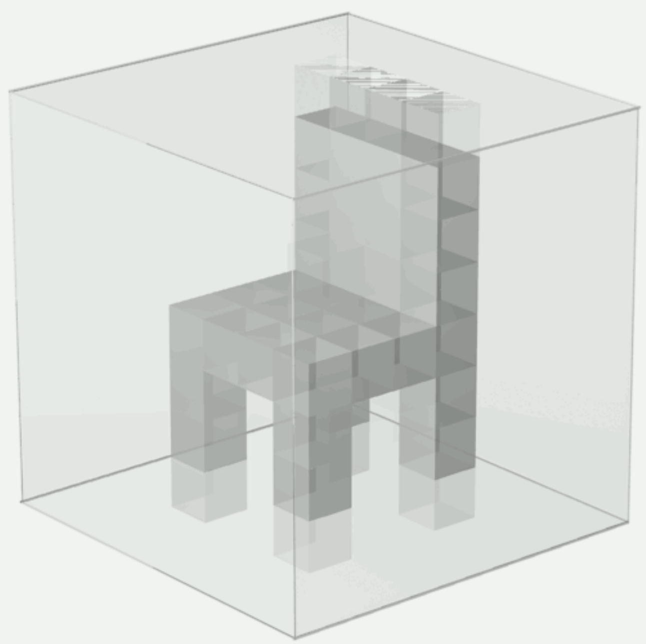
x



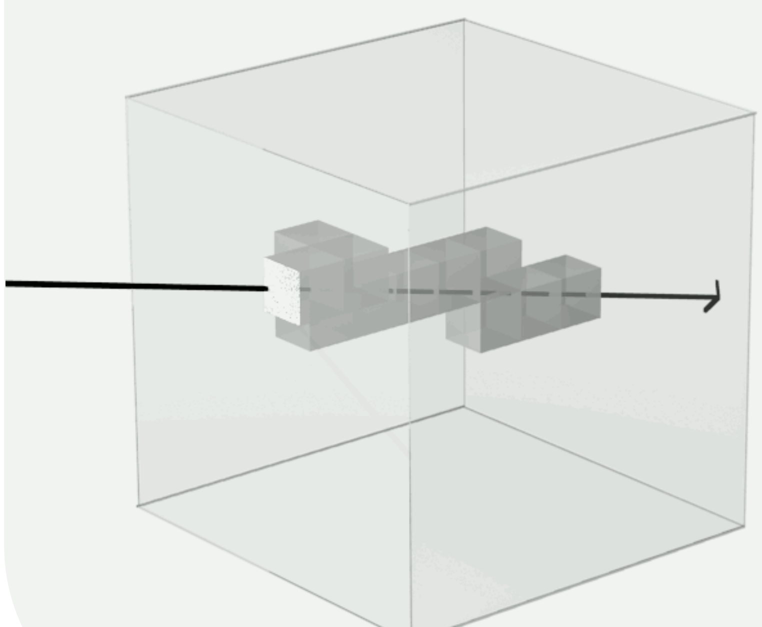
r

Possible Events

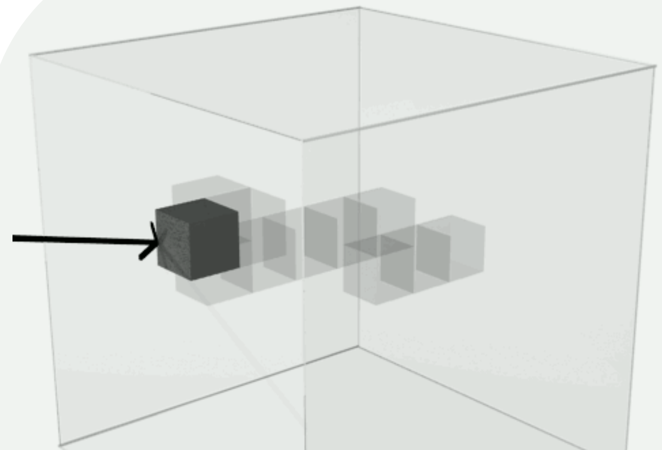
Probabilistic Ray Tracing



x



r

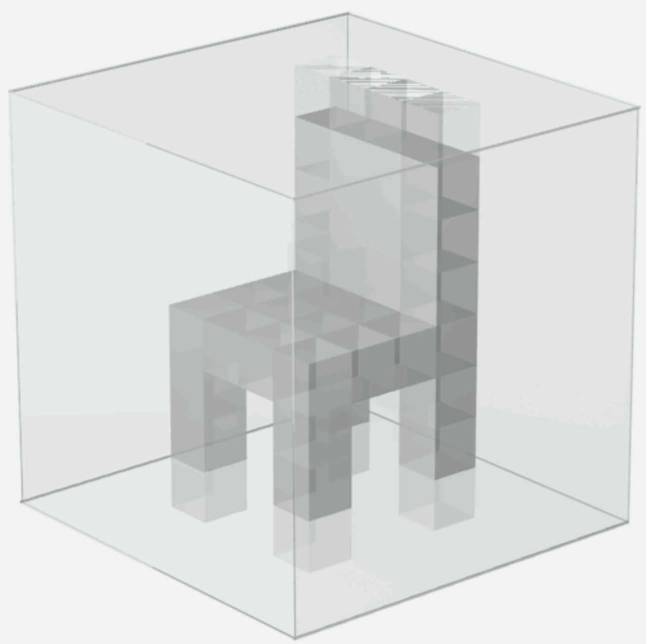


$$\text{Event}: \; z_r = 1$$

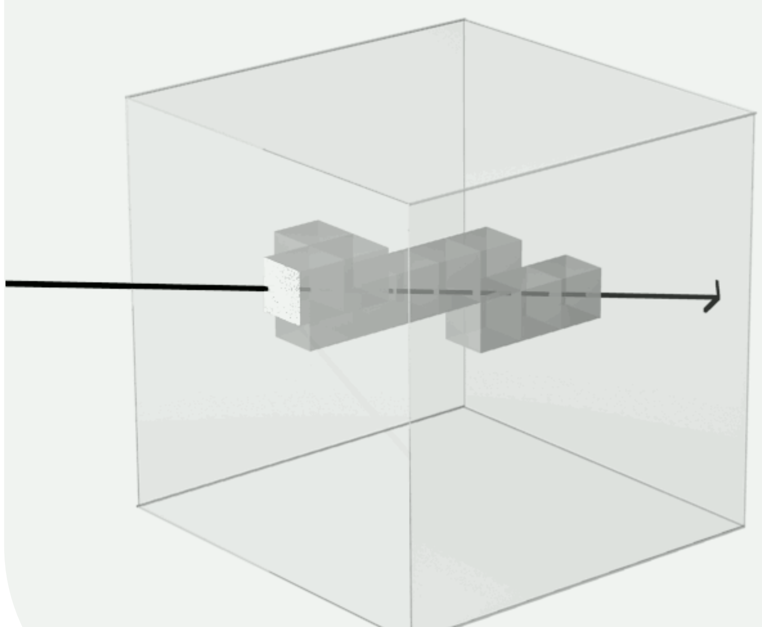
$$p(z_r = 1) = (1 - x_1^r)$$

Possible Events

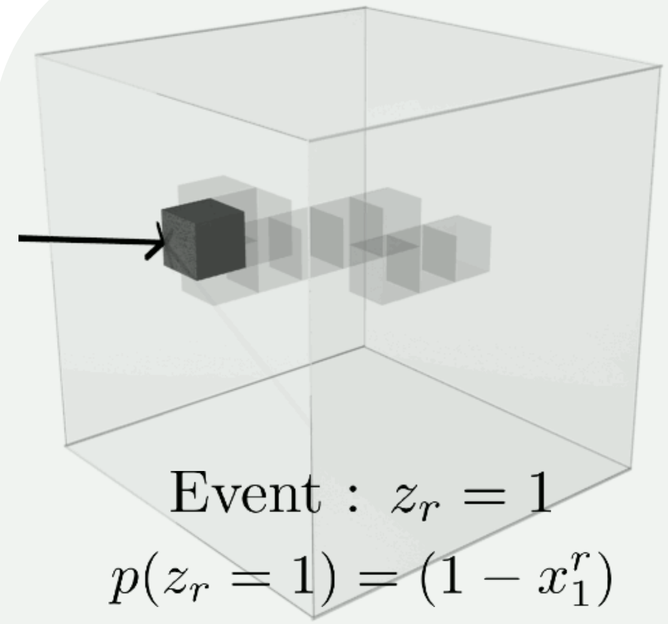
Probabilistic Ray Tracing

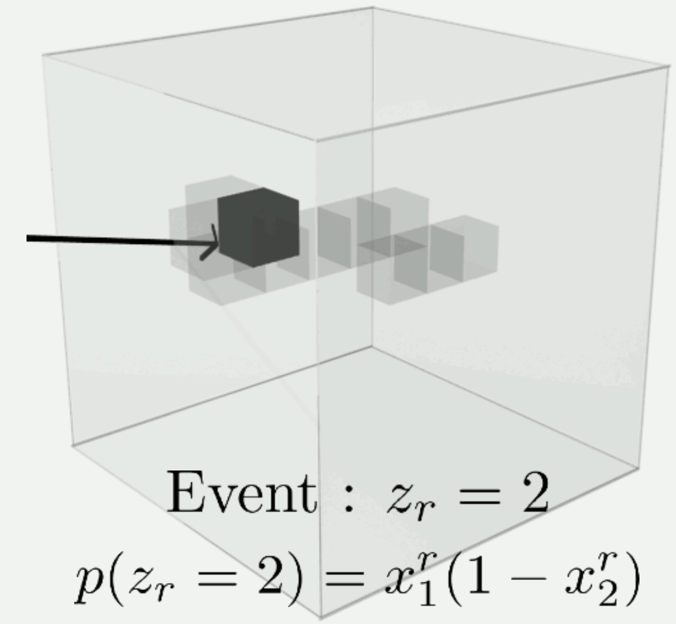


x



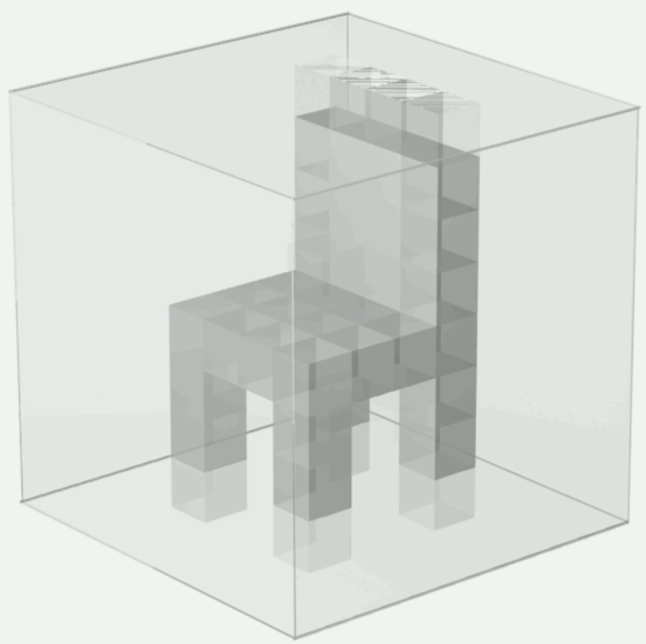
r



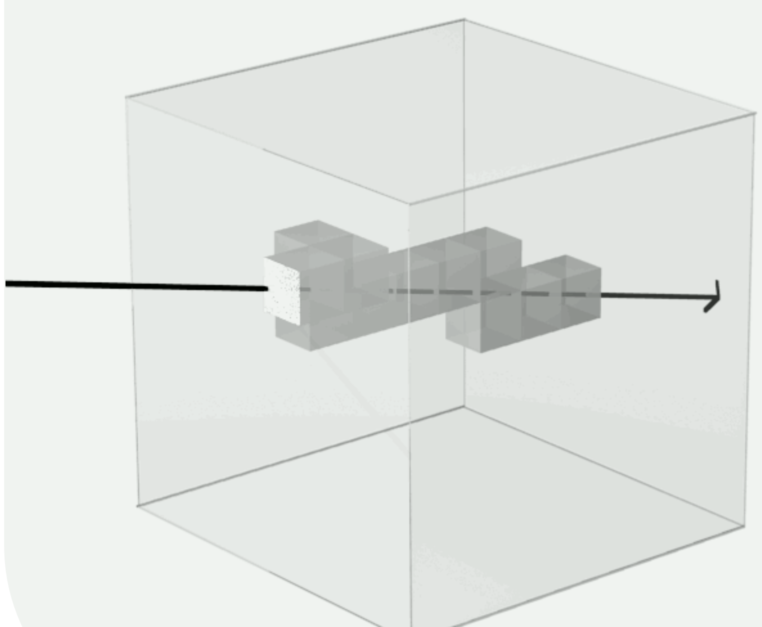


Possible Events

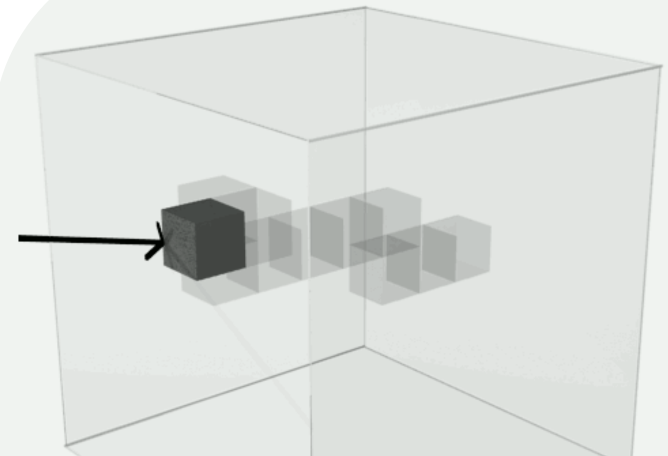
Probabilistic Ray Tracing



x

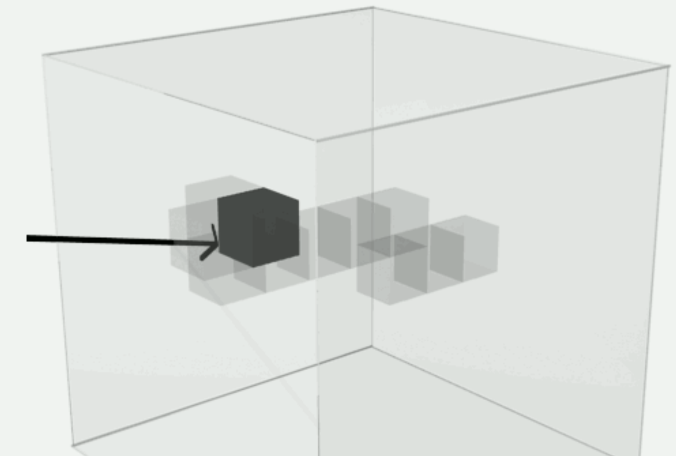


r



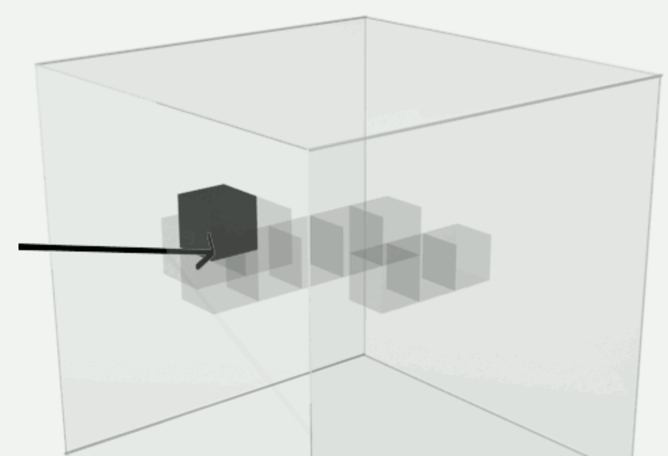
$\text{Event}: \ z_r = 1$

$p(z_r = 1) = (1 - x_1^r)$



$\text{Event}: \ z_r = 2$

$p(z_r = 2) = x_1^r(1 - x_2^r)$

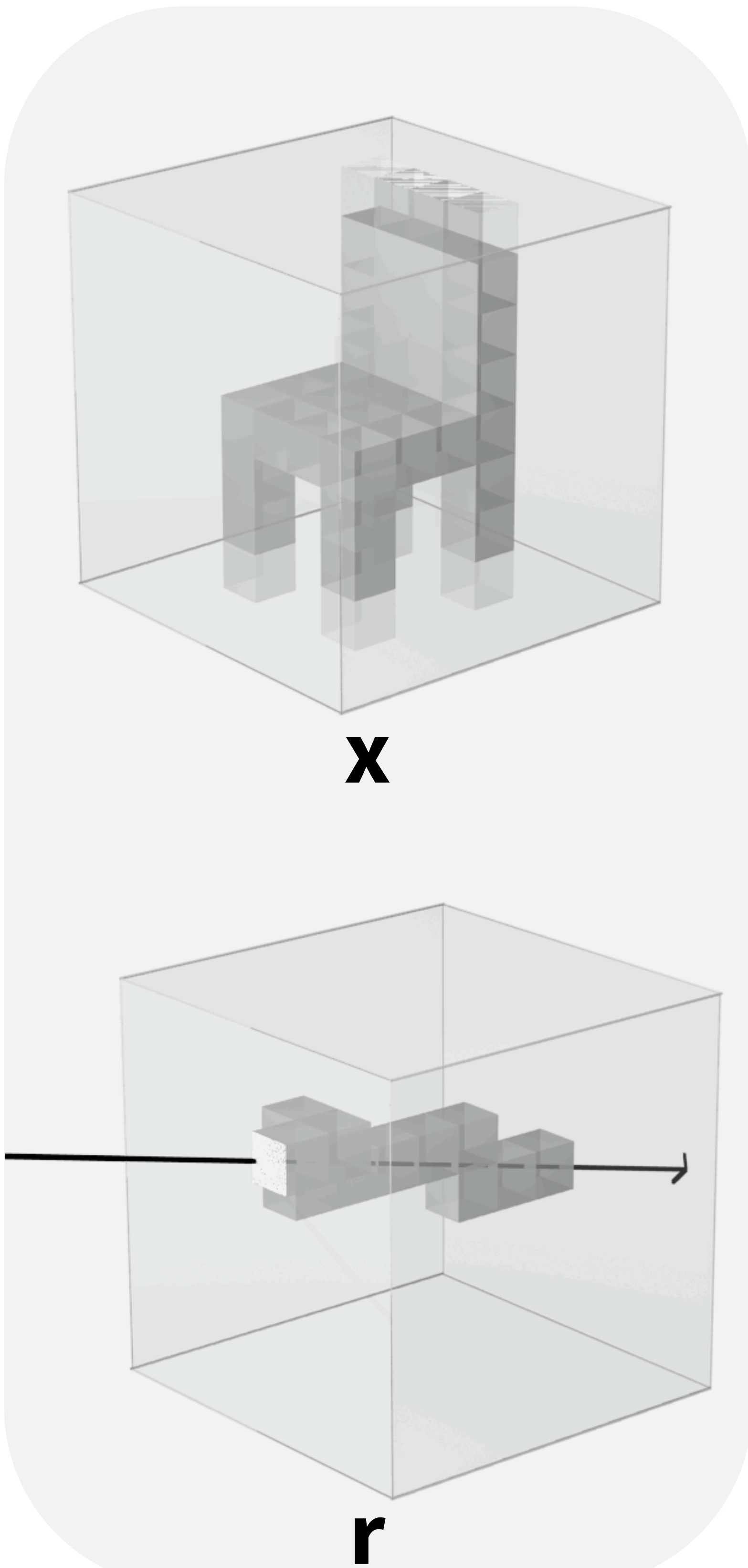


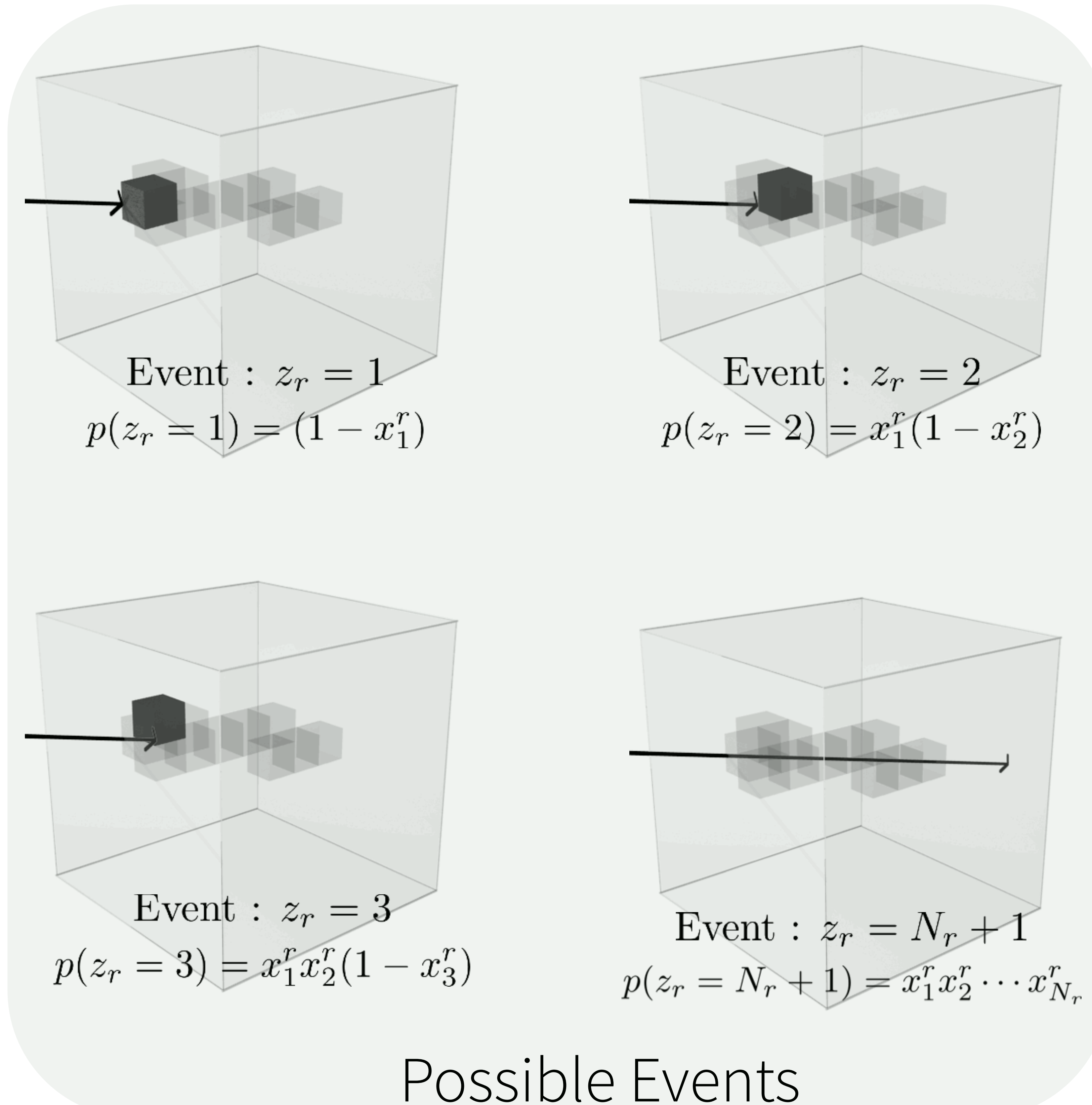
$\text{Event}: \ z_r = 3$

$p(z_r = 3) = x_1^r x_2^r(1 - x_3^r)$

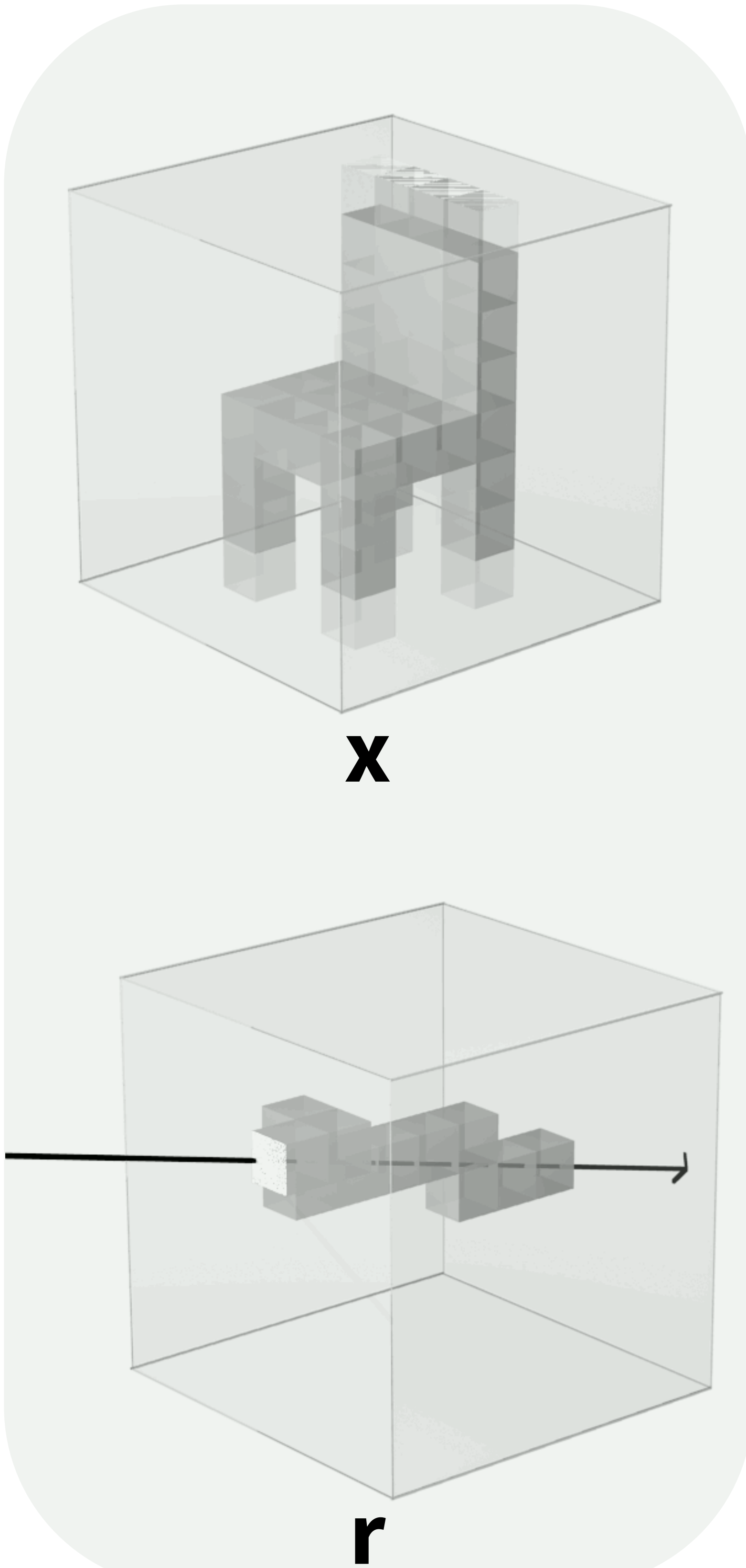
Possible Events

Probabilistic Ray Tracing



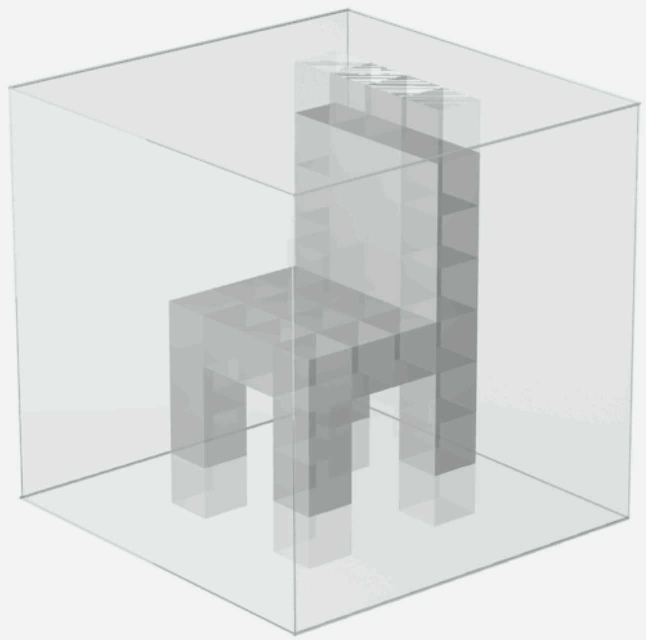


Probabilistic Ray Tracing

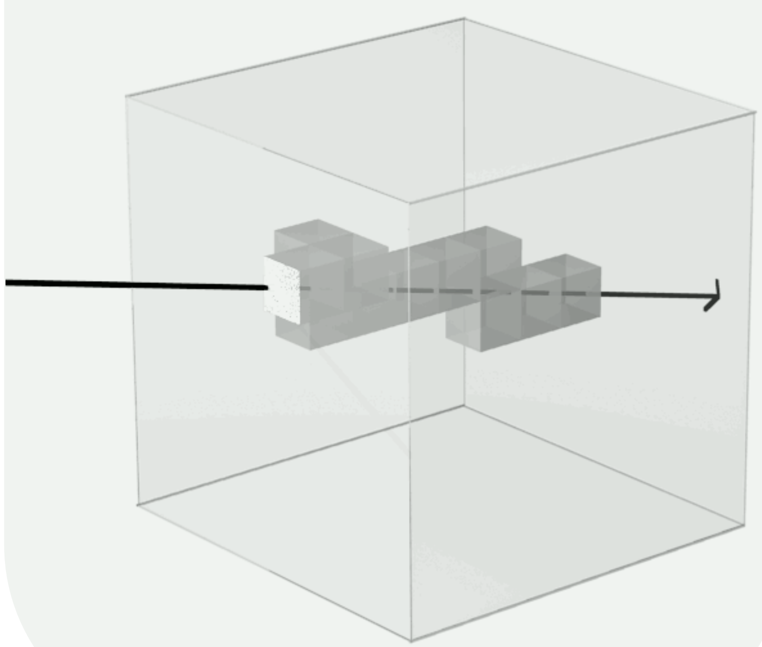


$$p(z_r = i) = \begin{cases} (1 - x_i^r) \prod_{j=1}^{i-1} x_j^r, & \text{if } i \leq N_r \\ \prod_{j=1}^{N_r} x_j^r, & \text{if } i = N_r + 1 \end{cases}$$

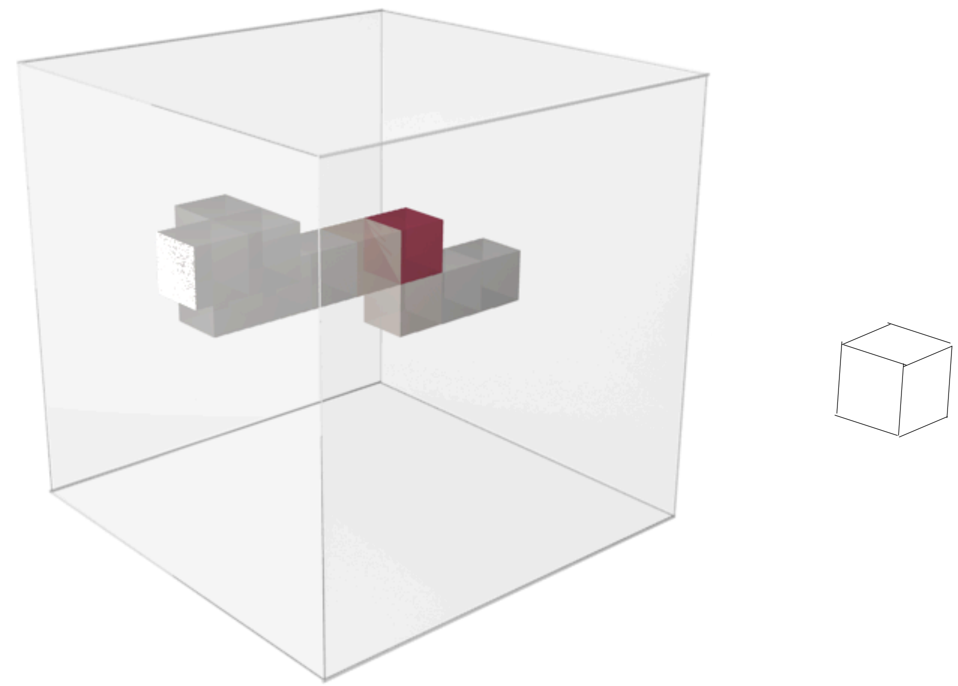
Probabilistic Ray Tracing



x

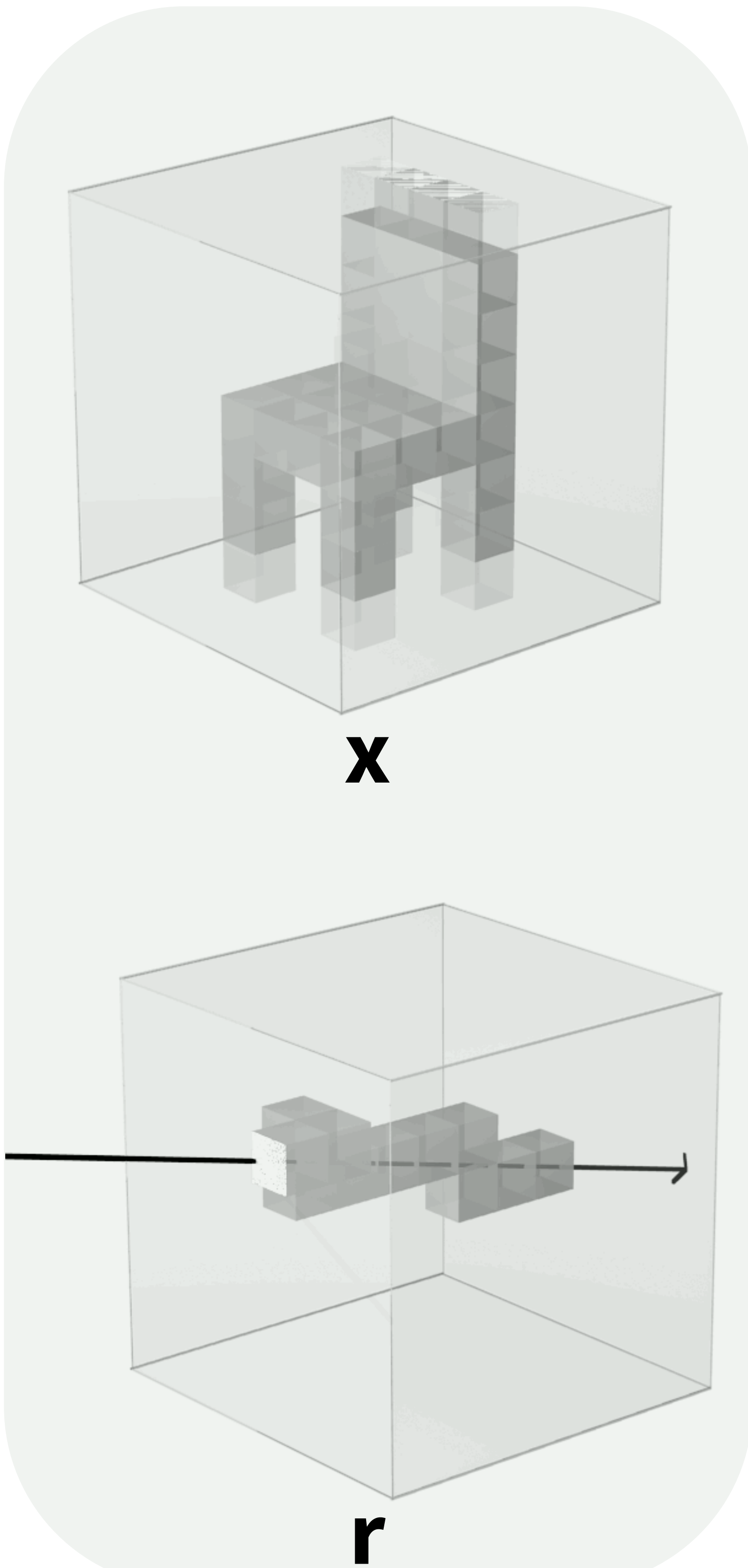


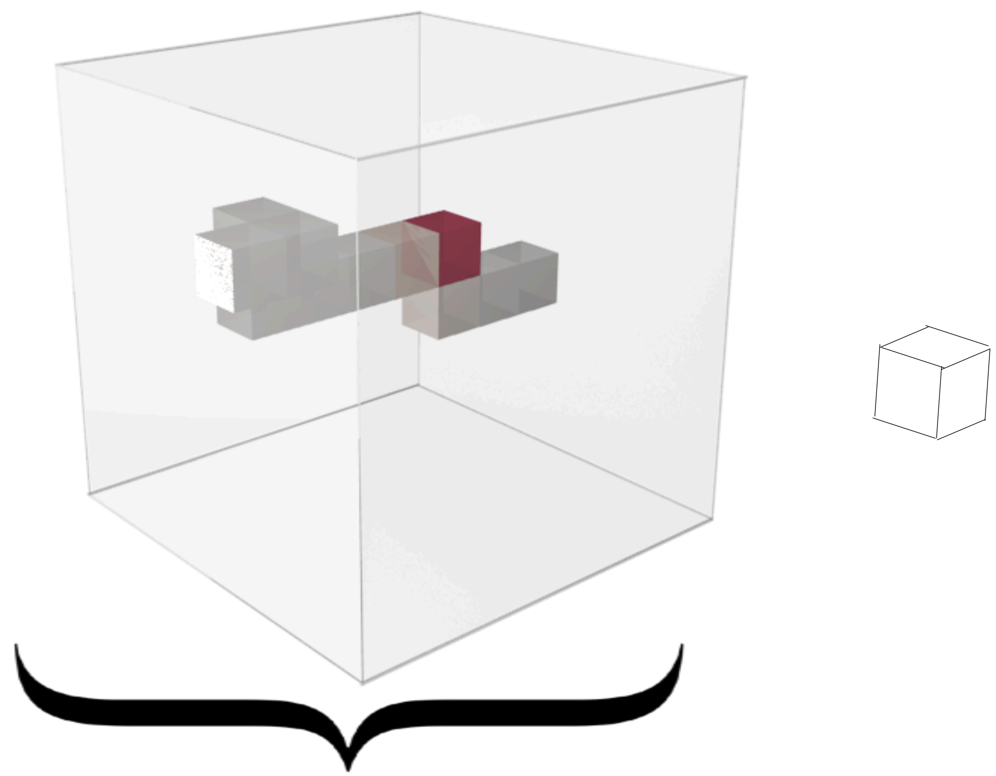
r



$$p(z_r = i) = \begin{cases} (1 - x_i^r) \prod_{j=1}^{i-1} x_j^r, & \text{if } i \leq N_r \\ \prod_{j=1}^{N_r} x_j^r, & \text{if } i = N_r + 1 \end{cases}$$

Probabilistic Ray Tracing

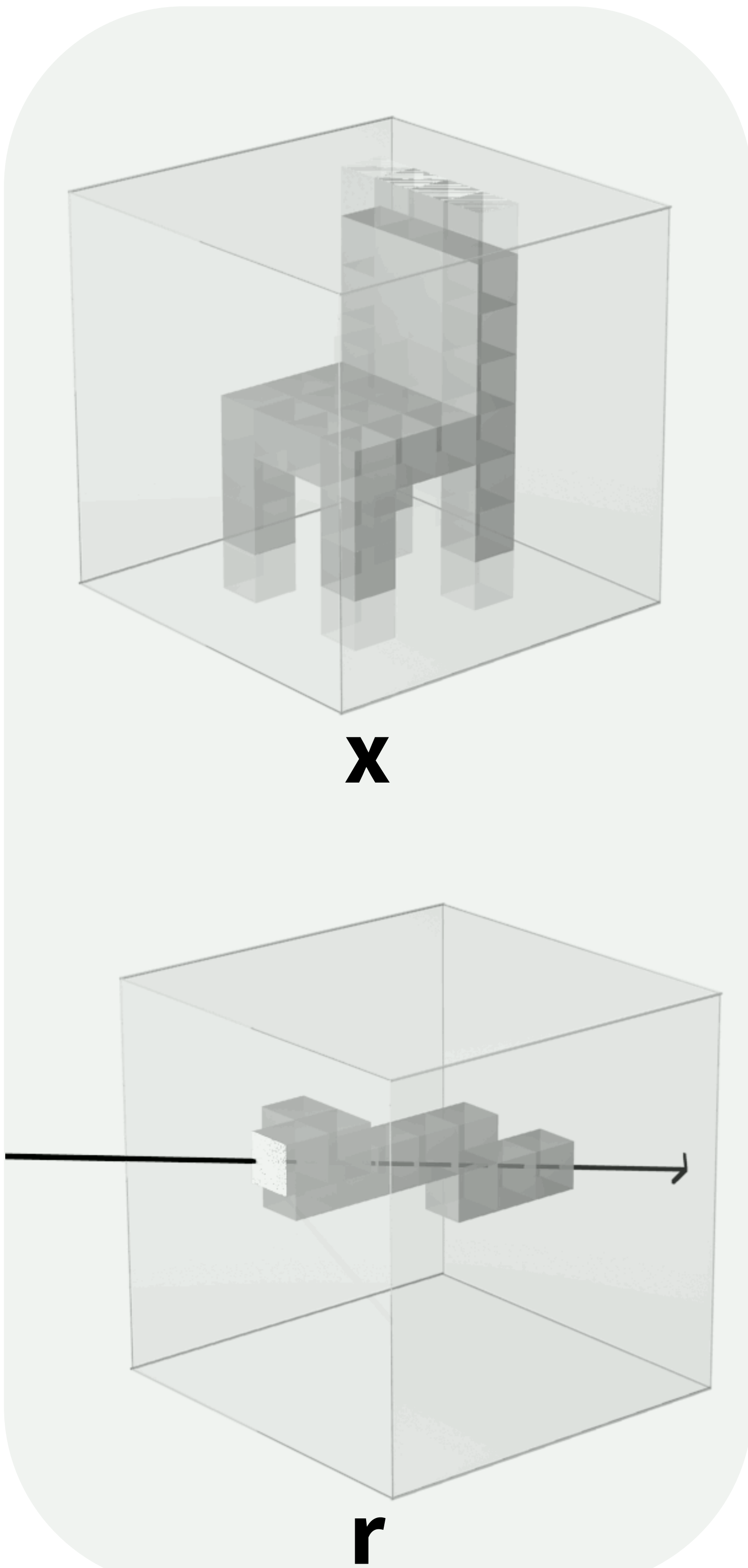


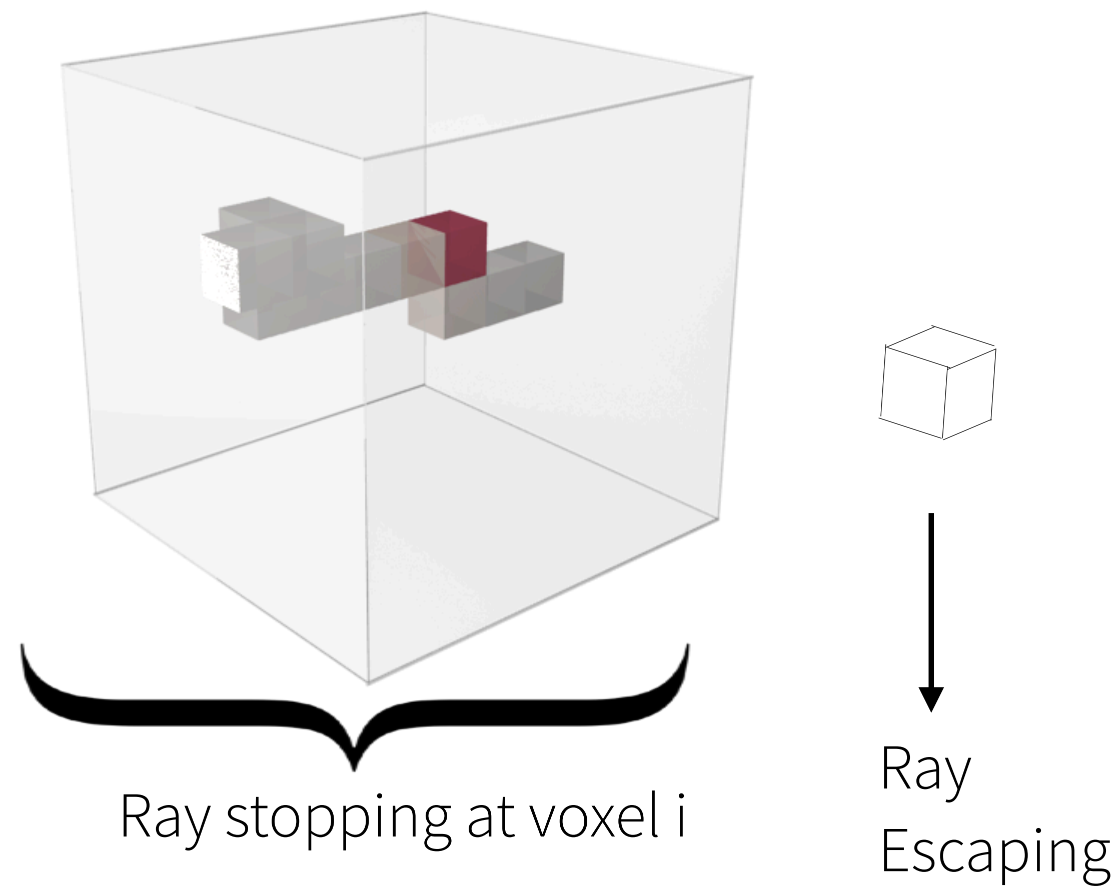


Ray stopping at voxel i

$$p(z_r = i) = \begin{cases} (1 - x_i^r) \prod_{j=1}^{i-1} x_j^r, & \text{if } i \leq N_r \\ \prod_{j=1}^{N_r} x_j^r, & \text{if } i = N_r + 1 \end{cases}$$

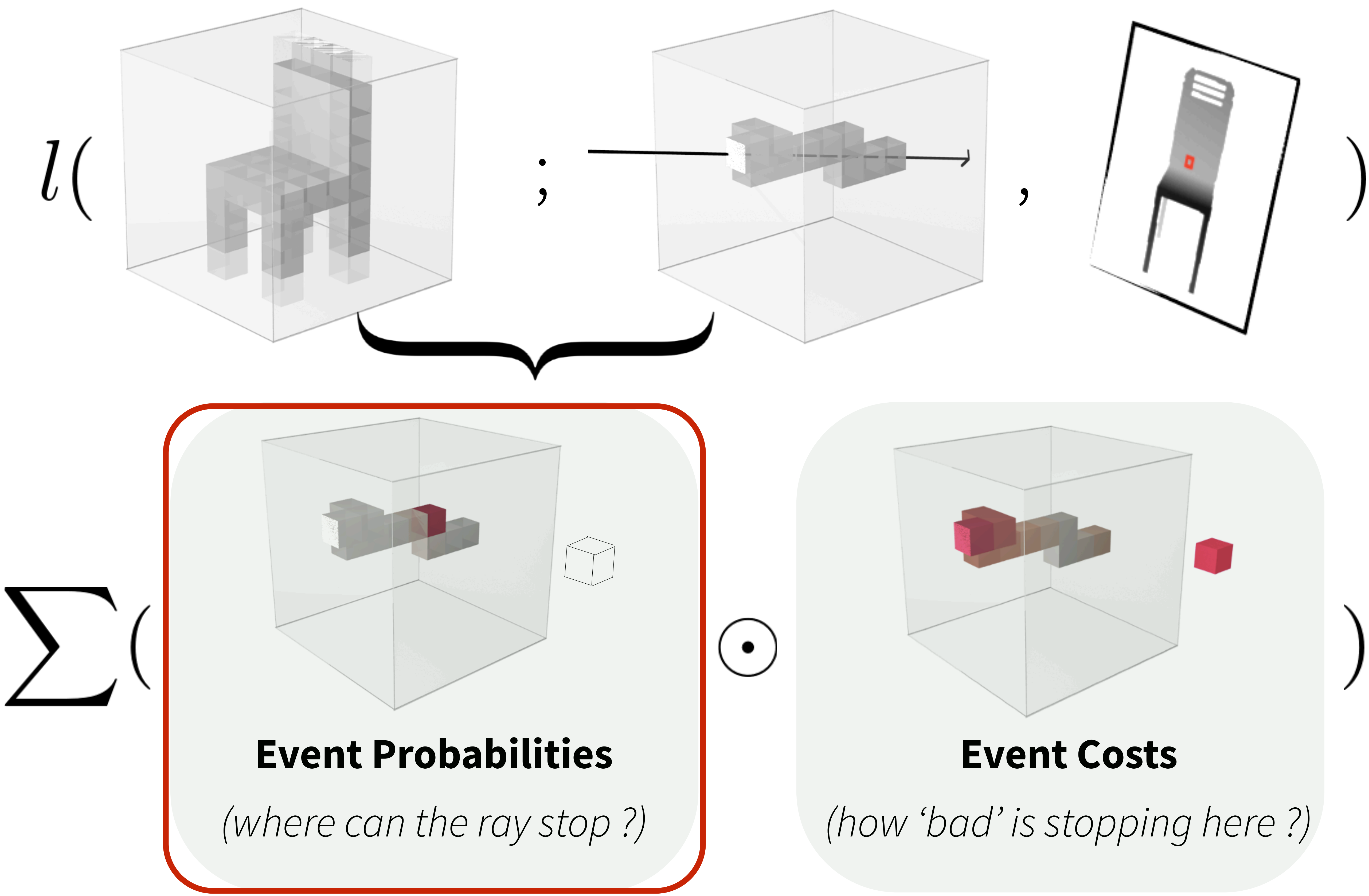
Probabilistic Ray Tracing



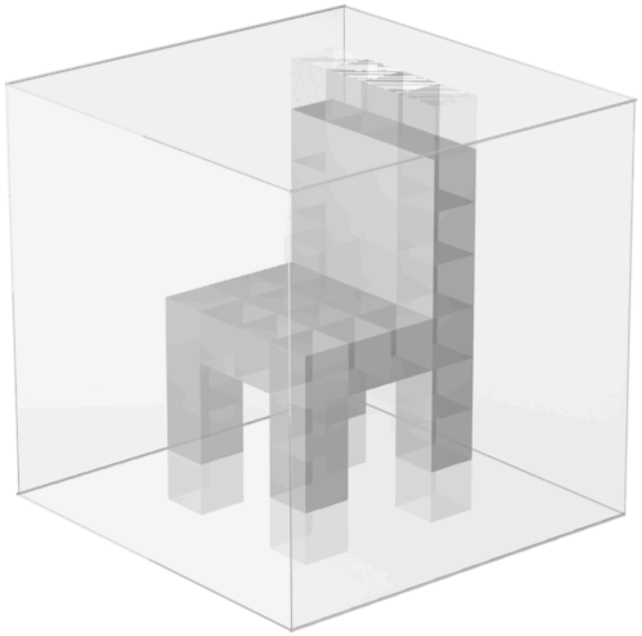
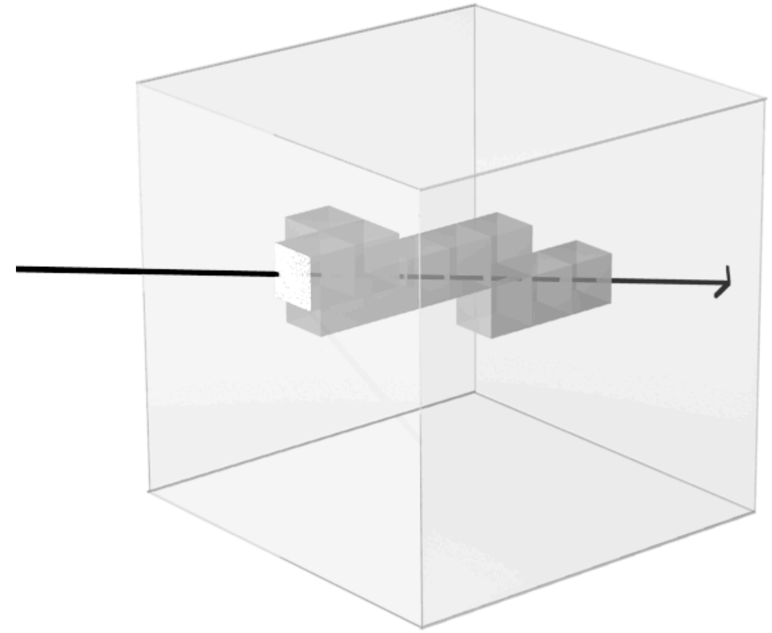
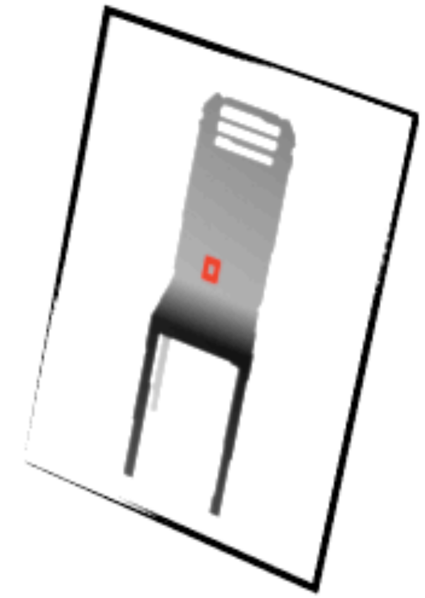


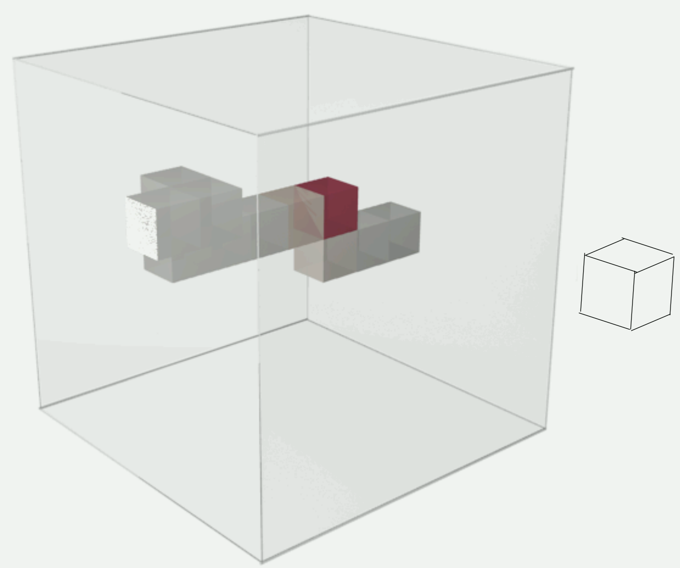
$$p(z_r = i) = \begin{cases} (1 - x_i^r) \prod_{j=1}^{i-1} x_j^r, & \text{if } i \leq N_r \\ \prod_{j=1}^{N_r} x_j^r, & \text{if } i = N_r + 1 \end{cases}$$

Differentiable Ray Consistency



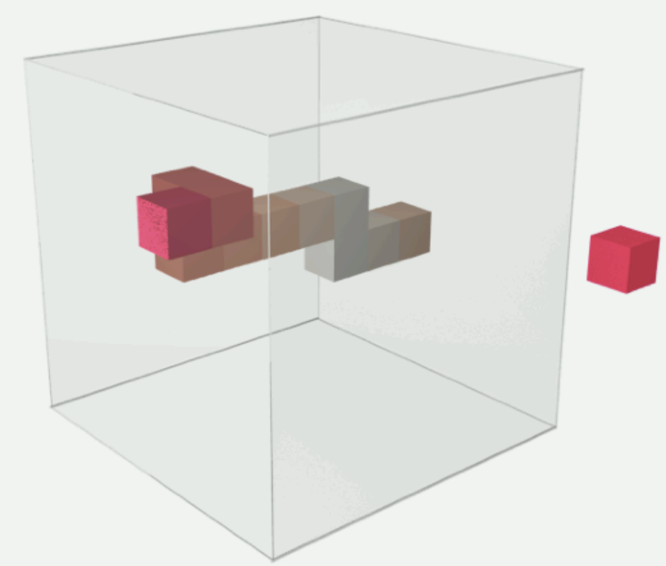
Differentiable Ray Consistency

$l($  $;$  $,$  $)$

$\sum($ 

Event Probabilities

(where can the ray stop ?)

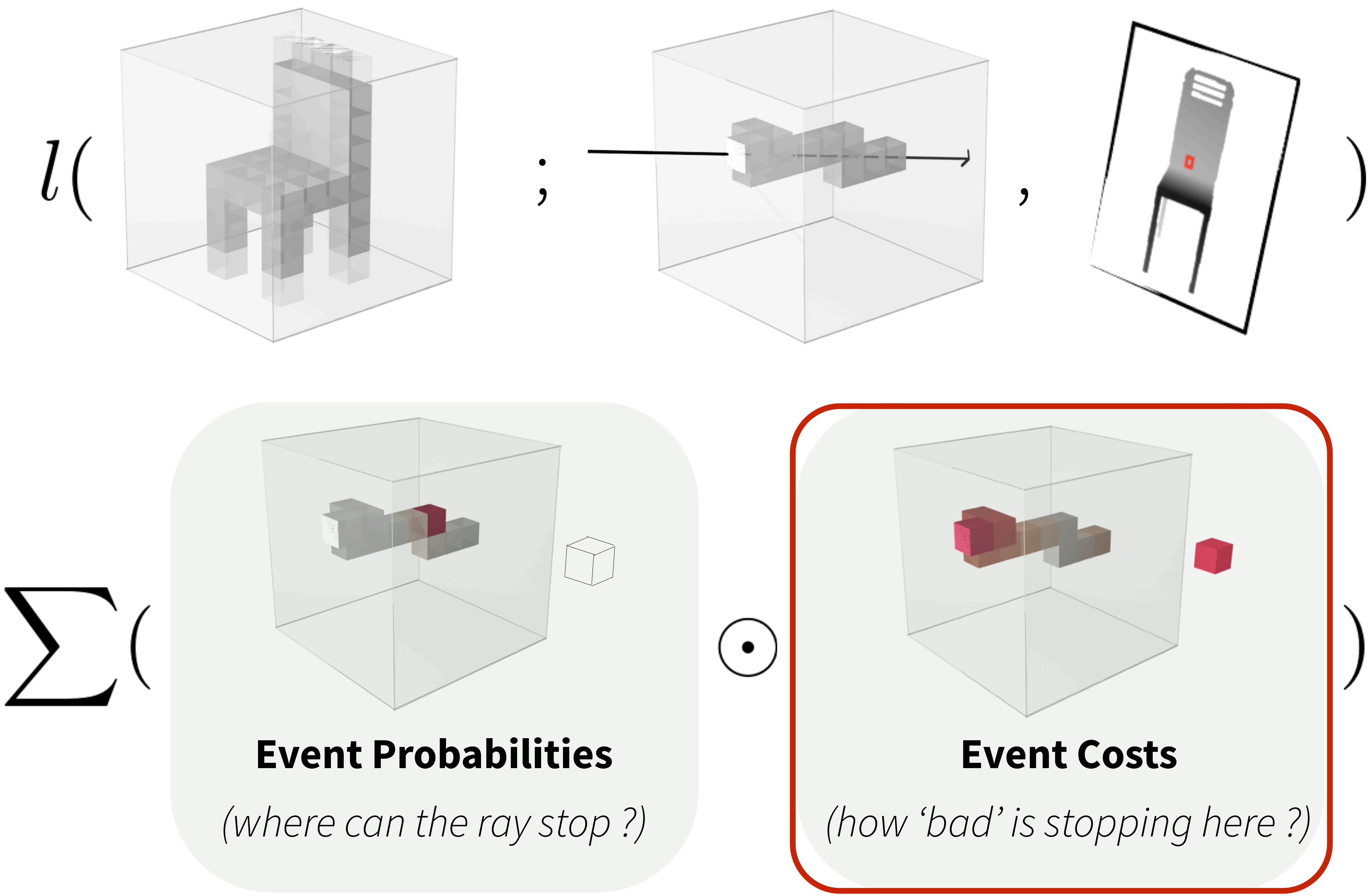
$\odot$ 

Event Costs

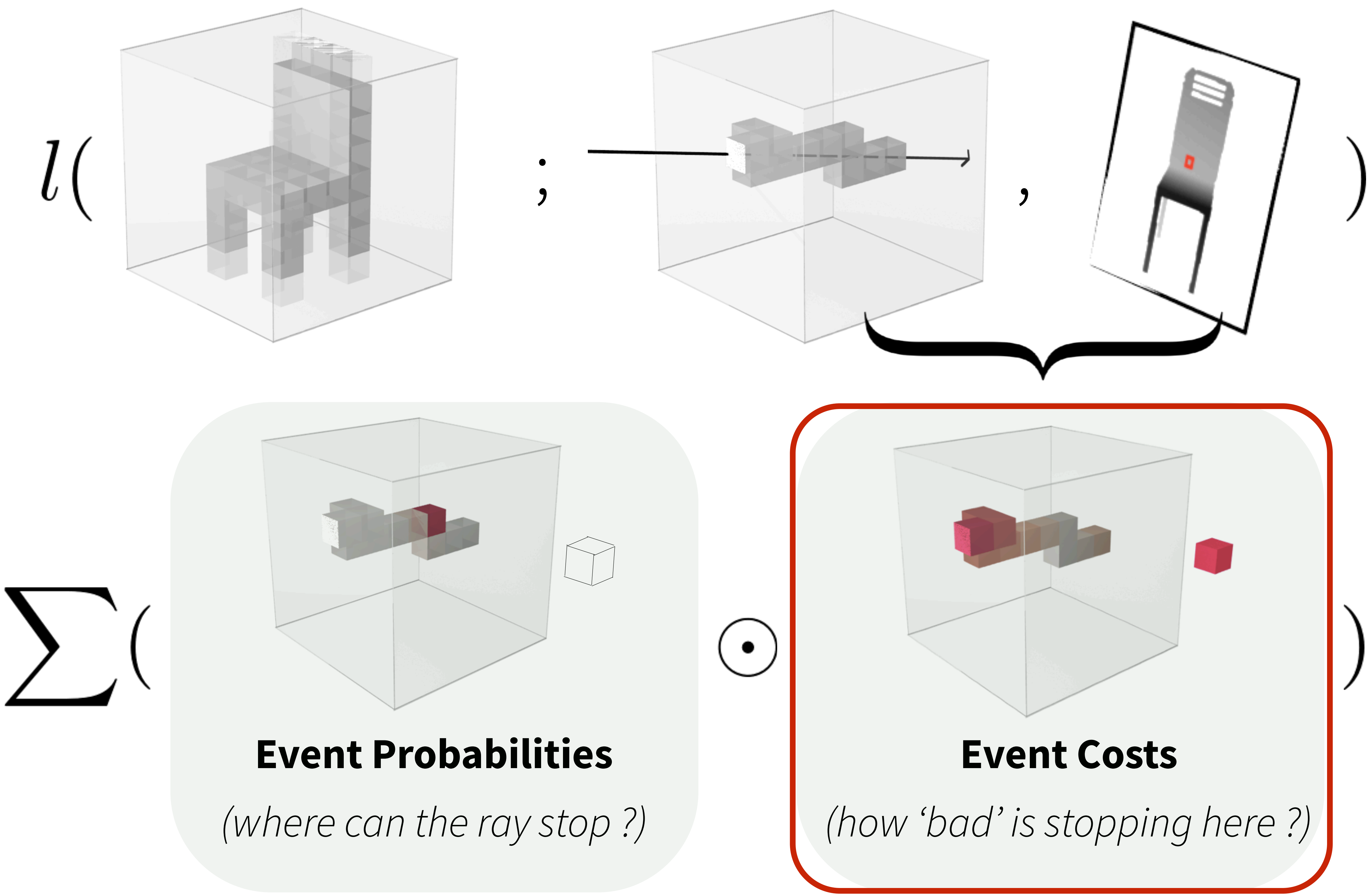
(how 'bad' is stopping here ?)

$)$

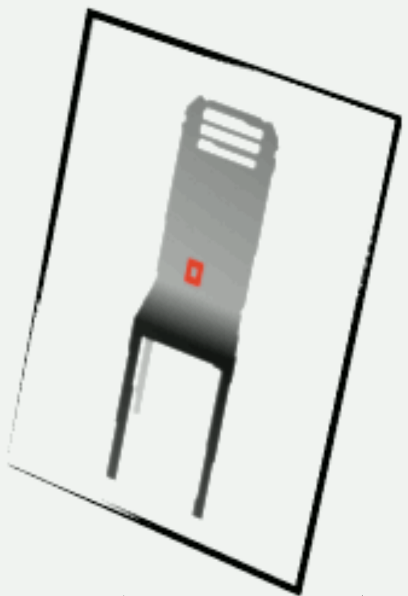
Differentiable Ray Consistency



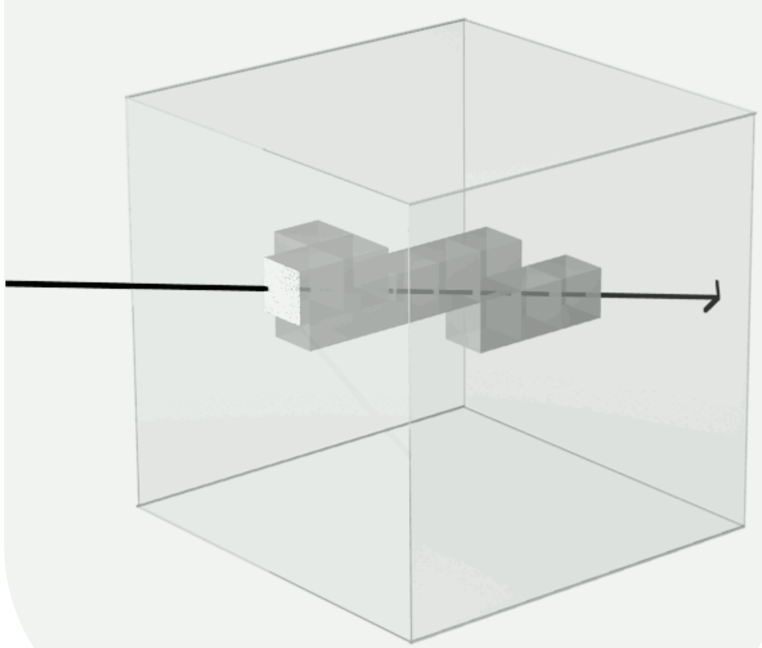
Differentiable Ray Consistency



Event Costs



$\mathbf{o_r}$ *(Depth)*

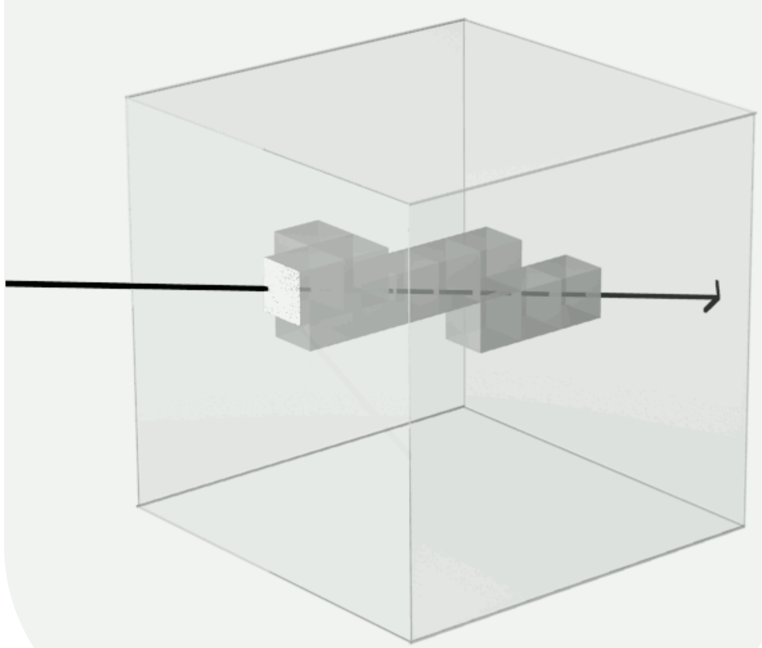


$\mathbf{r}$

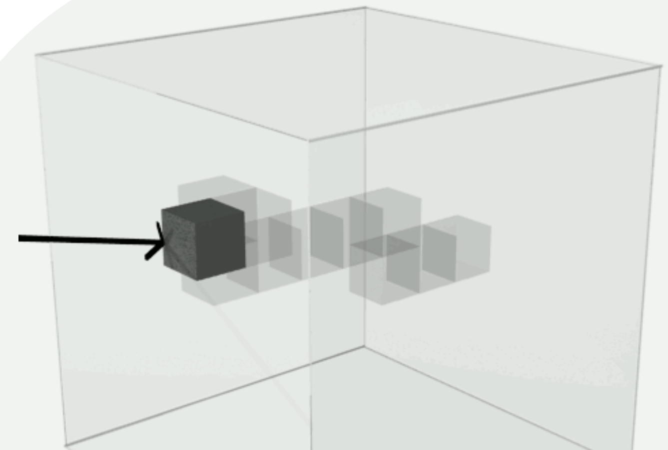
How inconsistent is each event w.r.t $\mathbf{o_r}$?

Event Costs

$\mathbf{o_r}$ *(Depth)*



r



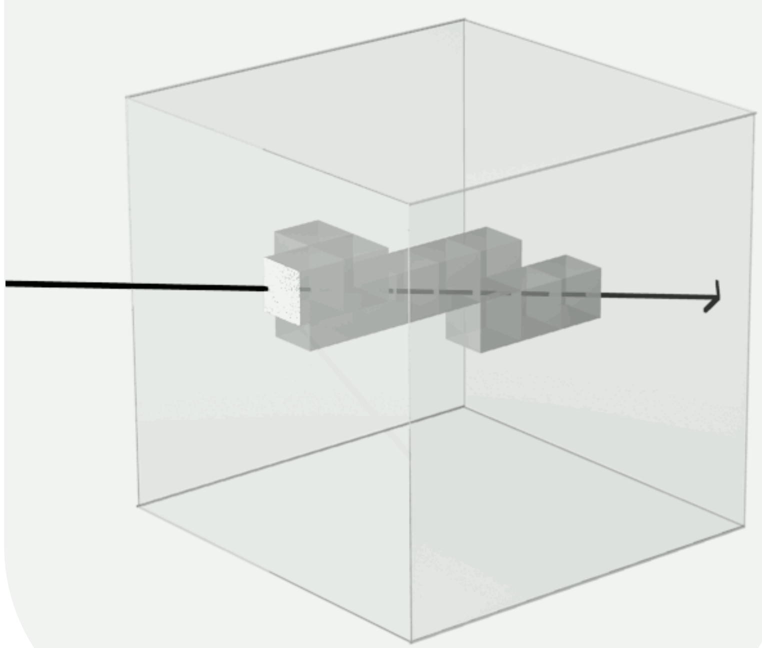
$\text{Event}: \; z_r = 1$

High cost

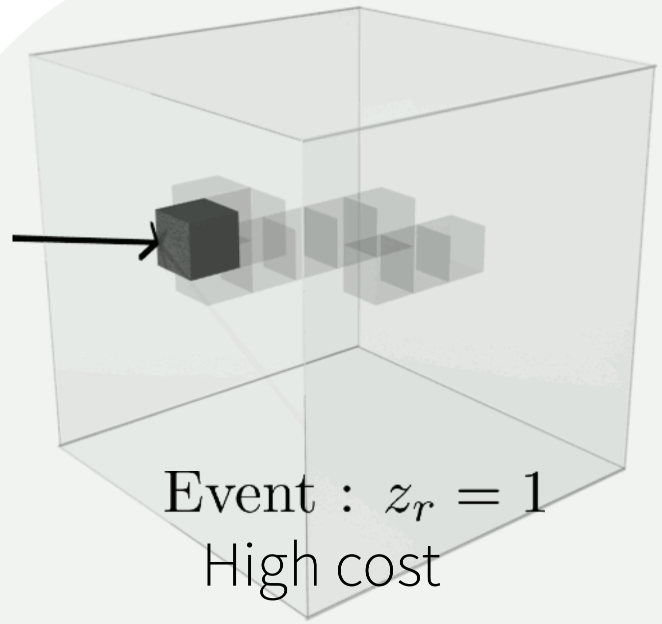
How inconsistent is each event w.r.t $\mathbf{o_r}$?

Event Costs

$\mathbf{o_r}$ *(Depth)*

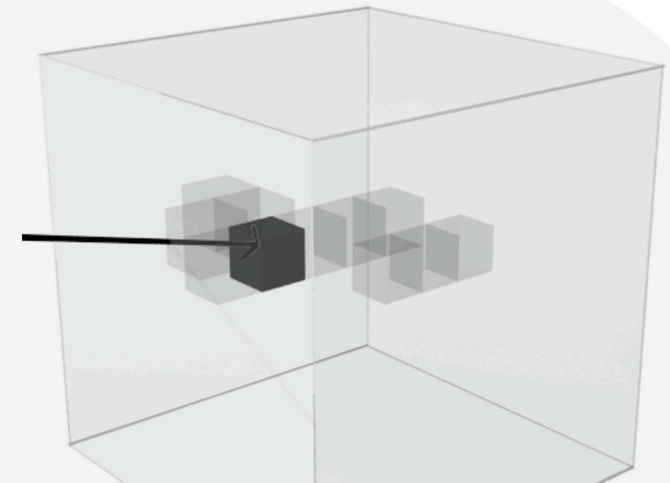


r



Event : $z_r = 1$

High cost



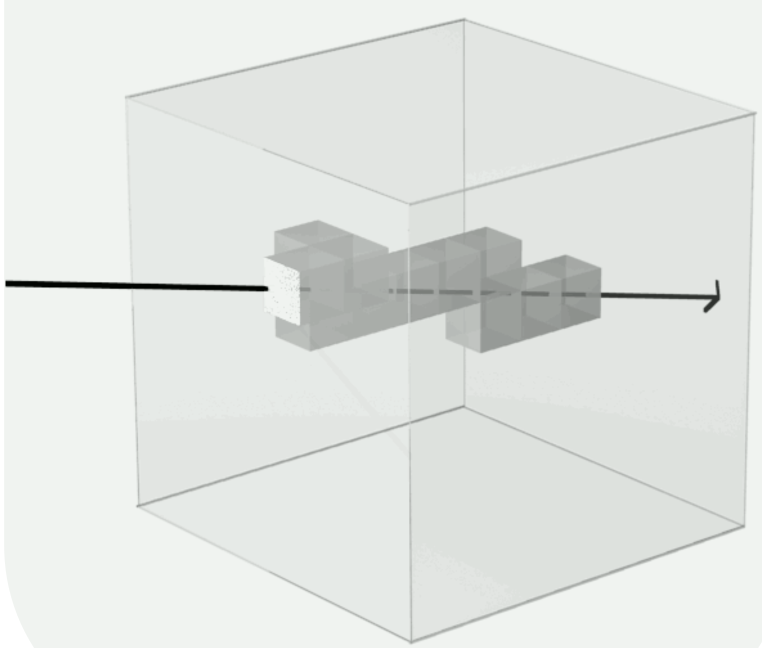
Event : $z_r = 5$

Medium Cost

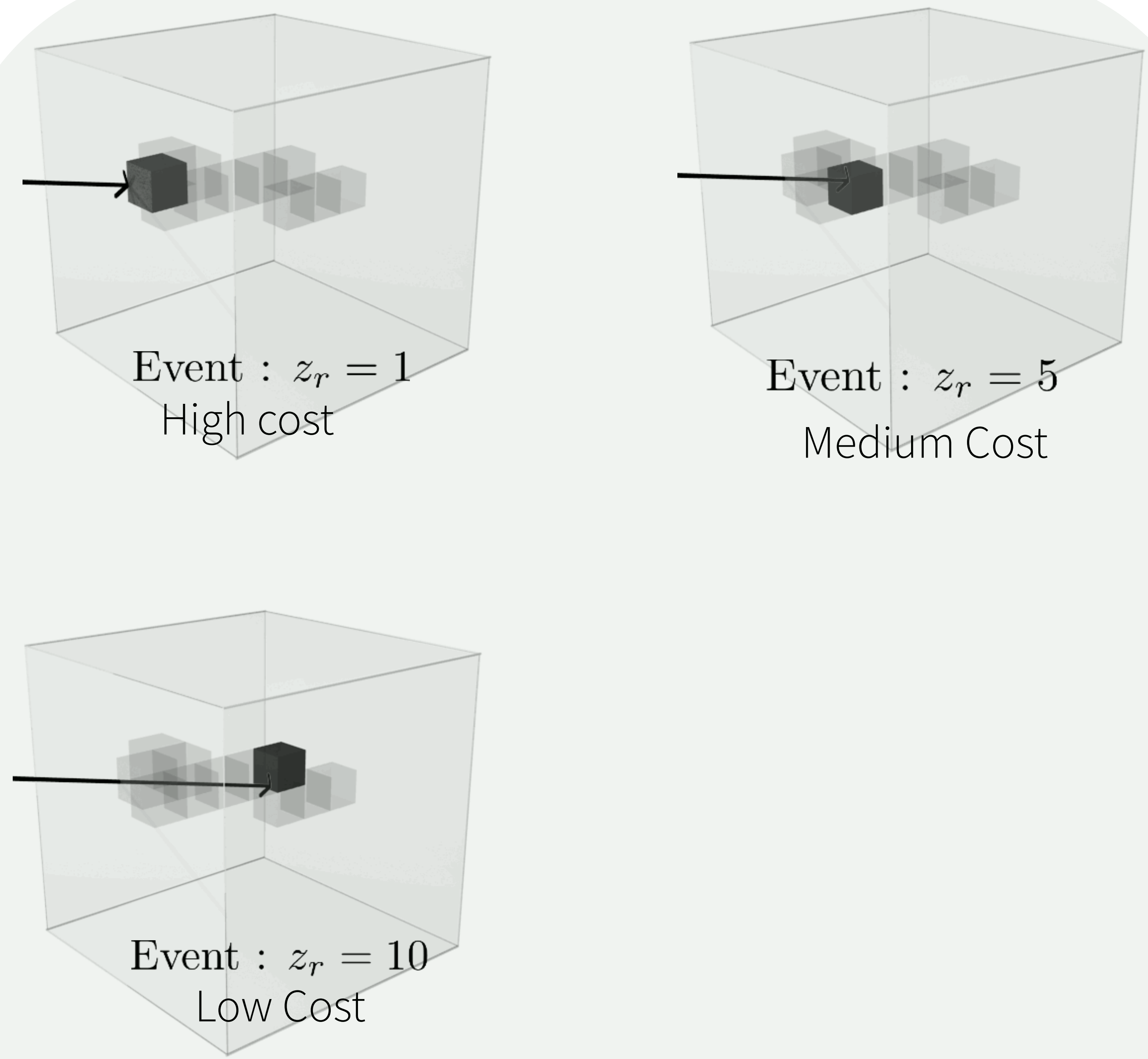
How inconsistent is each event w.r.t $\mathbf{o_r}$?

Event Costs

$\mathbf{o_r}$ *(Depth)*

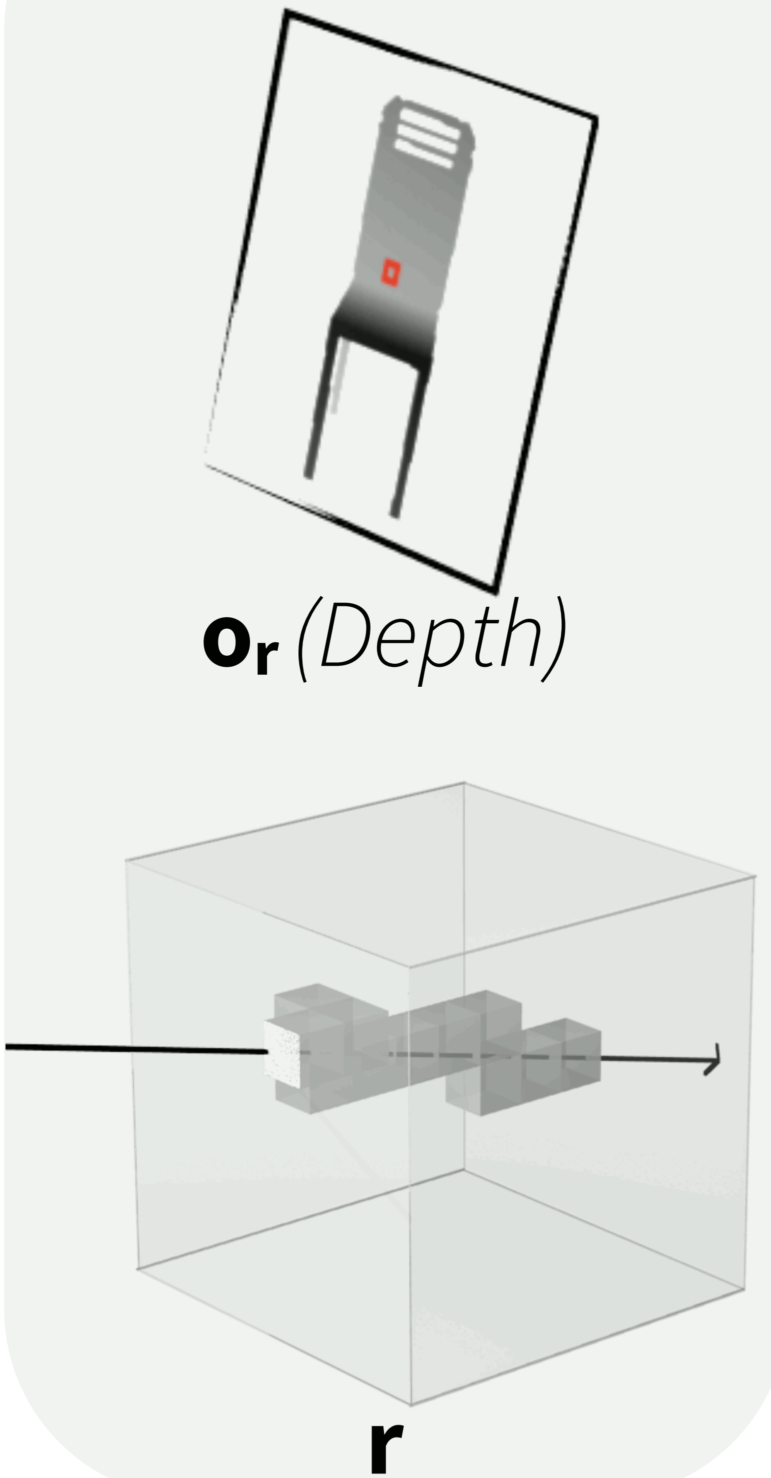


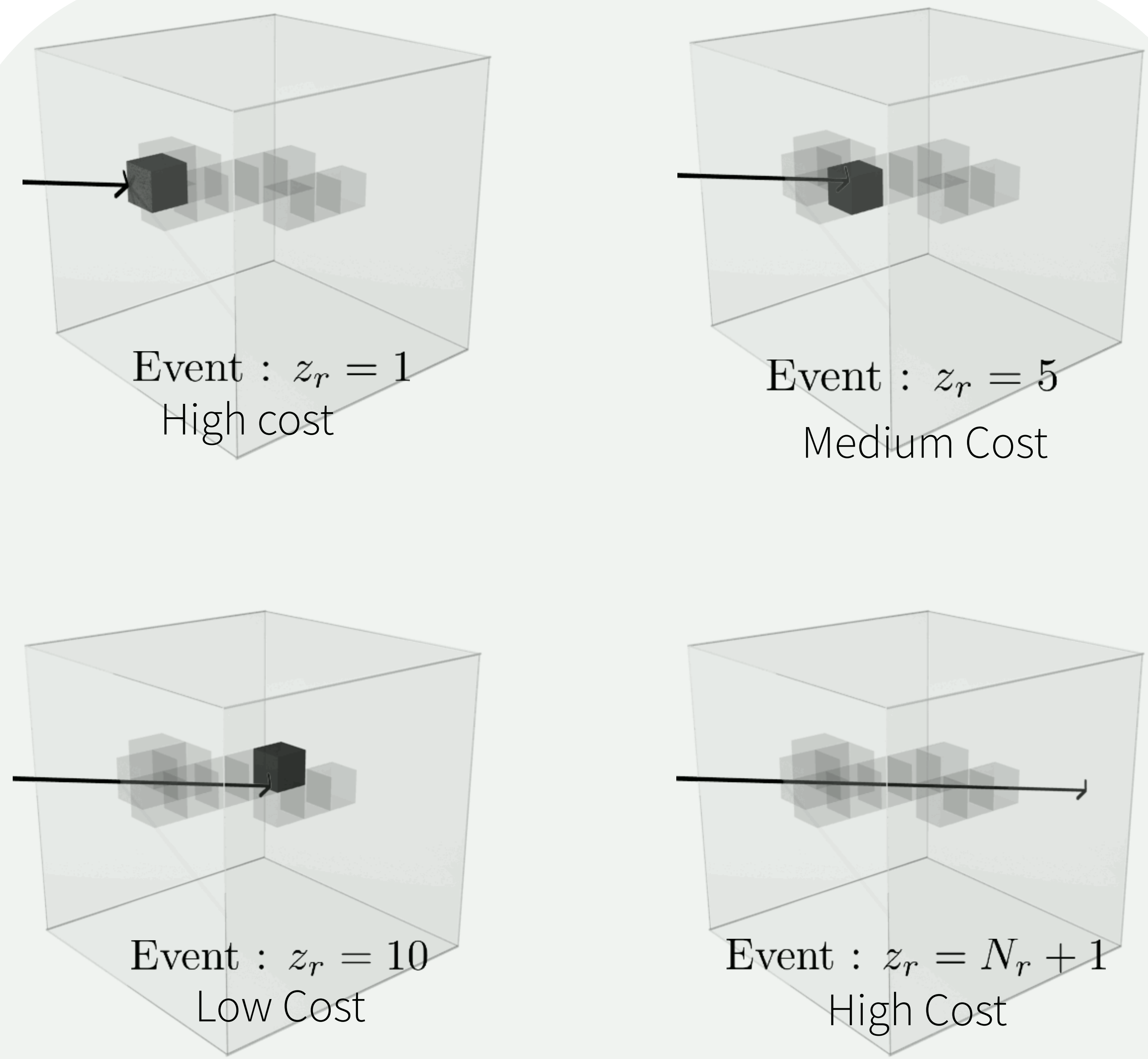
r



How inconsistent is each event w.r.t $\mathbf{o_r}$?

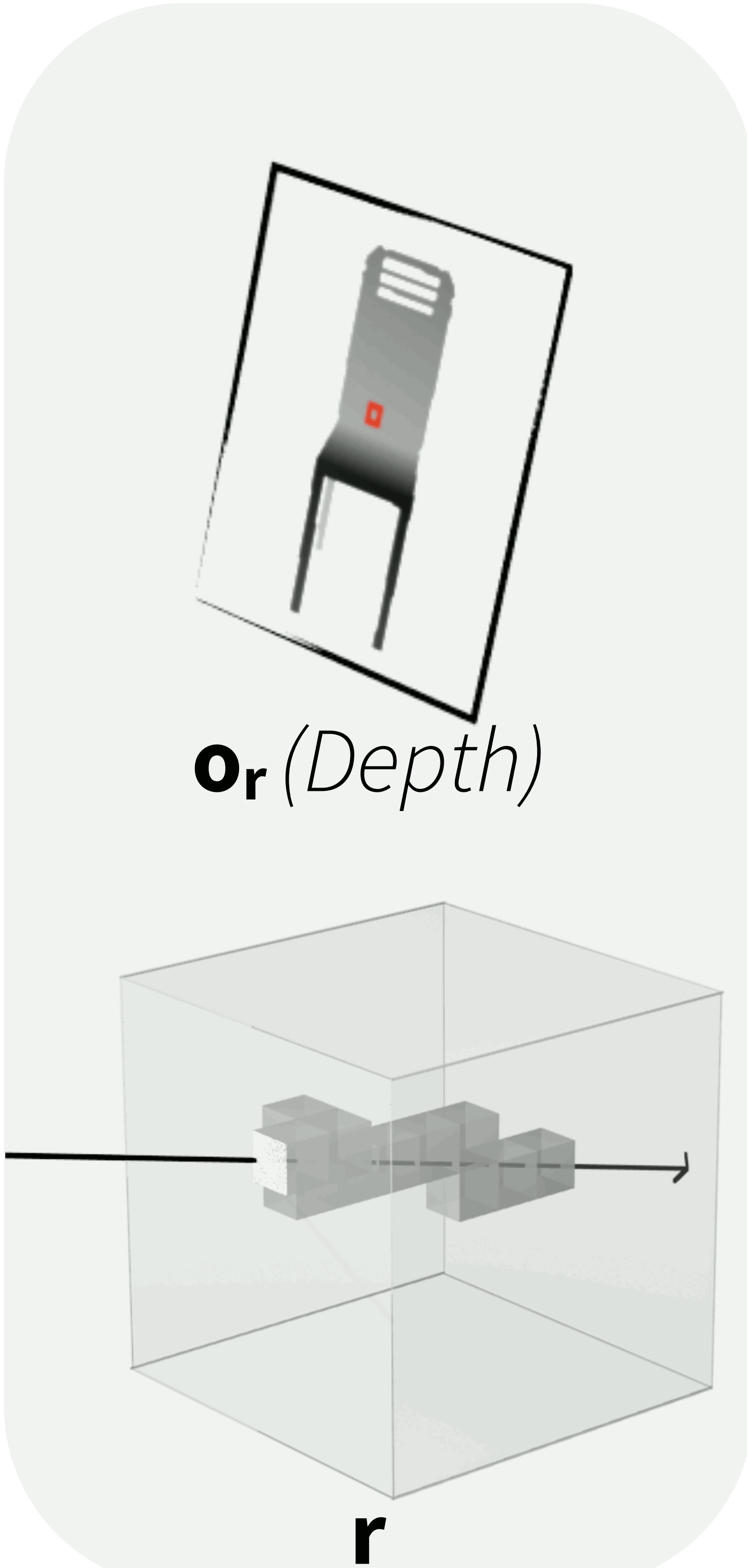
Event Costs





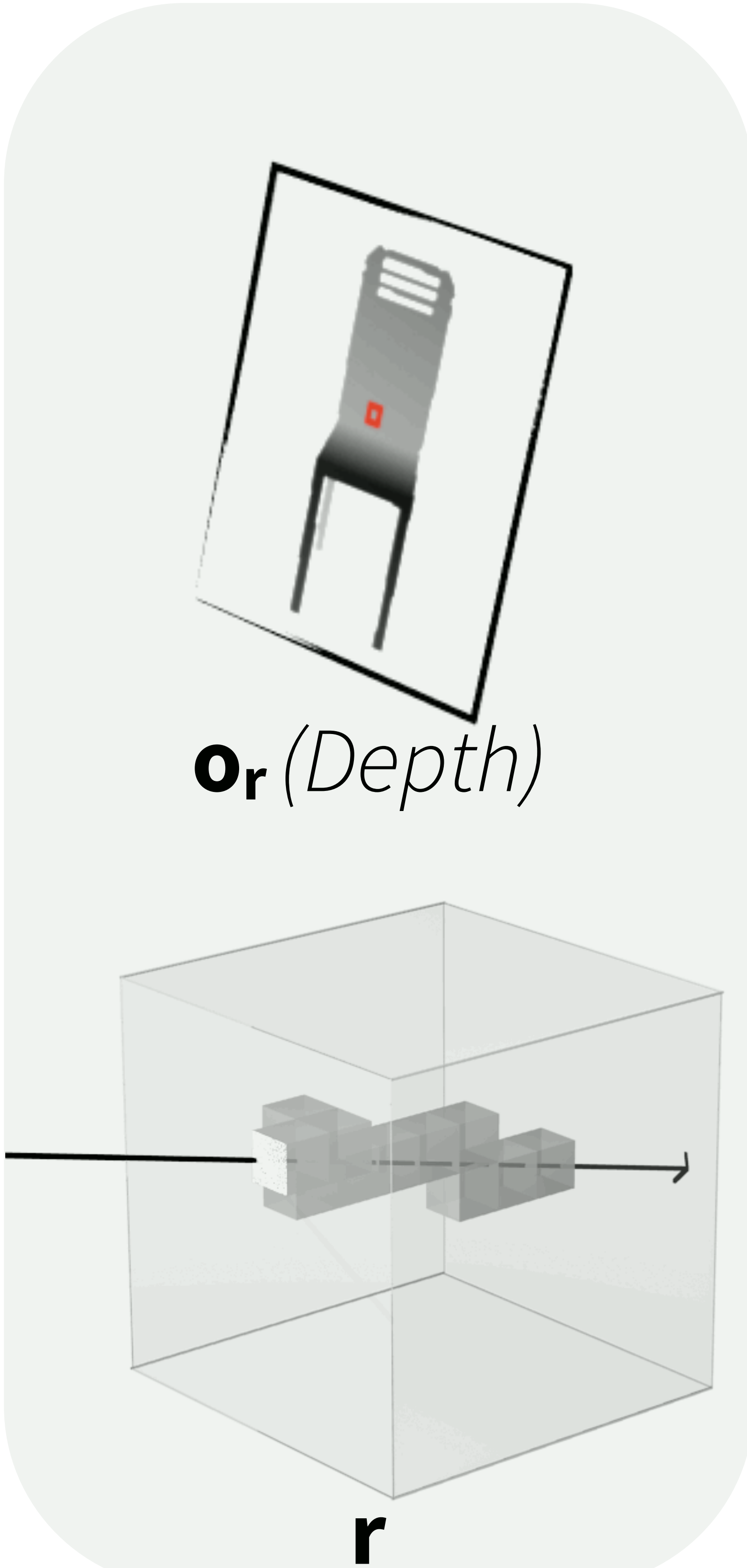
How inconsistent is each event w.r.t $\mathbf{o_r}$?

Event Costs



$$\psi_r^{depth}(i) = |d_i^r - d_r|$$

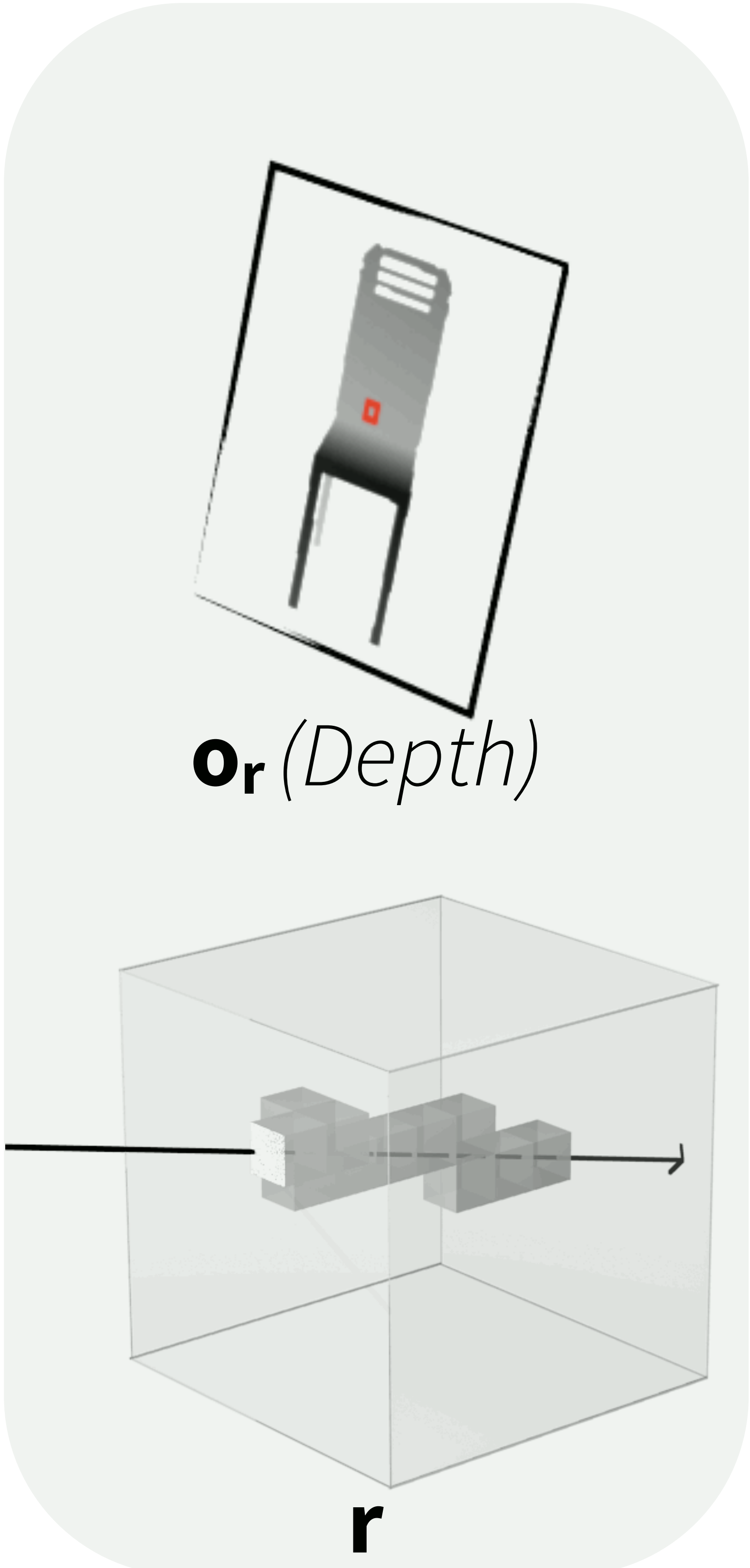
Event Costs



$$\psi_r^{depth}(i) = |d_i^r - \boxed{d_r}|$$

Observed Depth

Event Costs

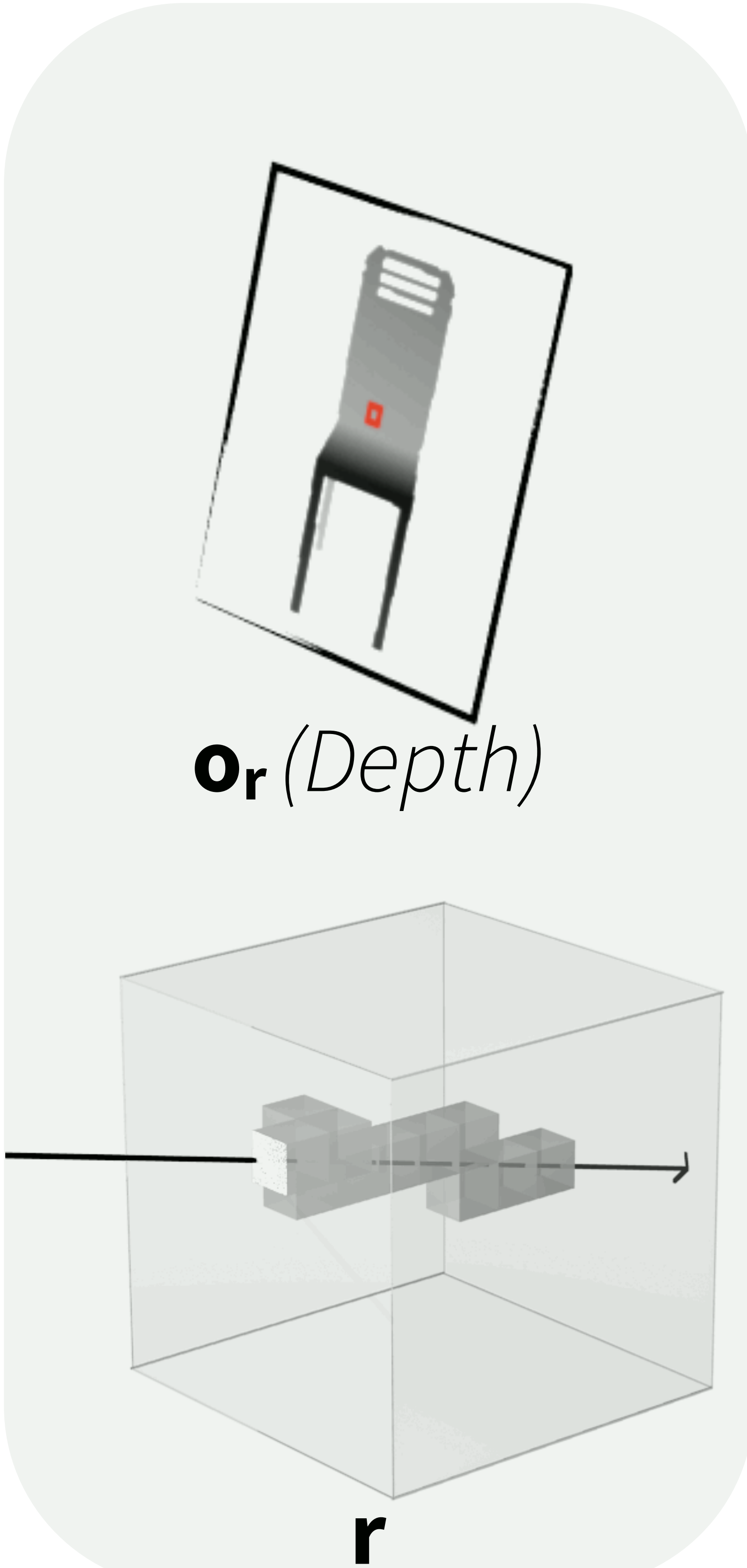


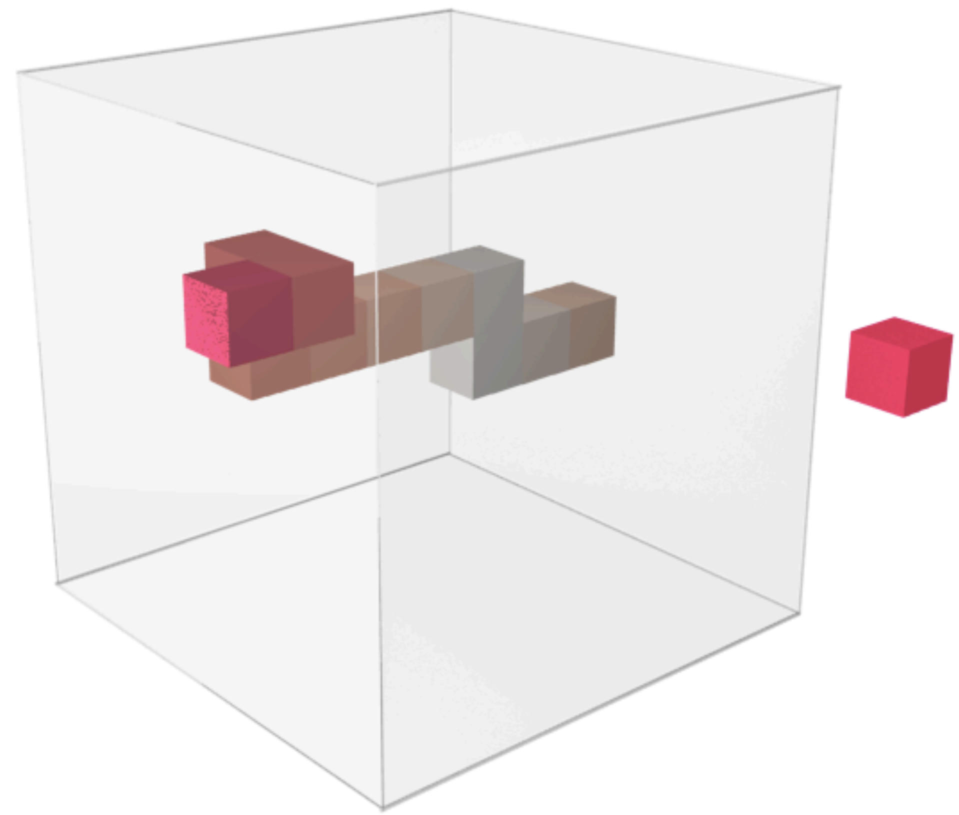
$$\psi_r^{depth}(i) = |\boxed{d_i^r} - \boxed{d_r}|$$

Depth under event i

Observed Depth

Event Costs





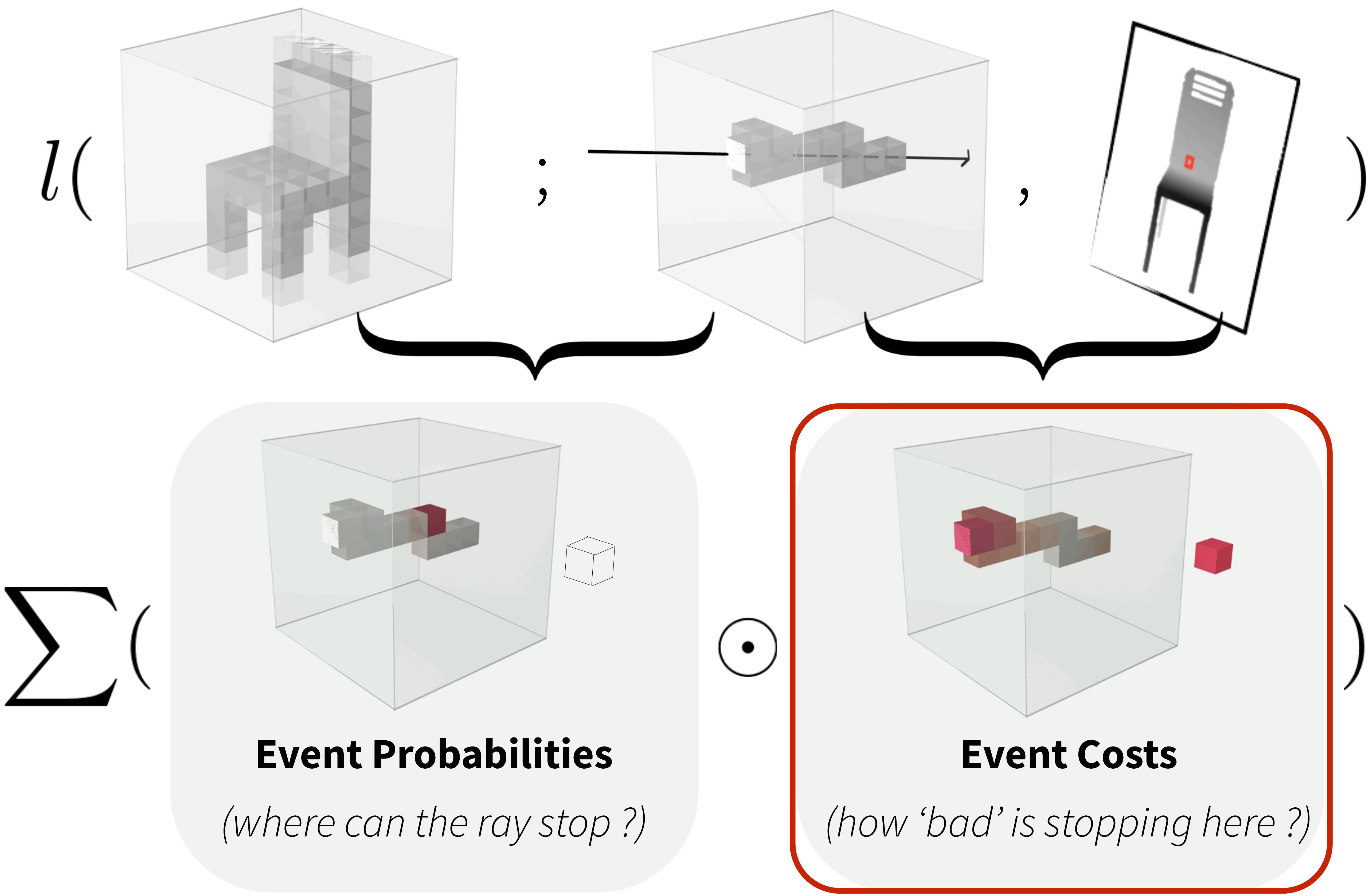
Event Costs

$$\psi_r^{depth}(i) = |\boxed{d_i^r} - \boxed{d_r}|$$

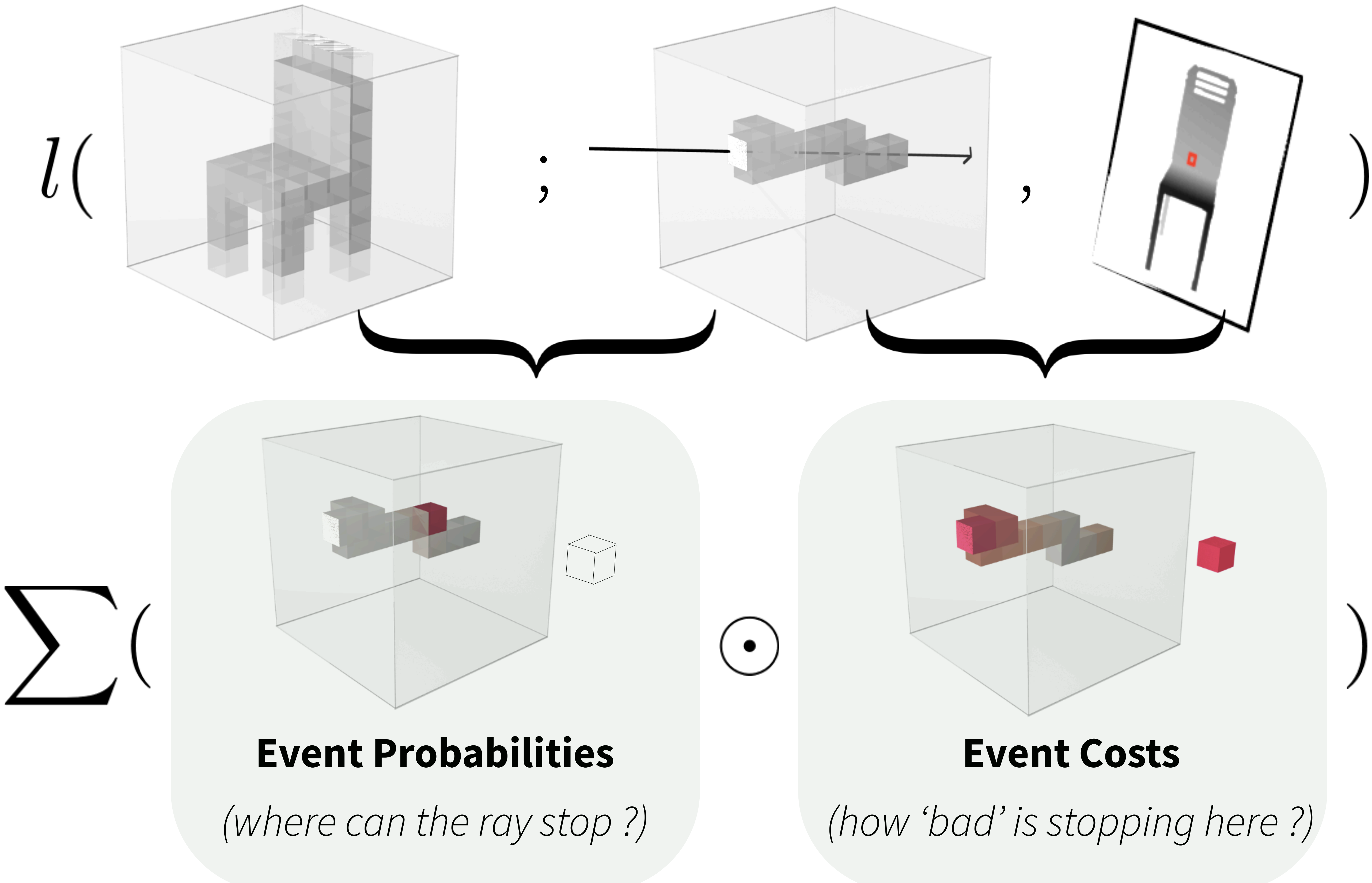
Depth under event i

Observed Depth

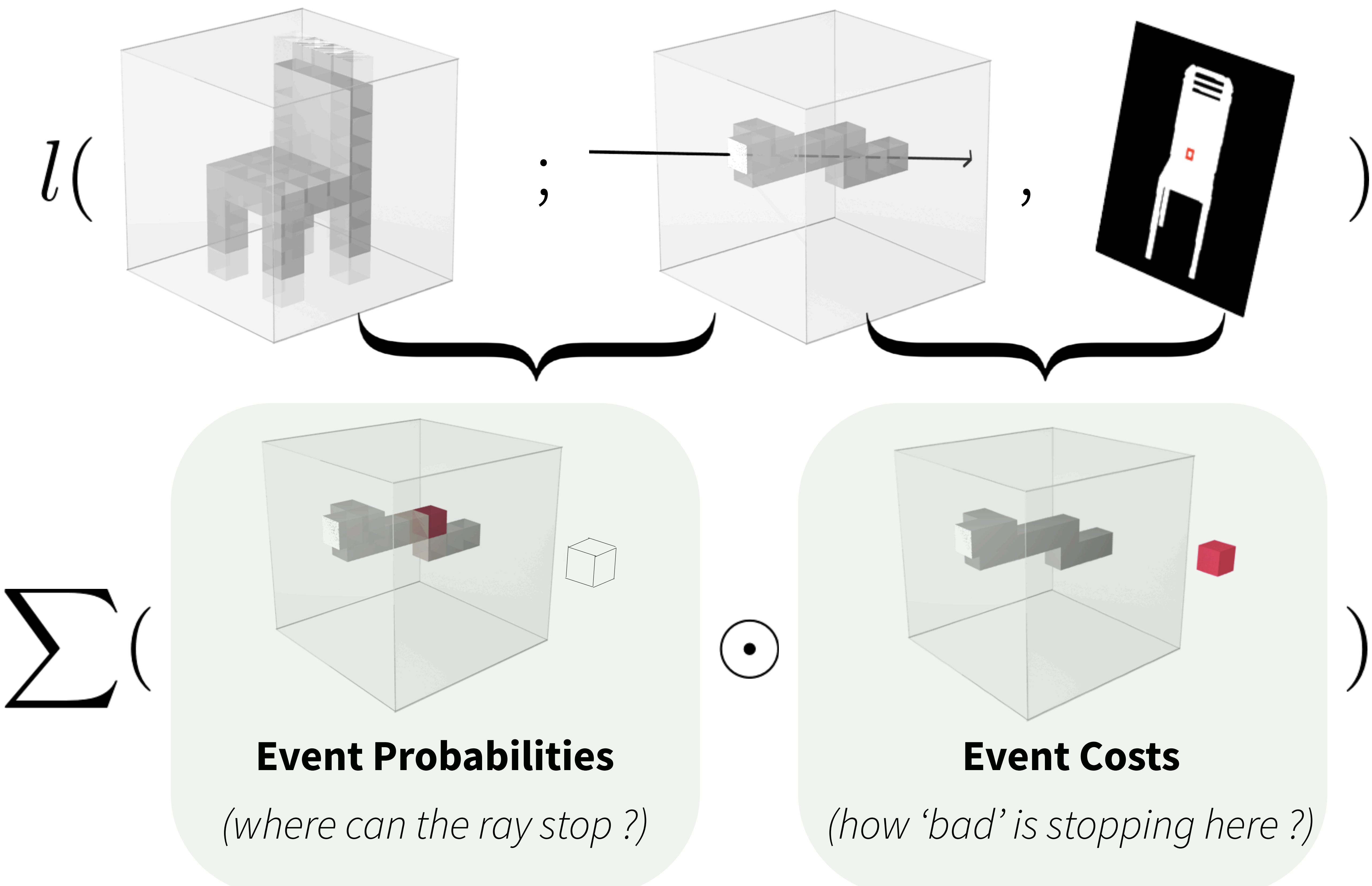
Differentiable Ray Consistency



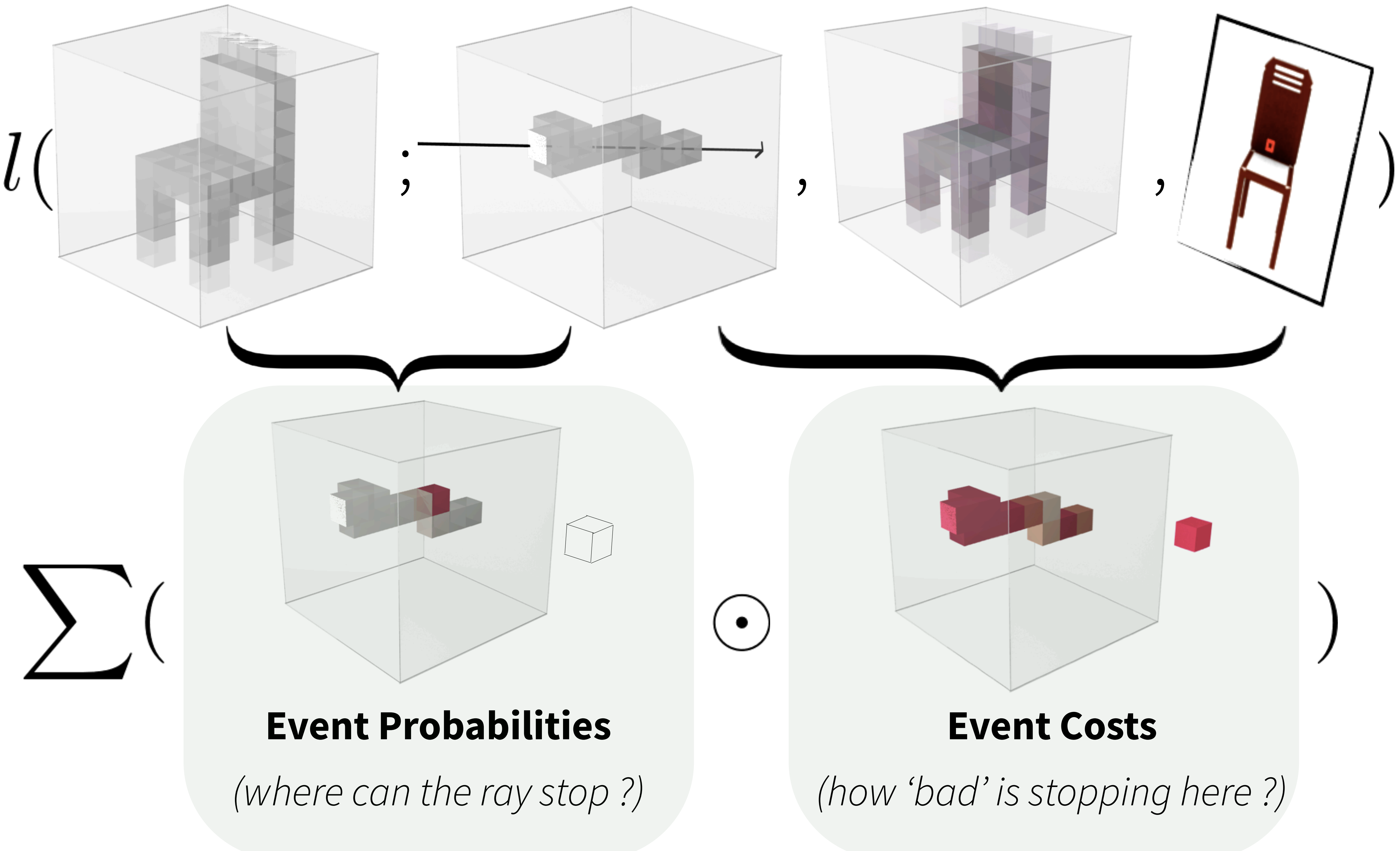
Differentiable Ray Consistency



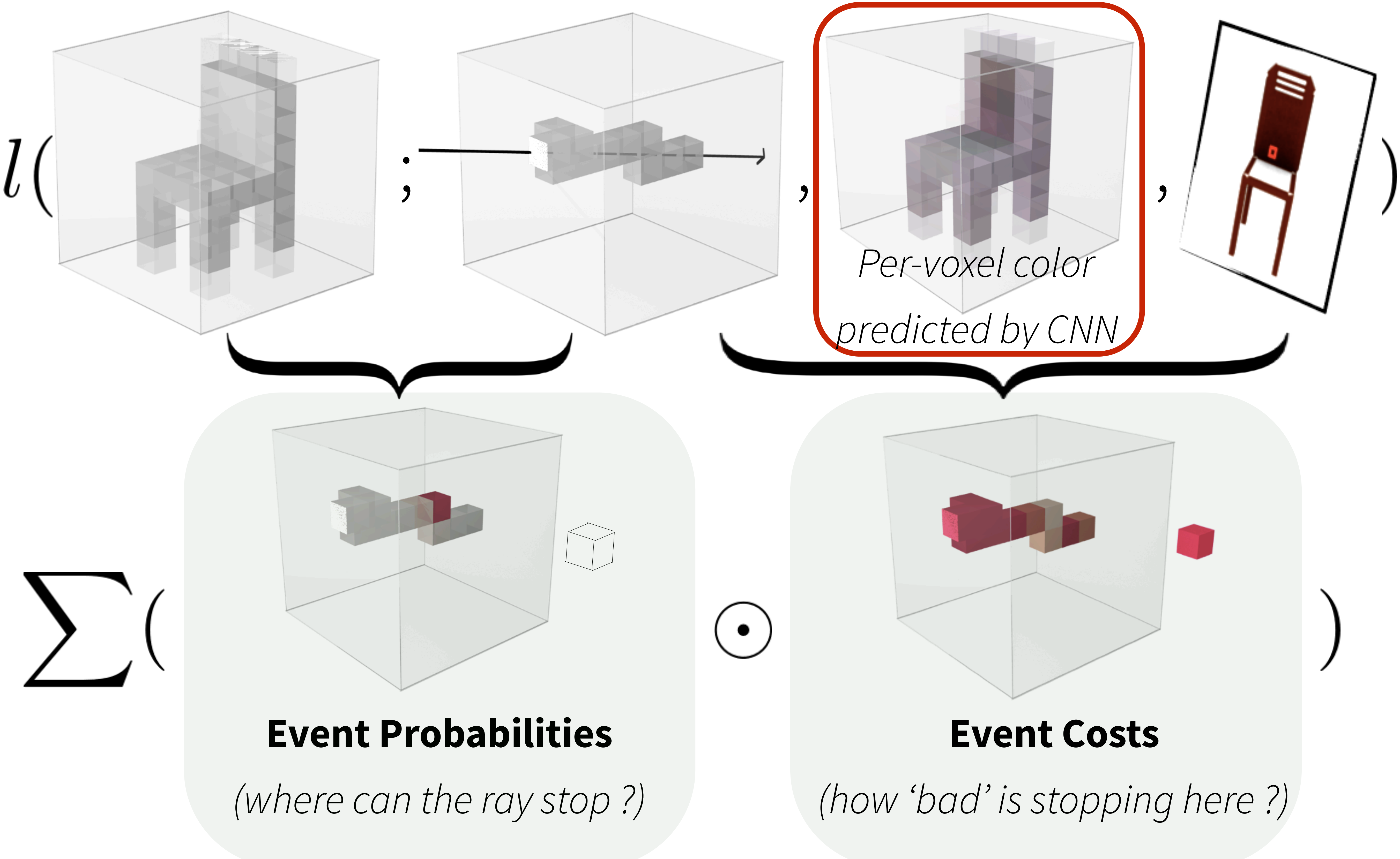
Mask Observation



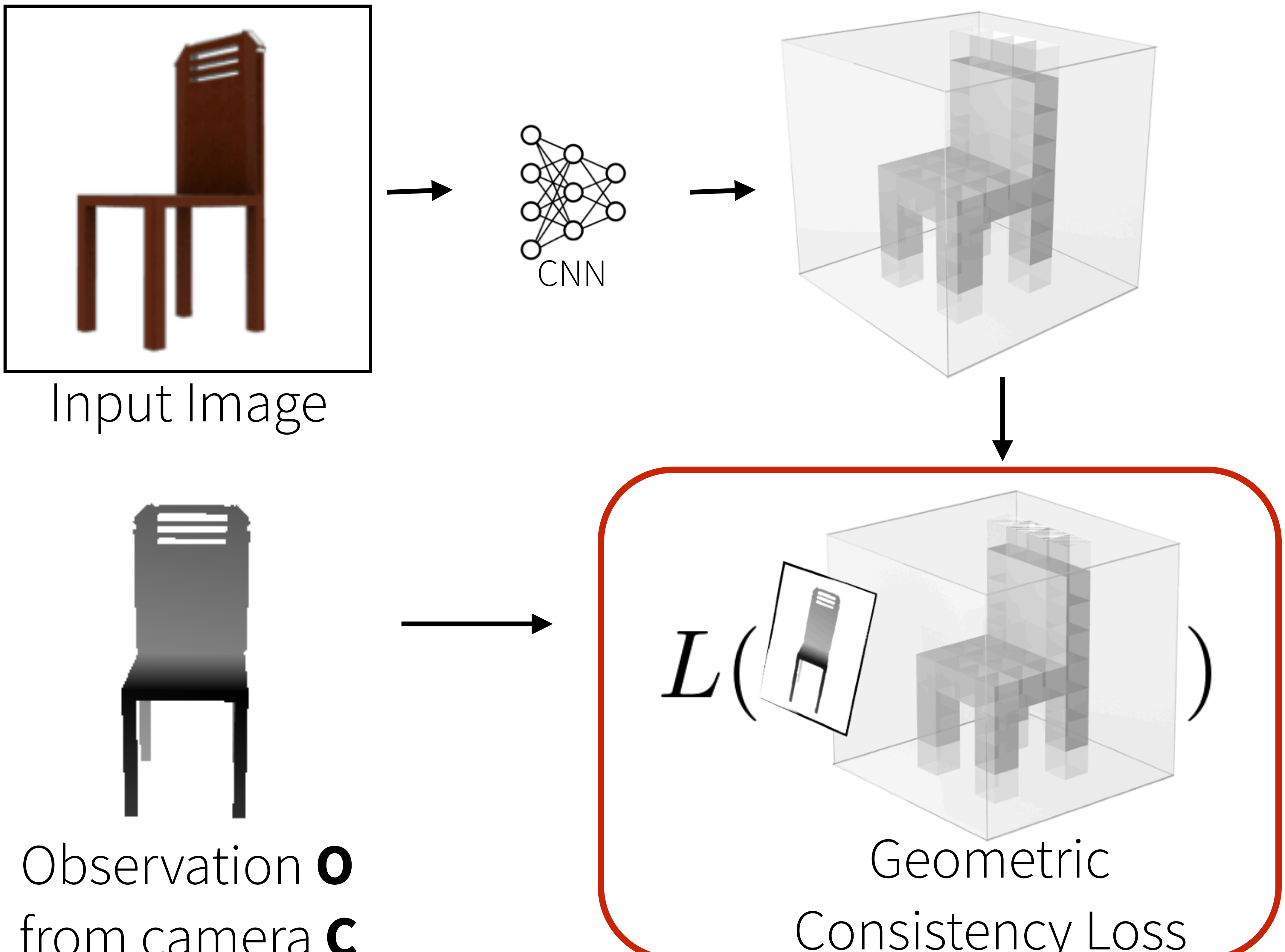
Color Observation



Color Observation

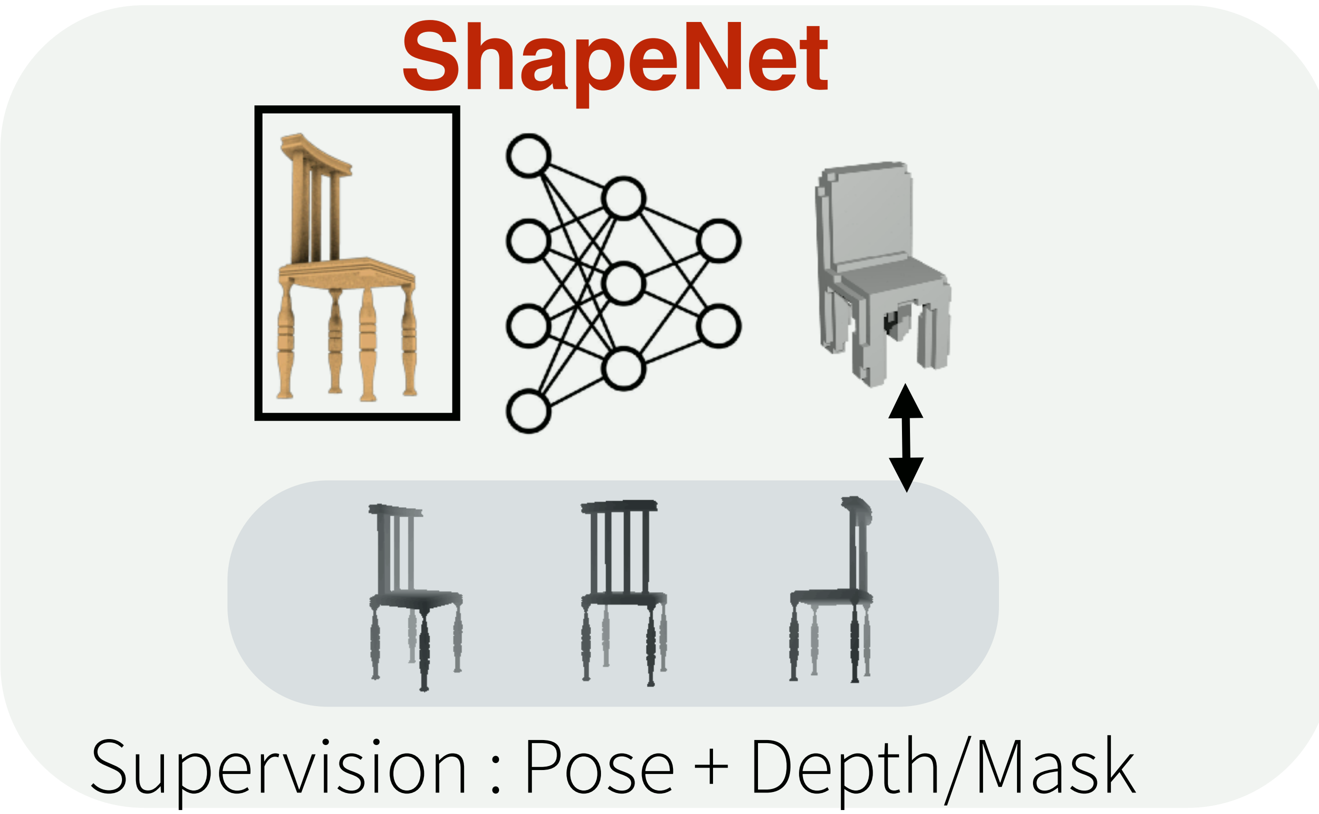


Learning via Geometric Consistency

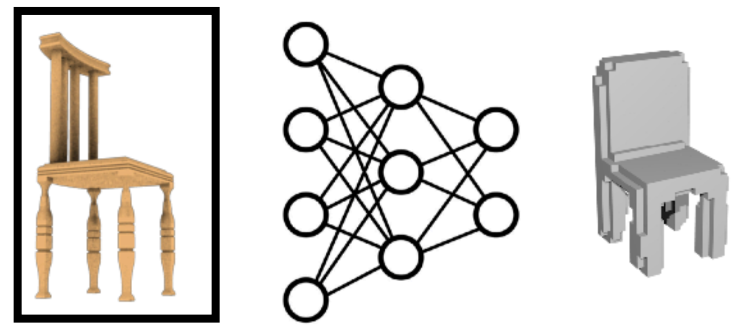


Learning Single-view Reconstruction

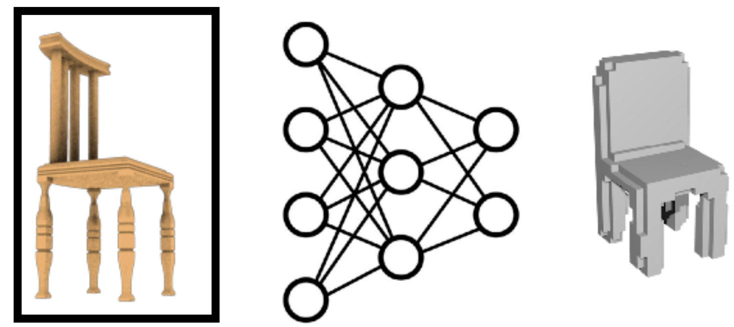
Learning Single-view Reconstruction



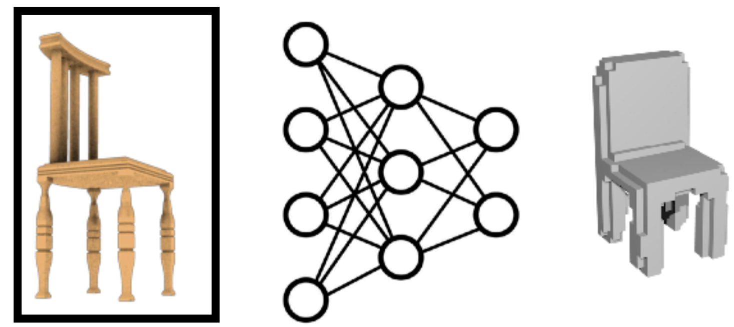
Experiments - ShapeNet



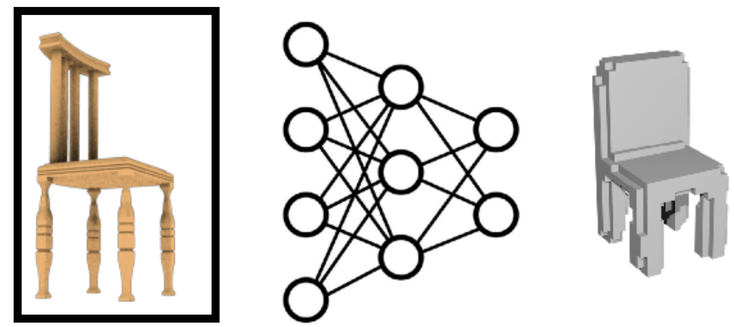
Experiments - ShapeNet

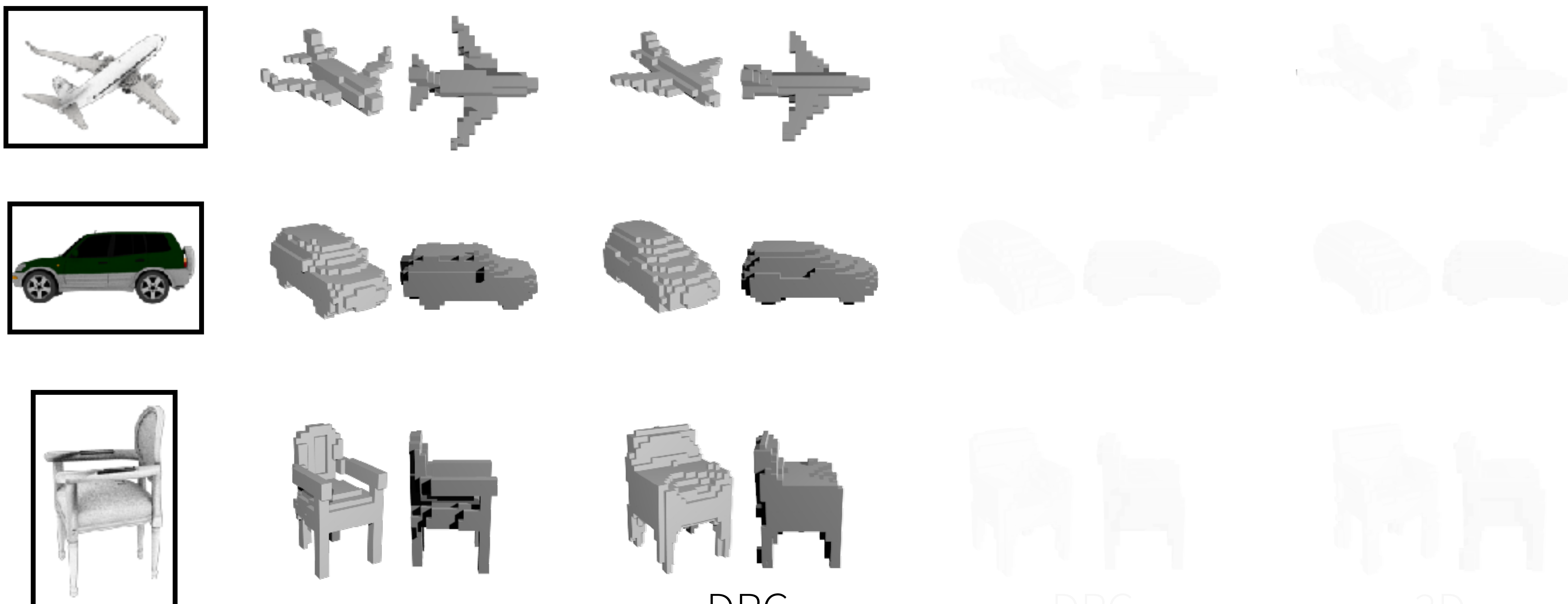


Experiments - ShapeNet



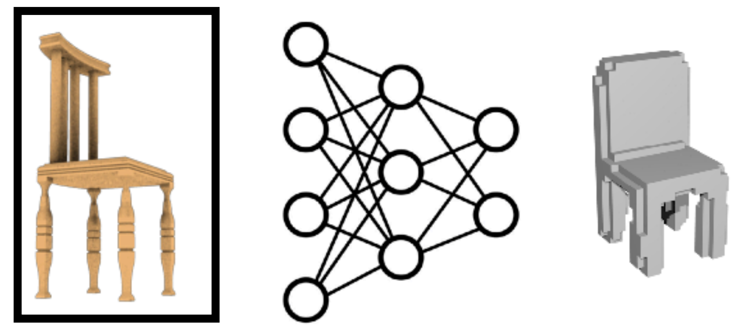
Experiments - ShapeNet

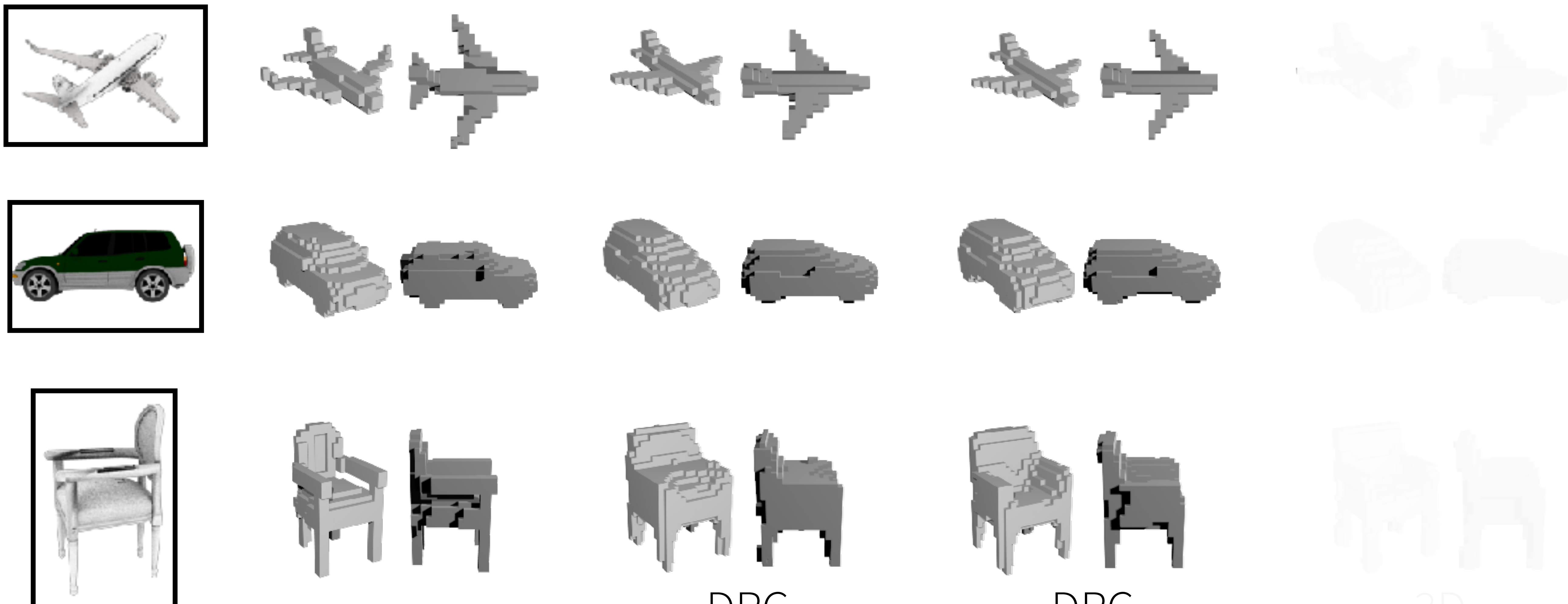




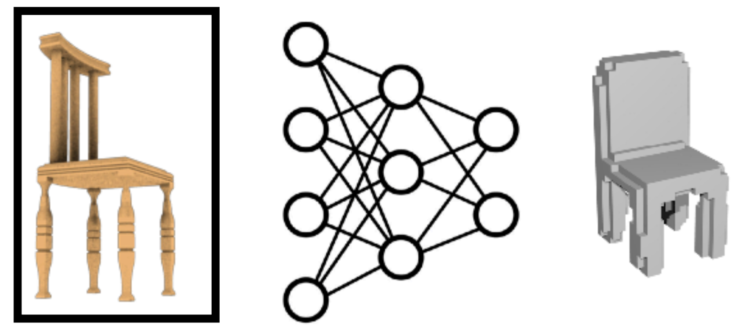
Input | GT | DRC (Mask) | DRC (Depth) | 3D Supervision

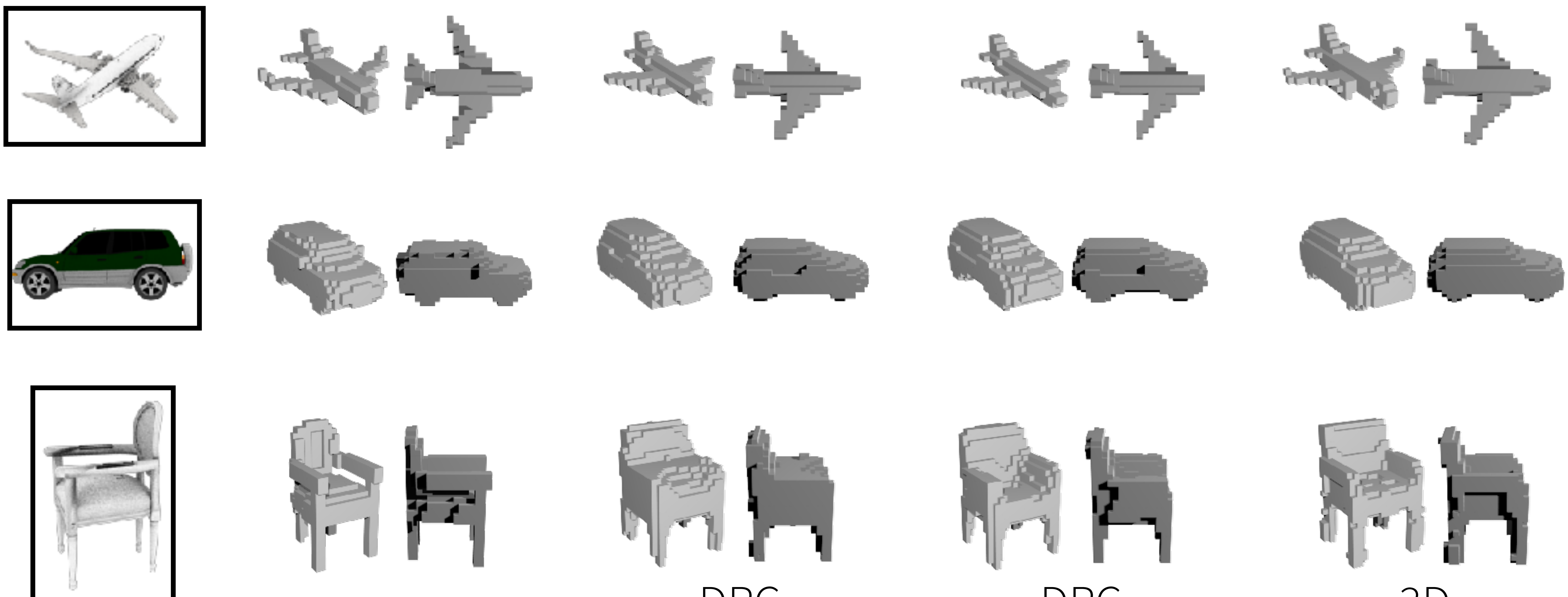
Experiments - ShapeNet





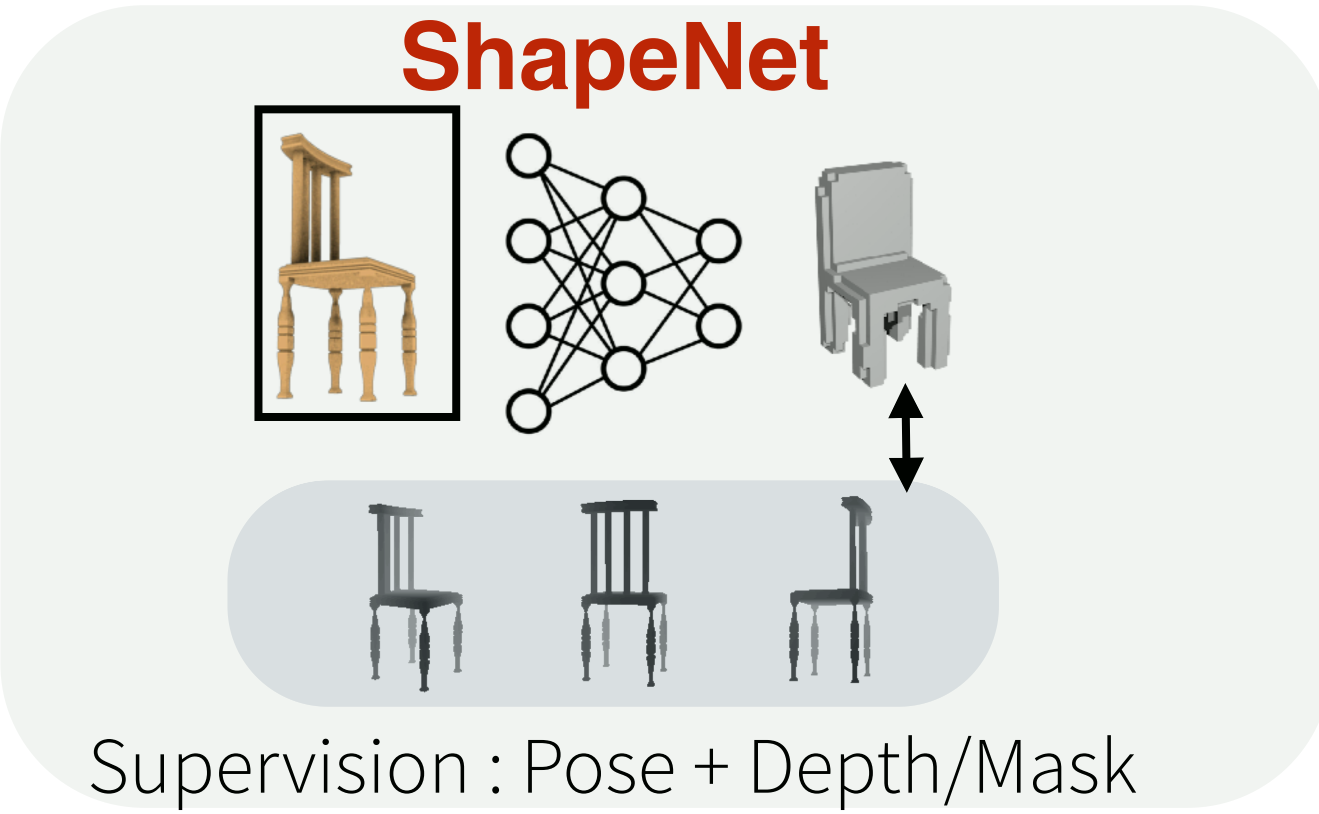
Experiments - ShapeNet



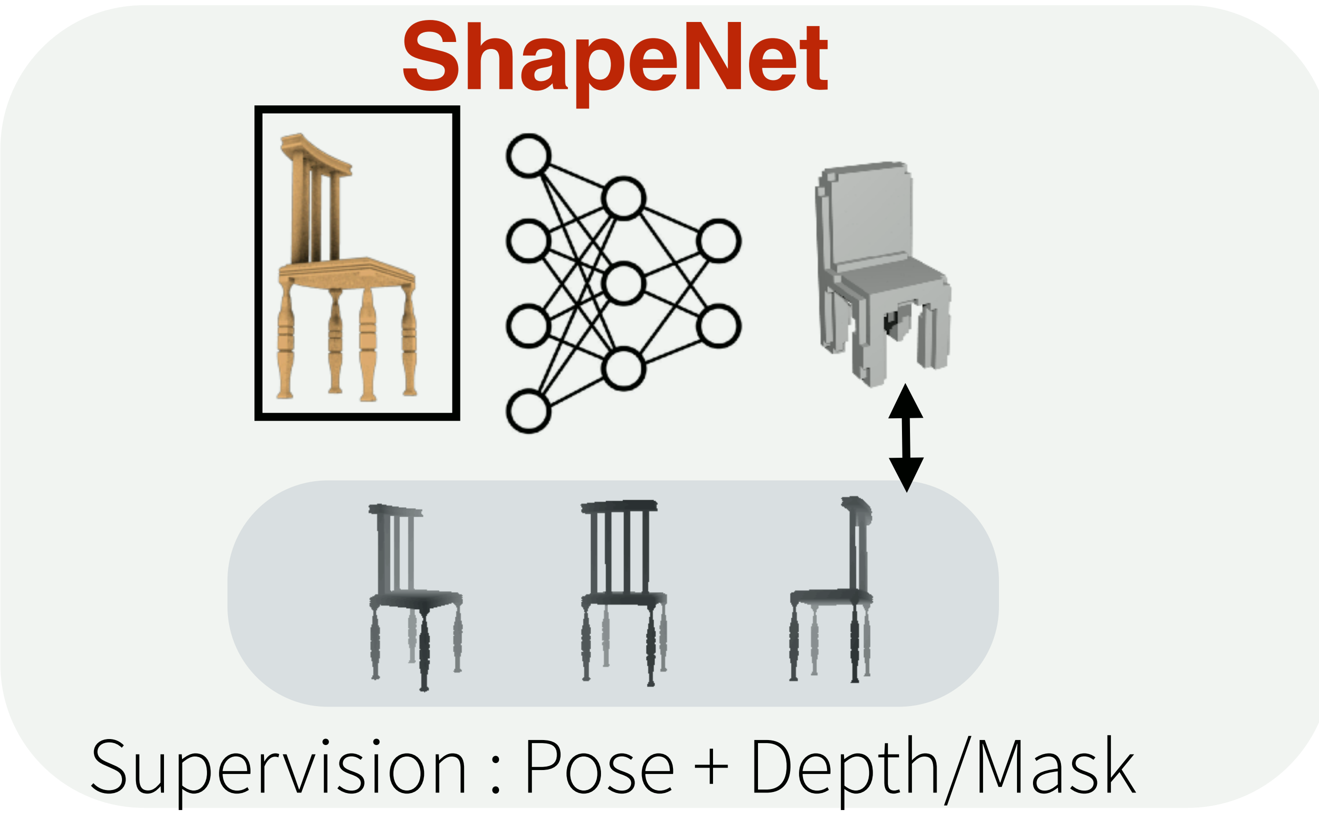


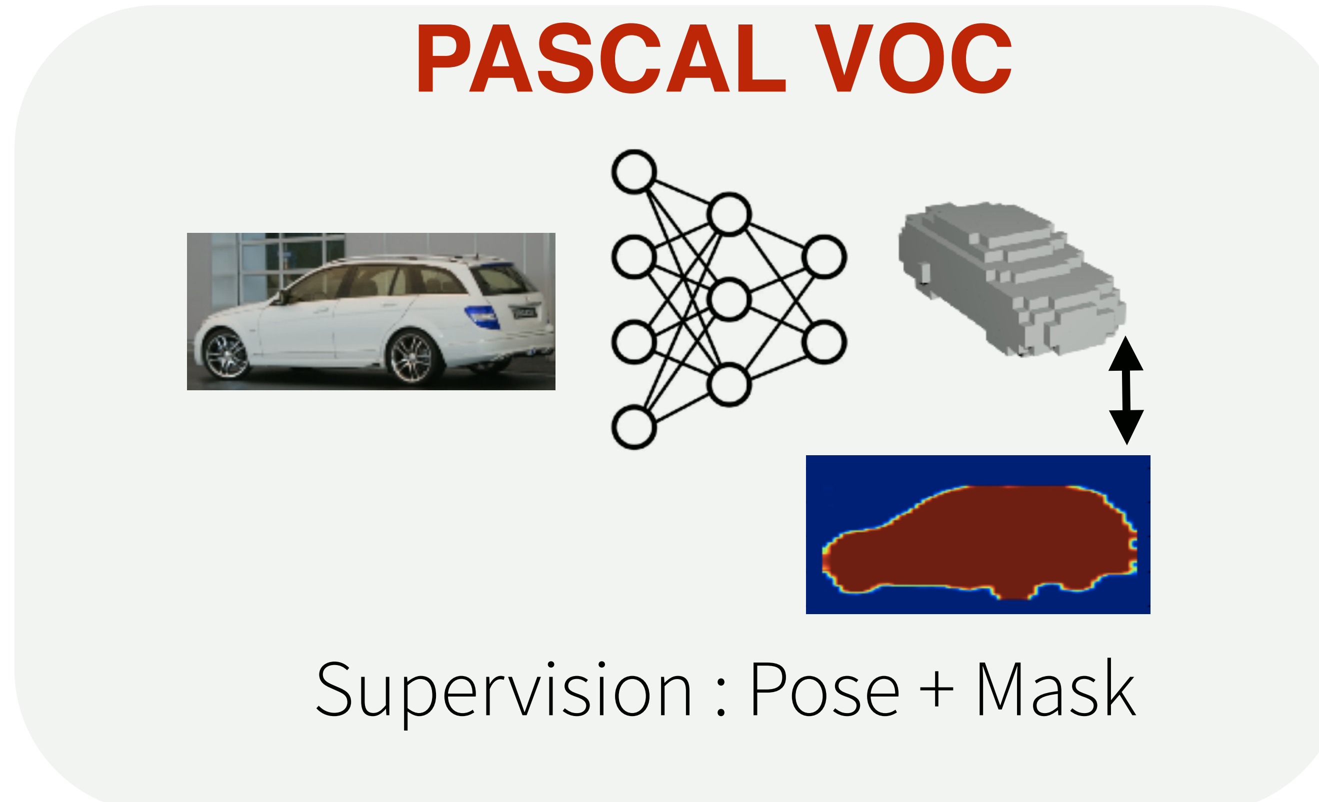
Input GT DRC (Mask) DRC (Depth) 3D Supervision

Learning Single-view Reconstruction

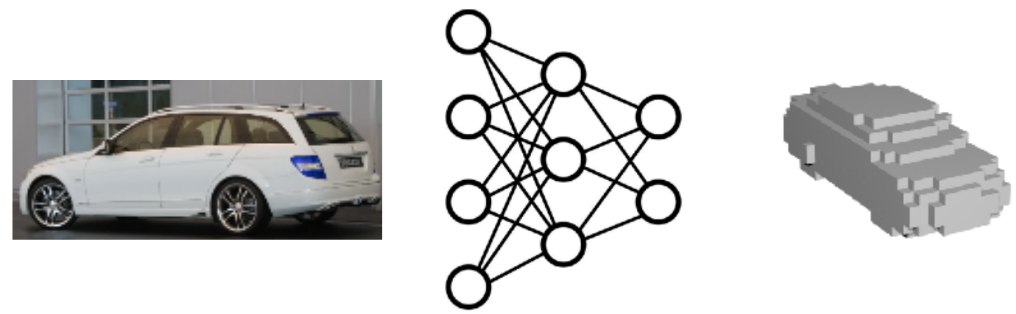


Learning Single-view Reconstruction

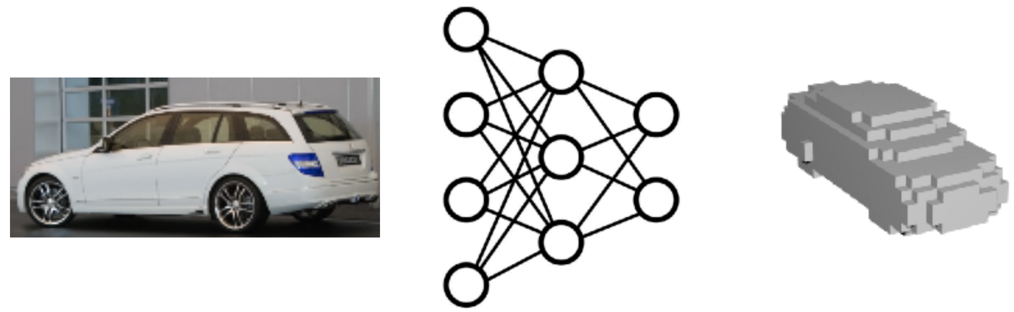




Experiments - PASCAL VOC



Experiments - PASCAL VOC



Input

CSDM (Kar et. al.)

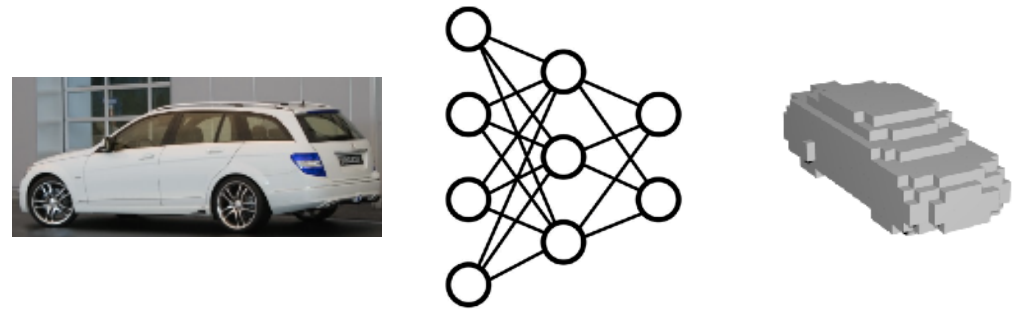
DRC (Pascal)

SNet 3D

DRC (Joint)

'Ground-Truth'

Experiments - PASCAL VOC



Input

CSDM
(Kar et. al.)

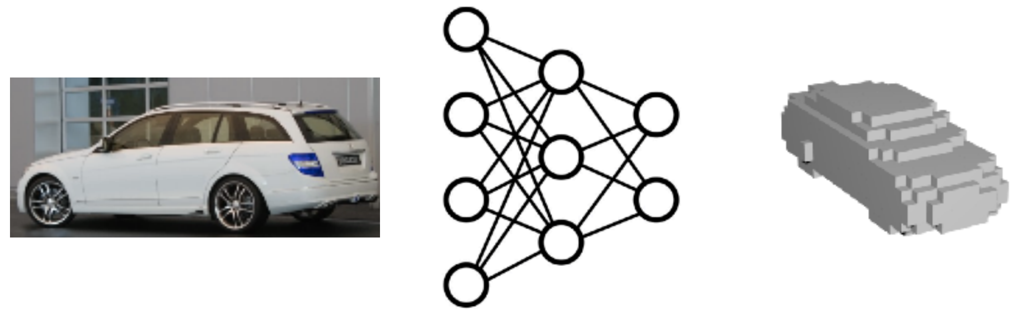
DRC
(Pascal)

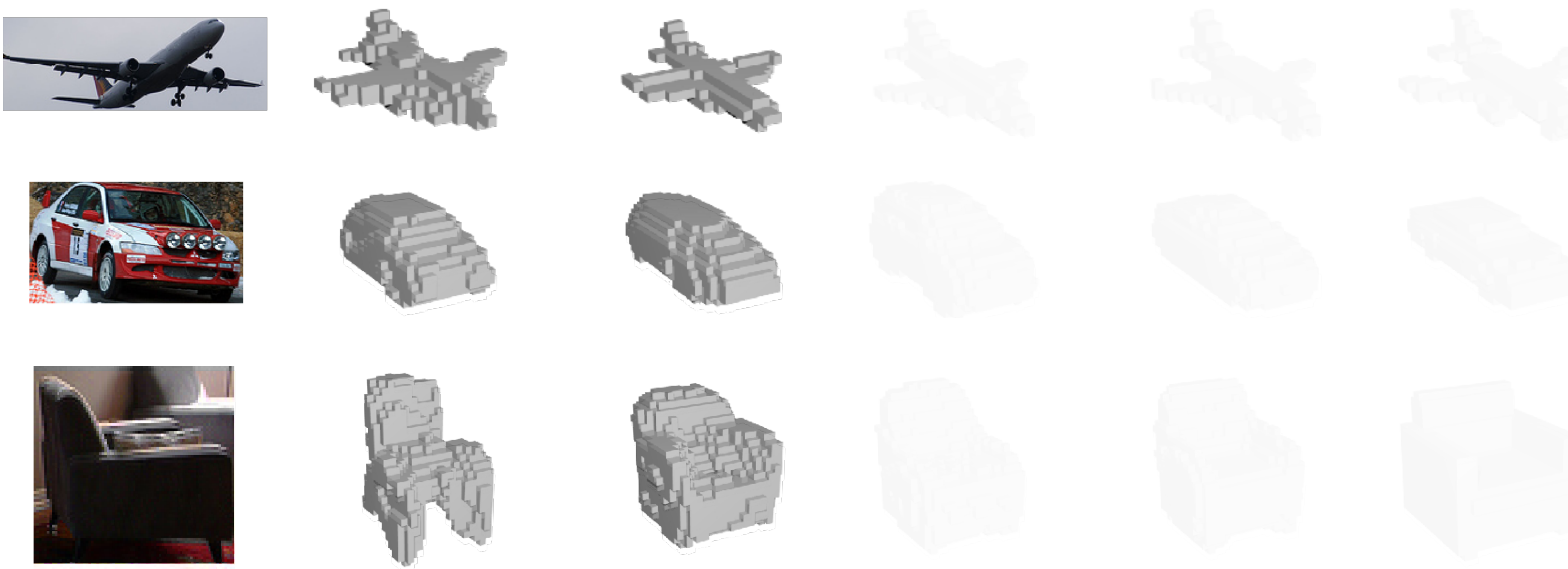
SNet 3D

DRC
(Joint)

'Ground-
Truth'

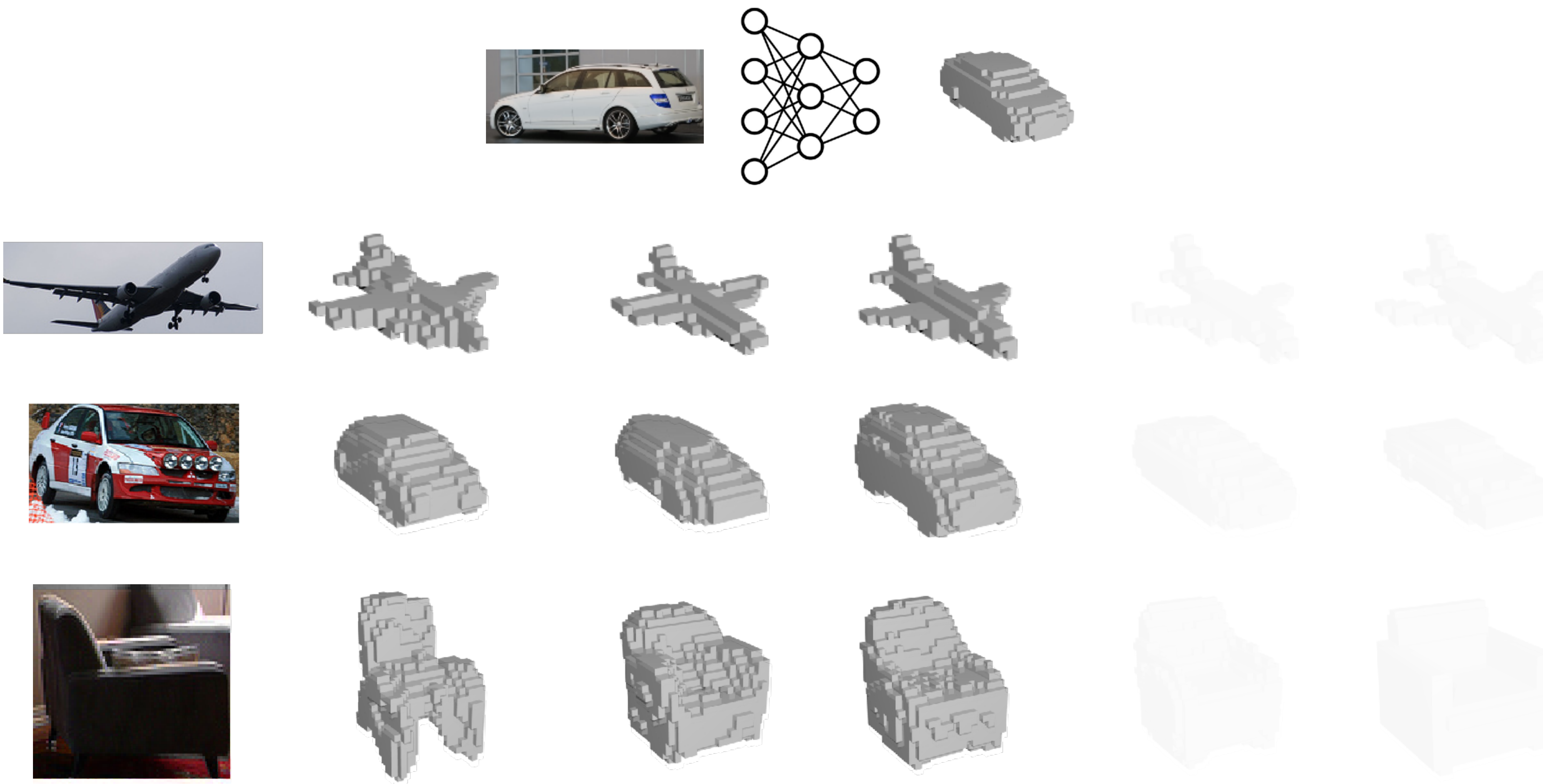
Experiments - PASCAL VOC





Input	CSDM (Kar et. al.)	DRC (Pascal)	SNet 3D	DRC (Joint)	‘Ground-Truth’

Experiments - PASCAL VOC



Input

CSDM
(Kar et. al.)

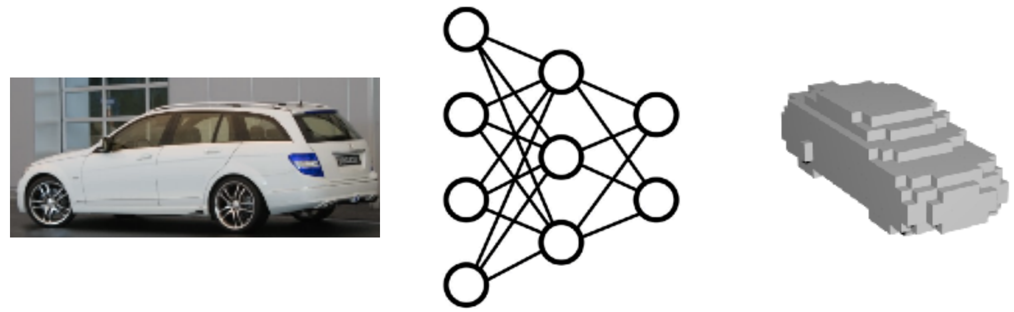
DRC
(Pascal)

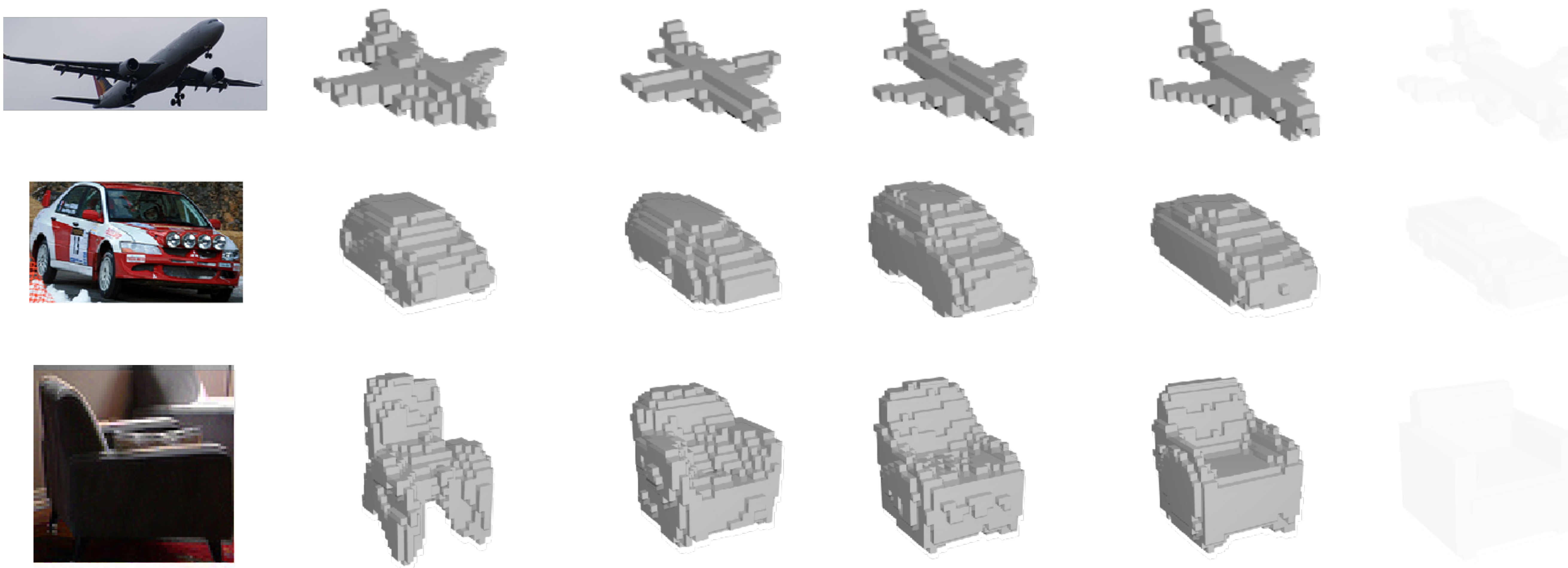
SNet 3D

DRC
(Joint)

'Ground-
Truth'

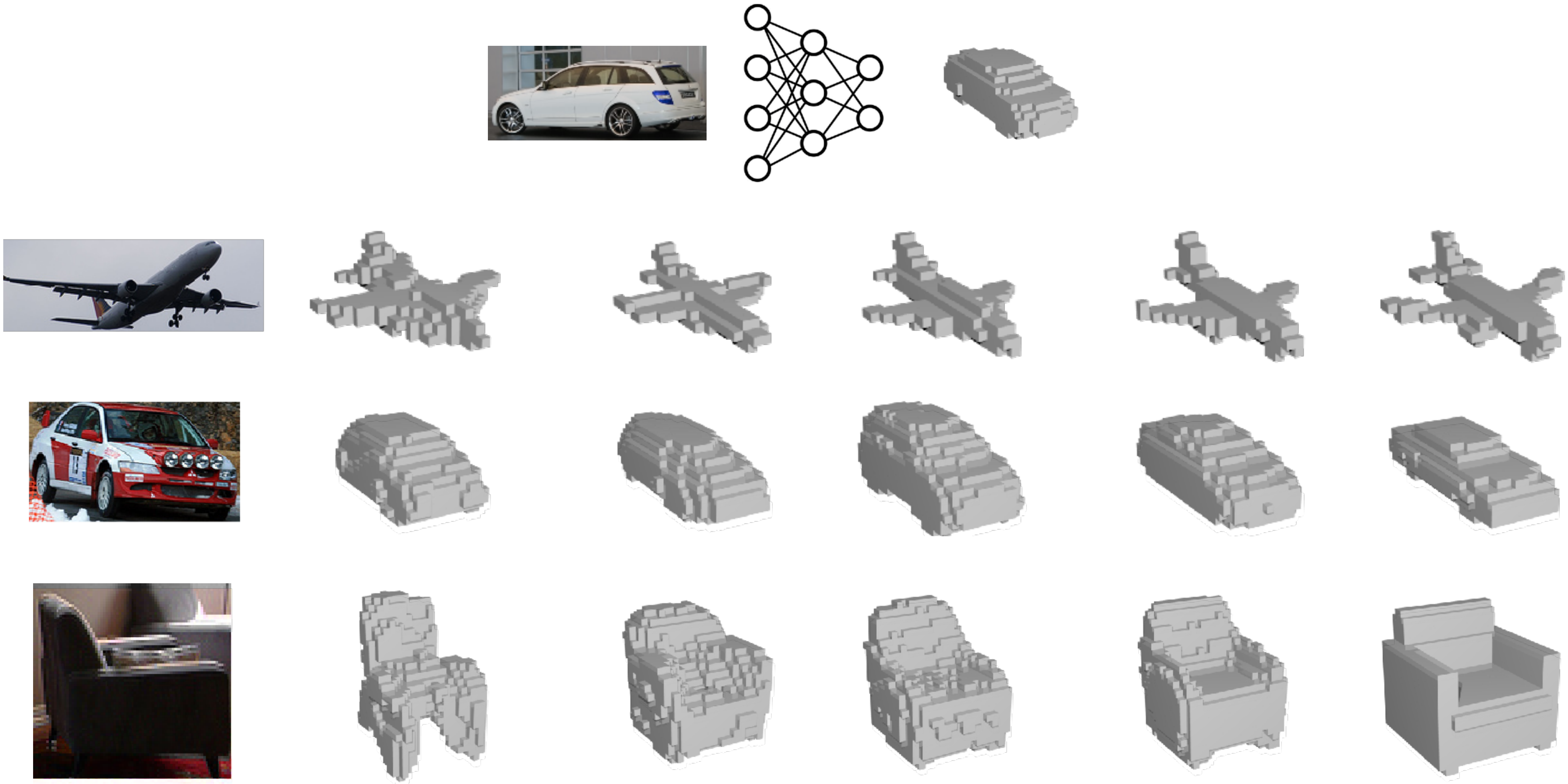
Experiments - PASCAL VOC





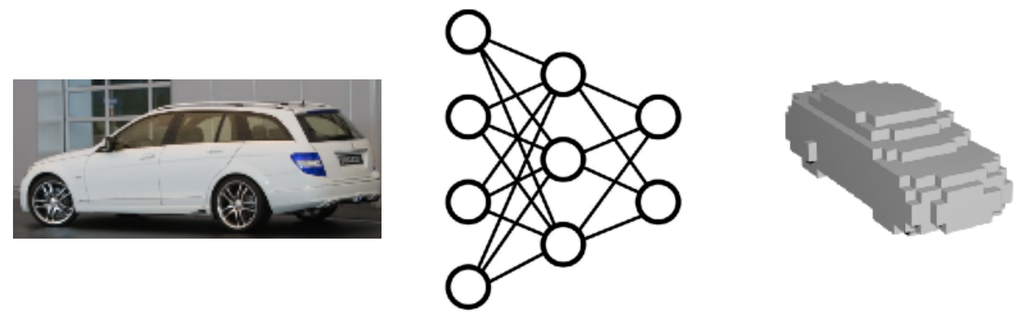
Input | CSDM (Kar et. al.) | DRC (Pascal) | SNet 3D | DRC (Joint) | ‘Ground-Truth’

Experiments - PASCAL VOC

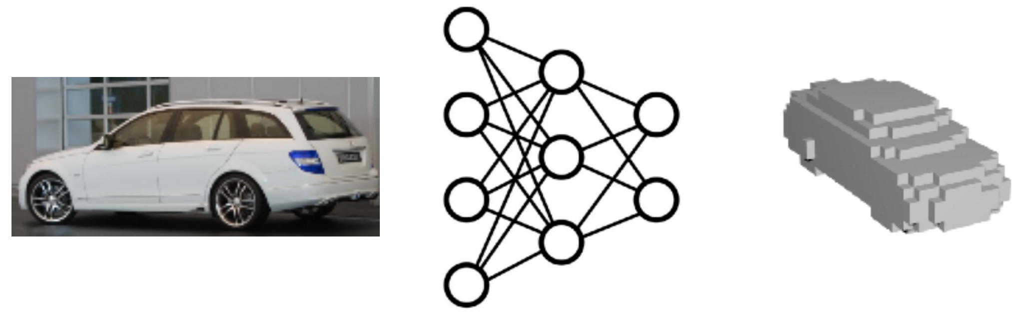


Input | CSDM (Kar et. al.) | DRC (Pascal) | SNet 3D | DRC (Joint) | ‘Ground-Truth’

Experiments - PASCAL VOC

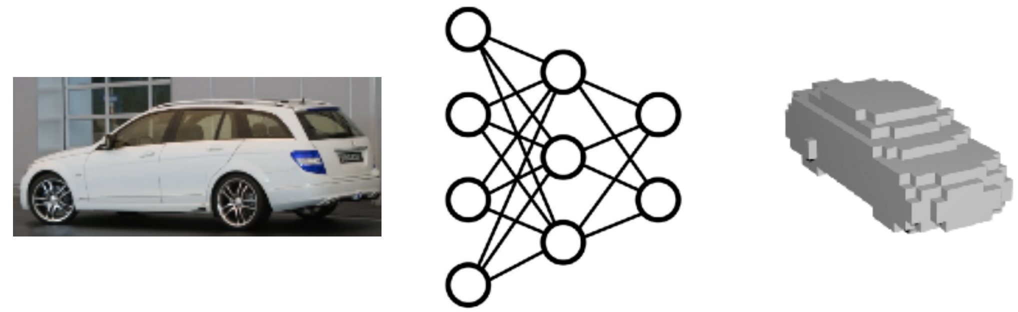


Experiments - PASCAL VOC



Input

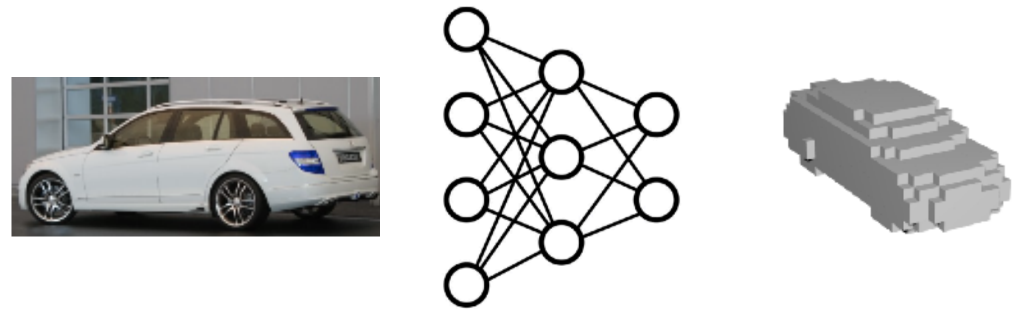
Experiments - PASCAL VOC



Input

Prediction

Experiments - PASCAL VOC

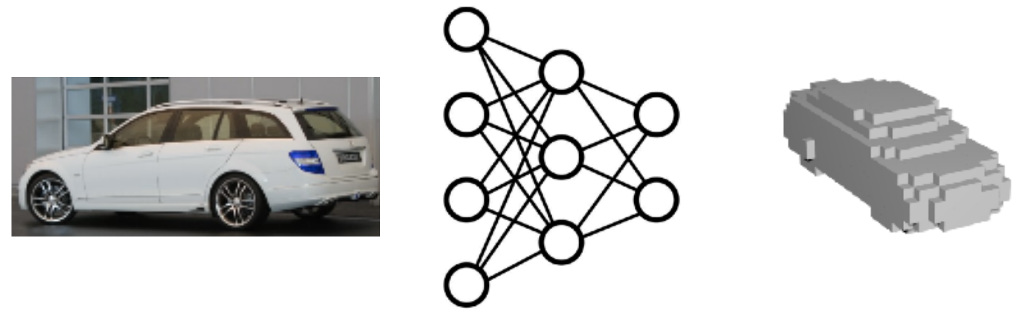


Input

Prediction

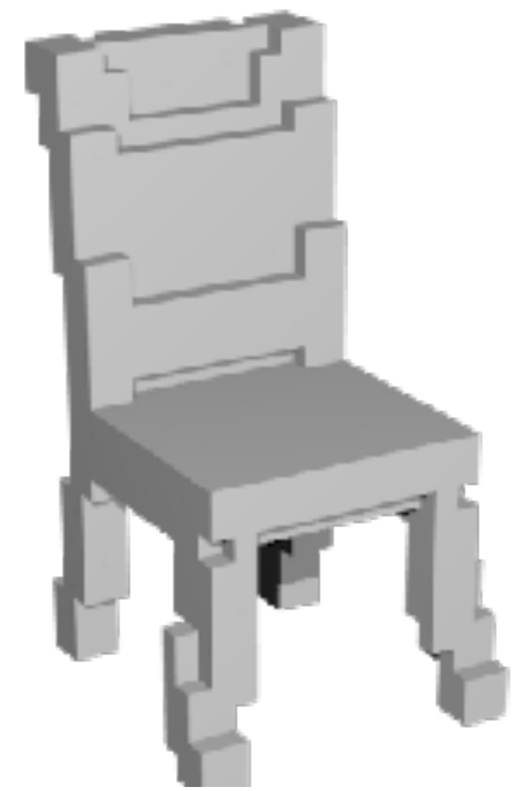
‘Ground-truth’

Experiments - PASCAL VOC



Input

Prediction

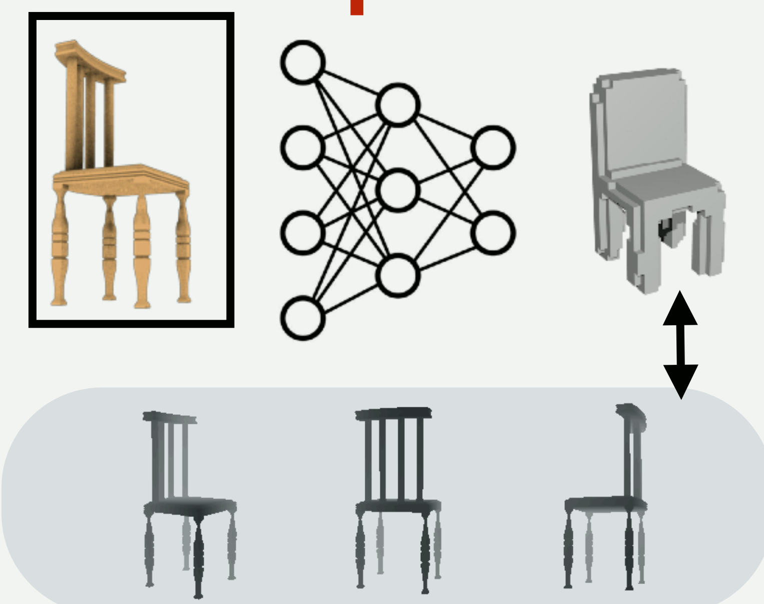


‘Ground-truth’

Collecting ‘ground-truth’ 3D is hard !

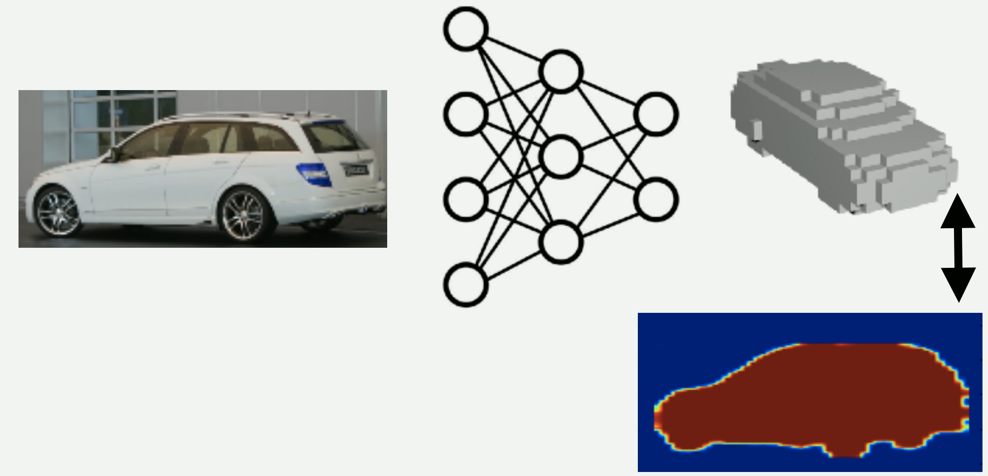
Learning Single-view Reconstruction

ShapeNet



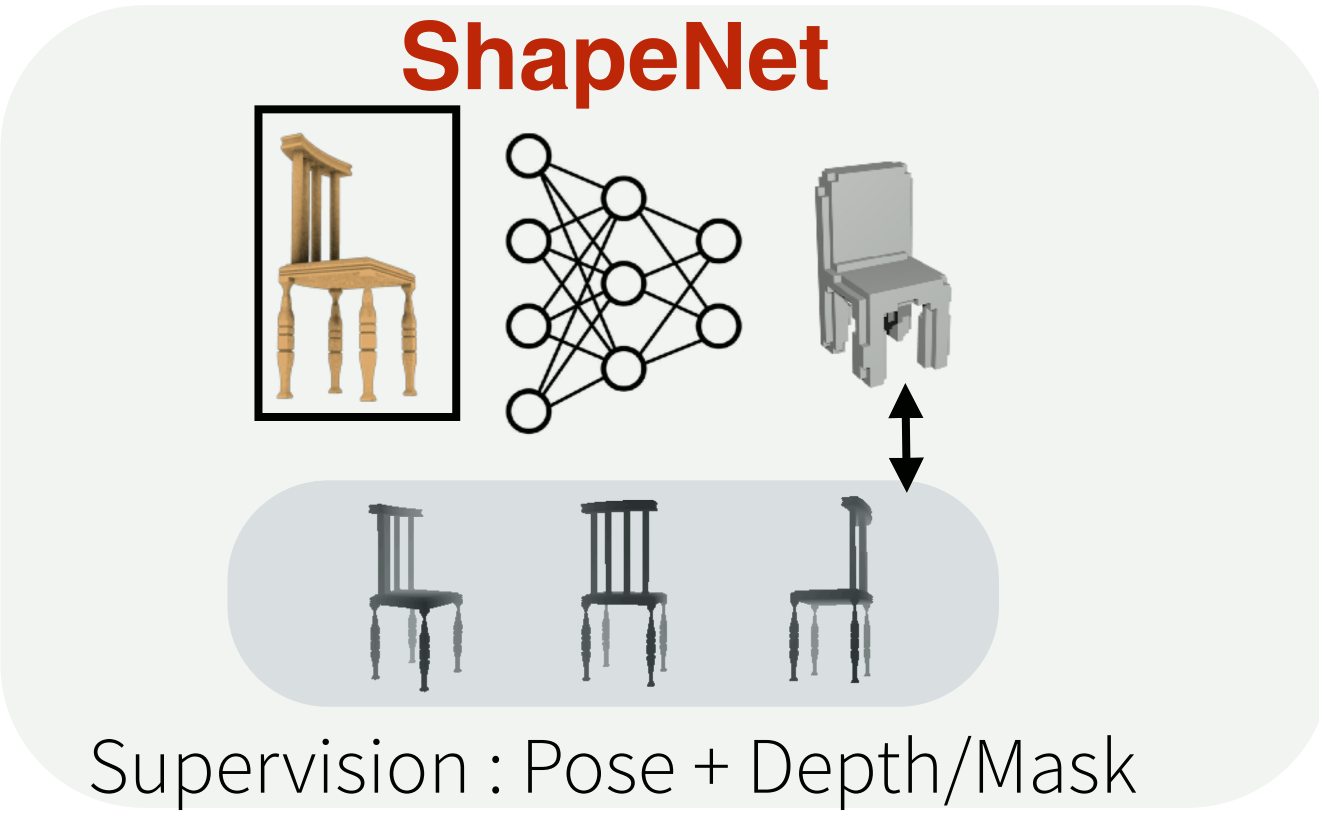
Supervision : Pose + Depth/Mask

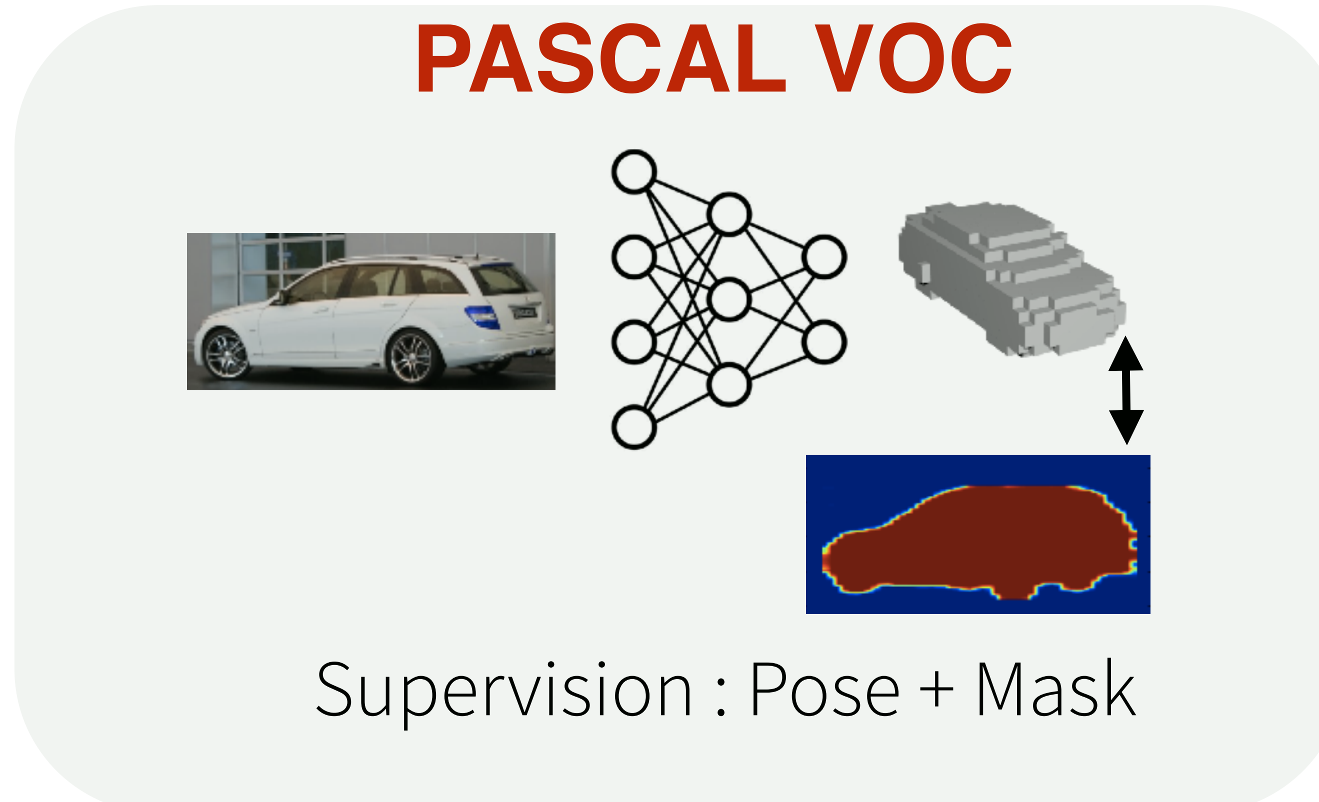
PASCAL VOC

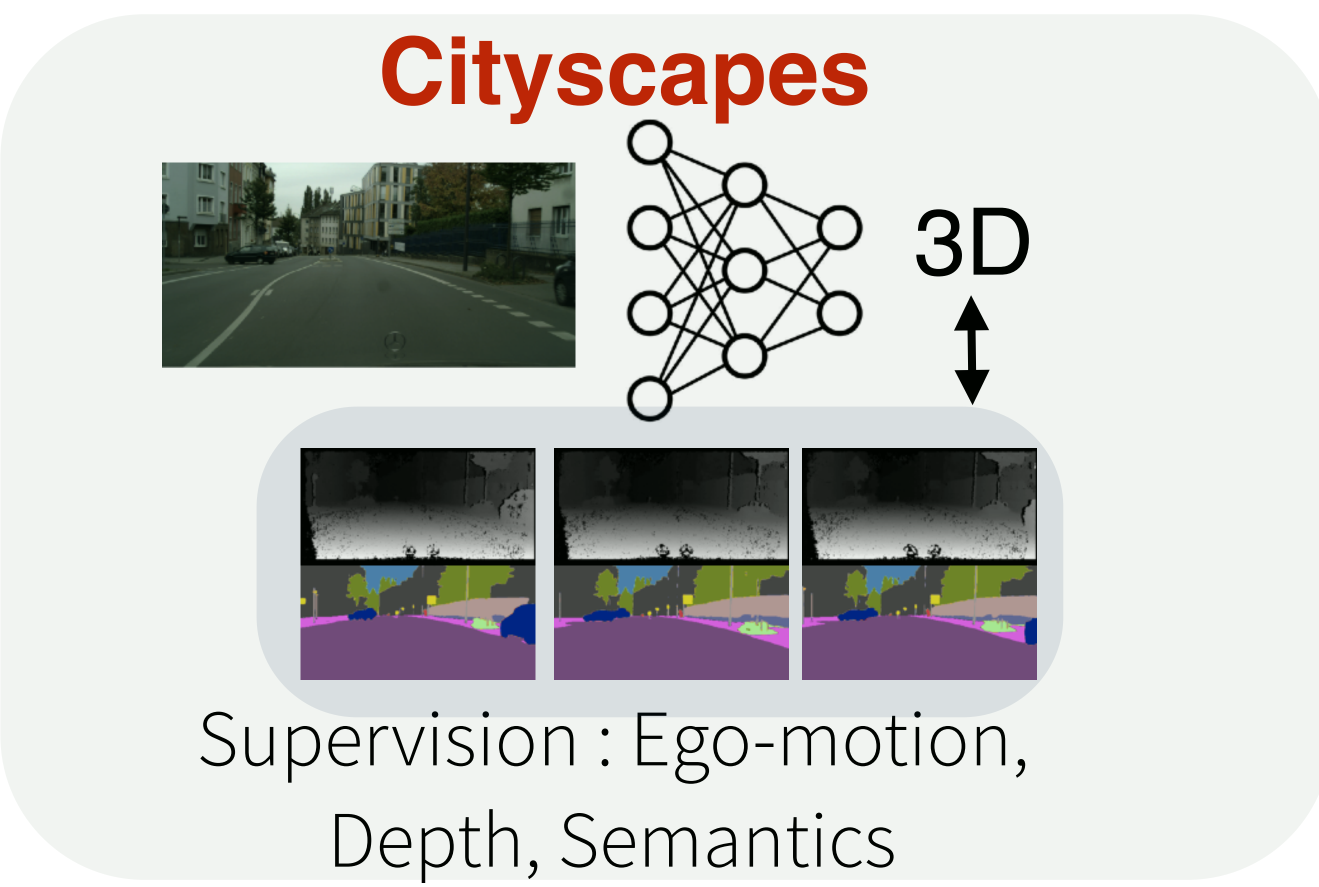


Supervision : Pose + Mask

Learning Single-view Reconstruction

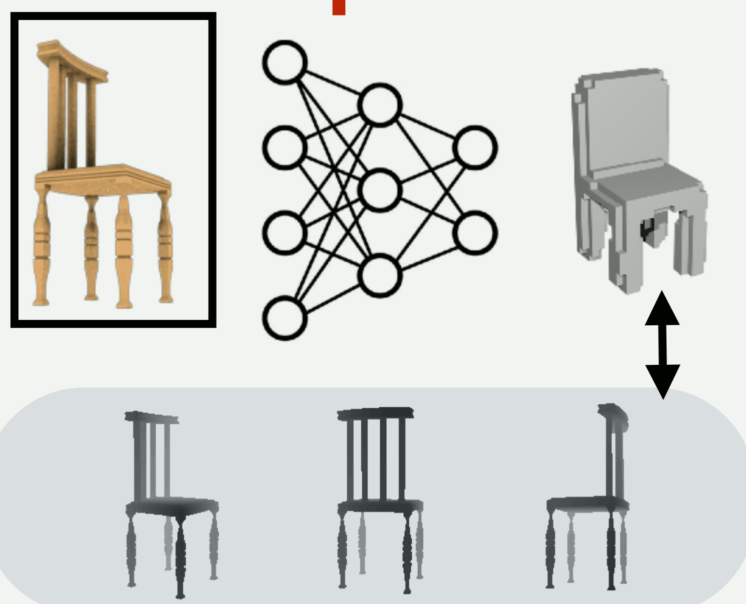






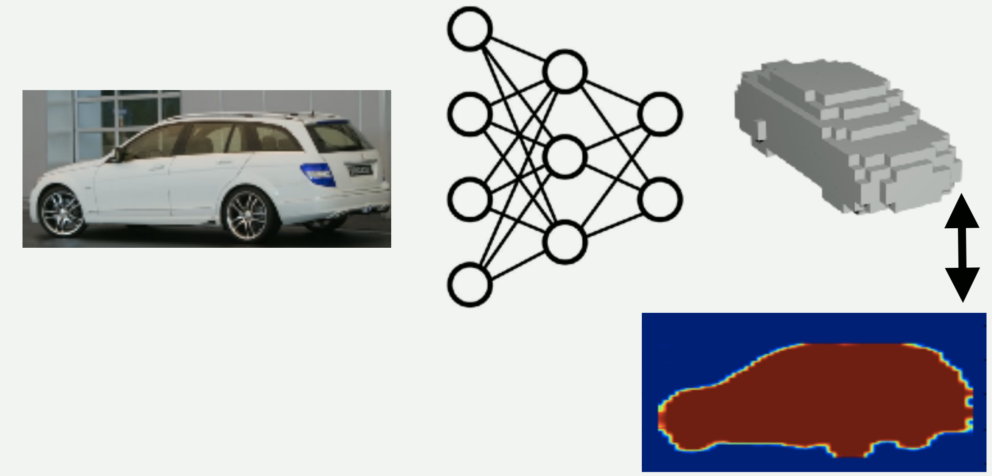
Learning Single-view Reconstruction

ShapeNet



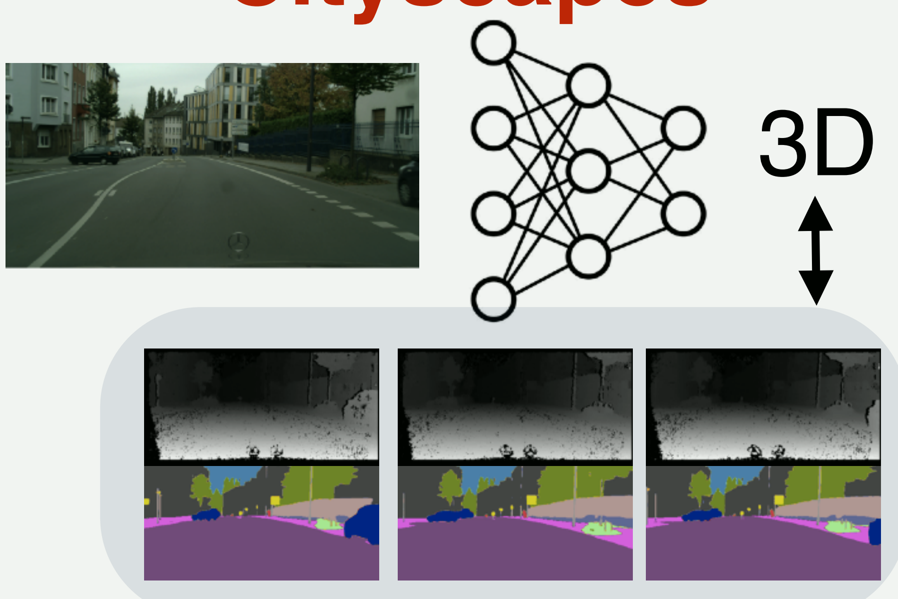
Supervision : Pose + Depth/Mask

PASCAL VOC



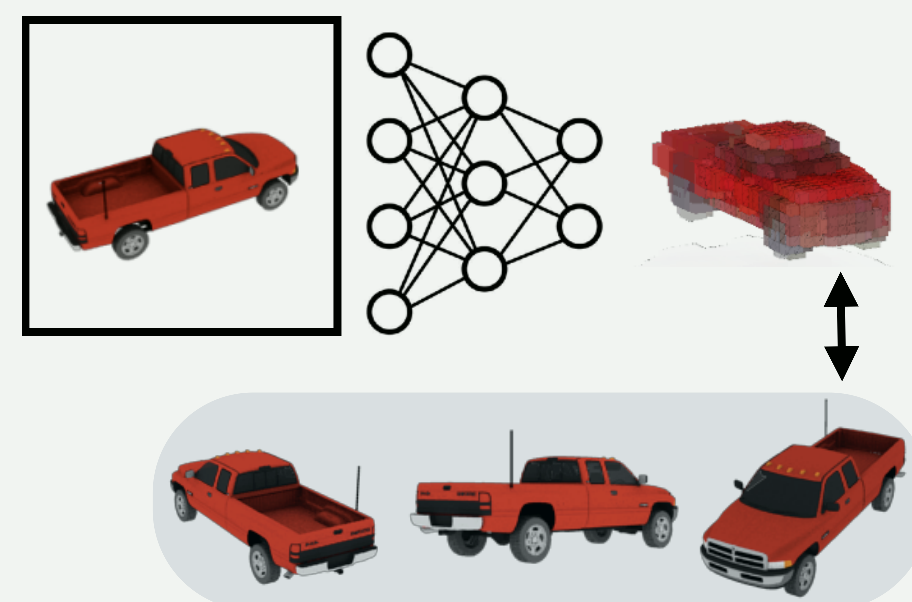
Supervision : Pose + Mask

Cityscapes



Supervision : Ego-motion,
Depth, Semantics

ShapeNet (color supervised)

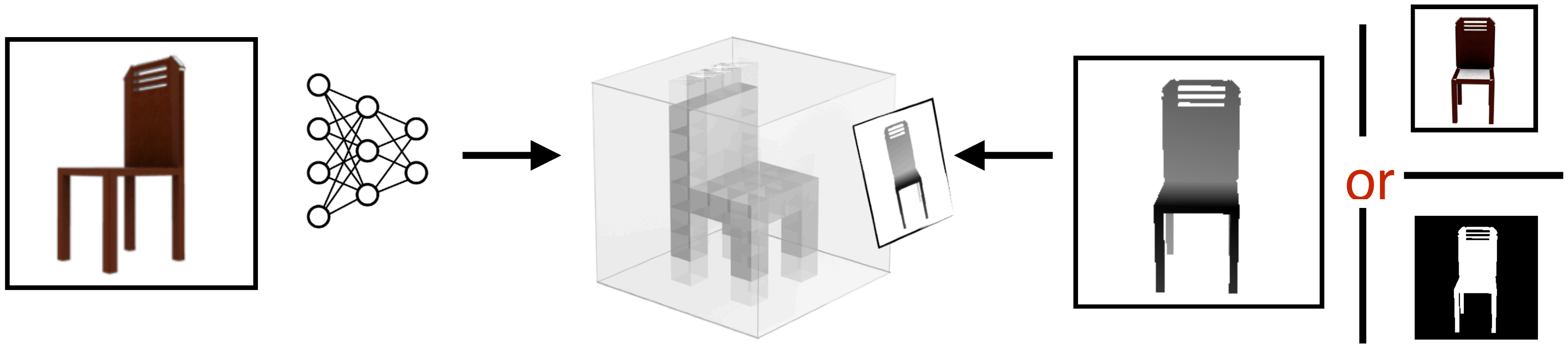


Supervision : Pose + RGB

Conclusion

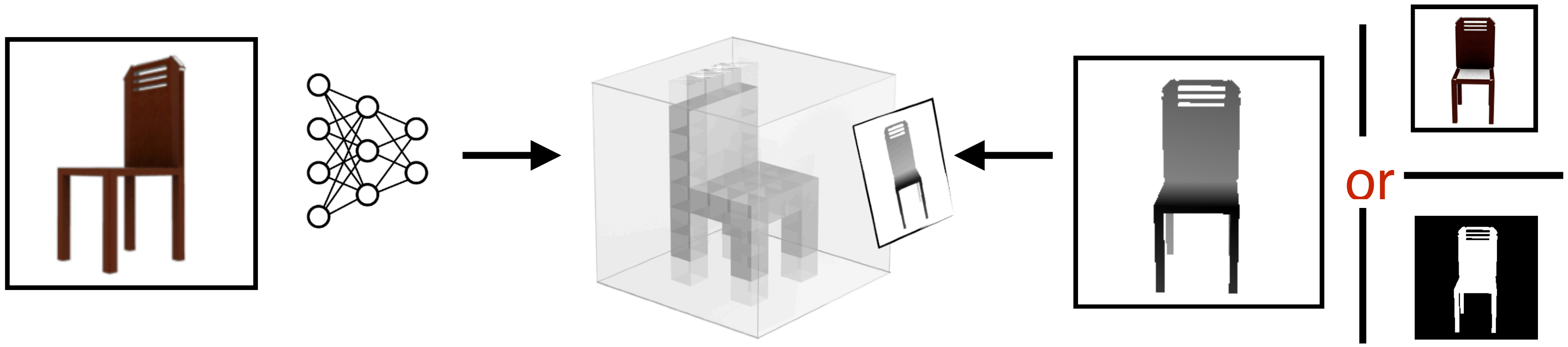
Conclusion

- Learning 3D via Geometric Consistency

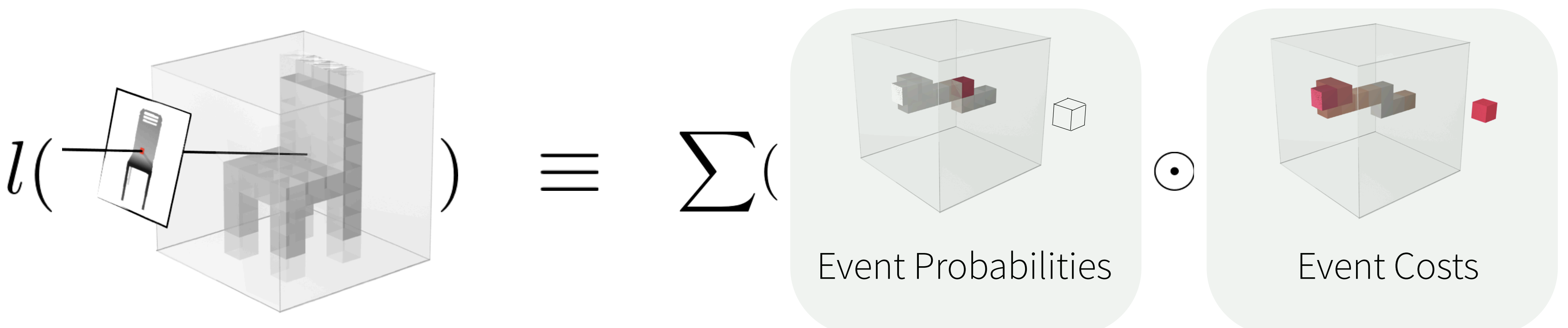


Conclusion

- Learning 3D via Geometric Consistency

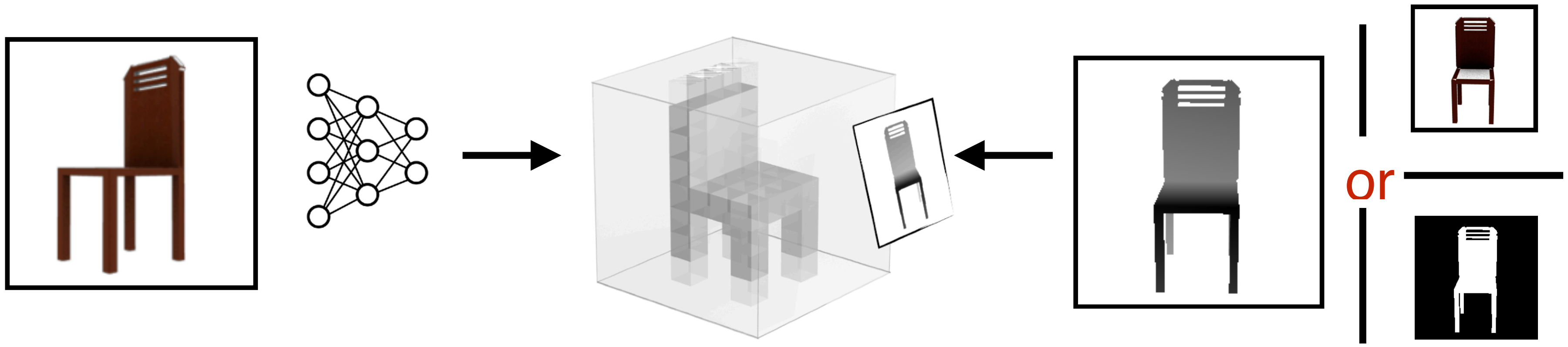


- Differentiable Ray Consistency Formulation

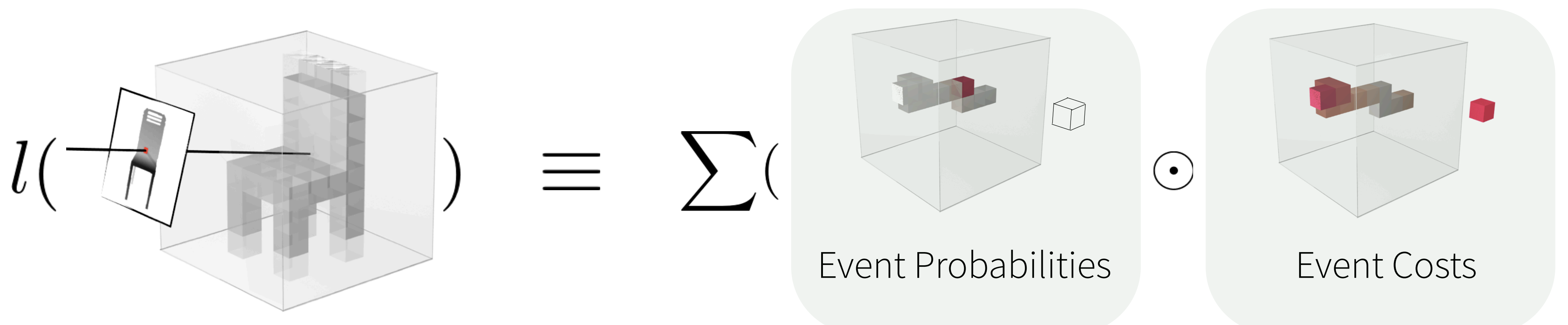


Conclusion

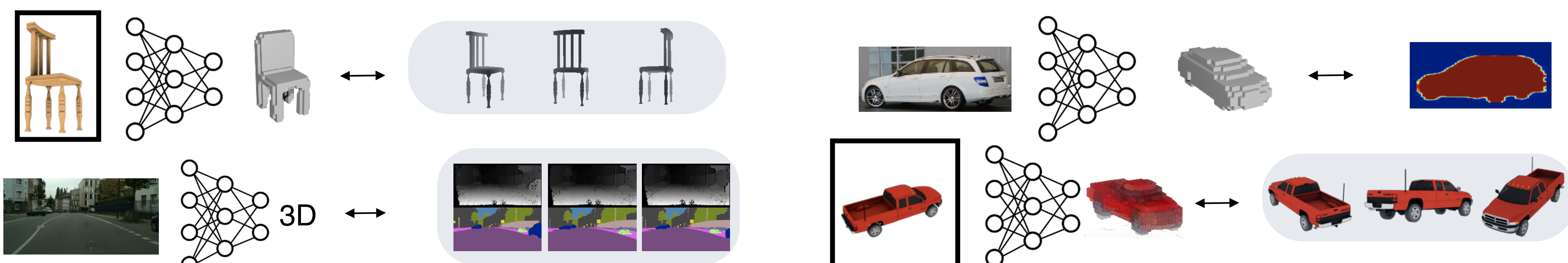
- Learning 3D via Geometric Consistency



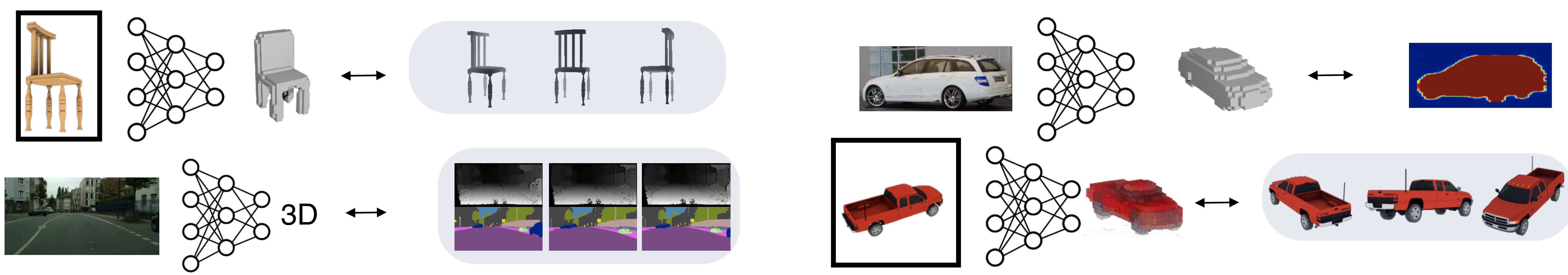
- Differentiable Ray Consistency Formulation



- Applications across scenarios



Thank You



Code : https://github.com/shubhtuls/drc