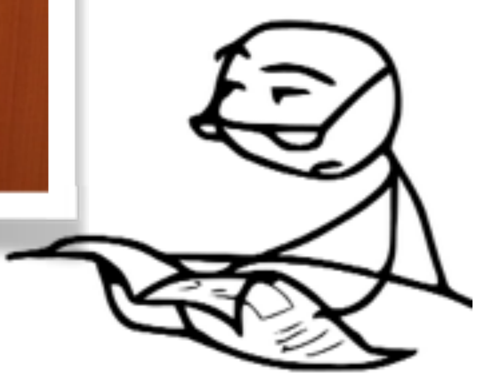
PRINCETON VISION GROUP

PanoContext

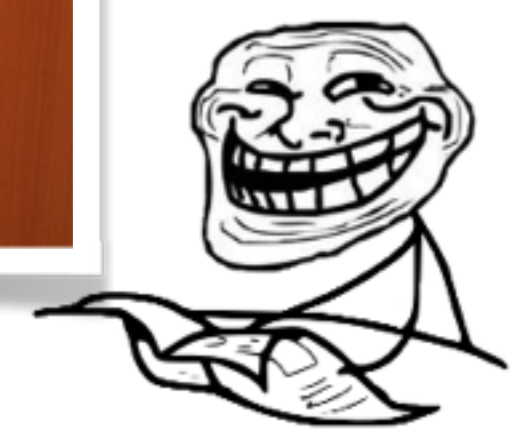
A Whole-room 3D Context Model for Panoramic Scene Understanding

Yinda Zhang Shuran Song Ping Tan[†] Jianxiong Xiao
Princeton University [†]Simon Fraser University

Context is important



Context is important



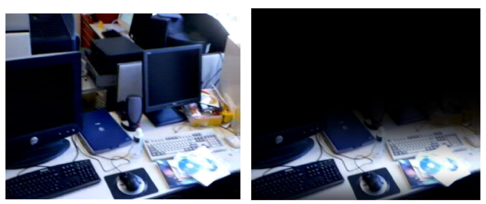
Context is important

Context is important

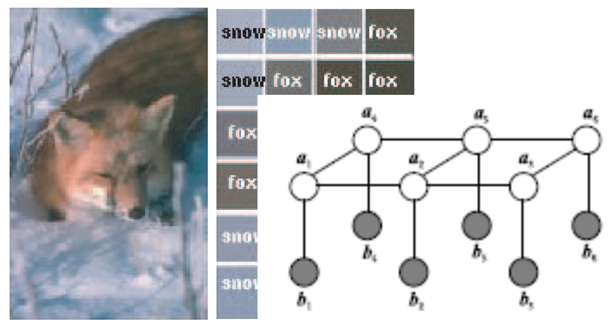
Are you kidding me?

Context models

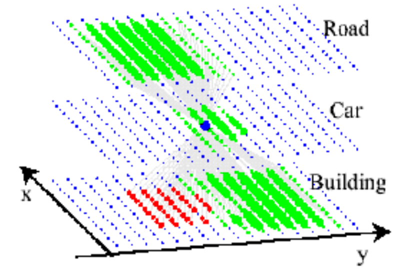
Torralba, Sinha (2001)



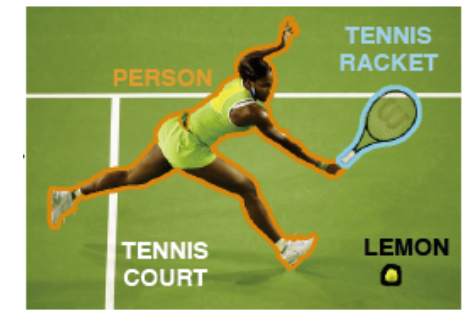
Carbonetto, de Freitas & Barnard (2004)



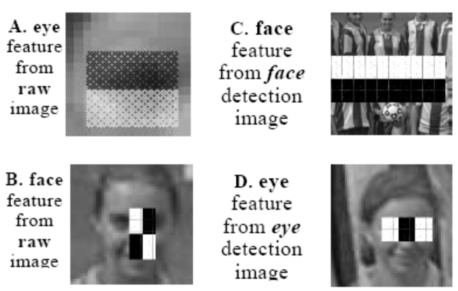
Torralba Murphy Freeman (2004)



Rabinovich et al (2007)

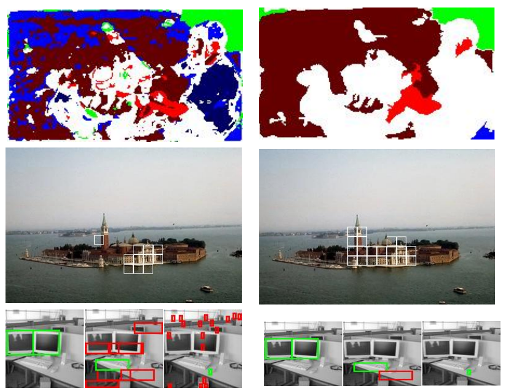


Fink & Perona (2003)

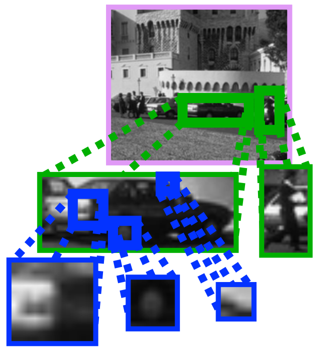


Heitz and Koller (2008)

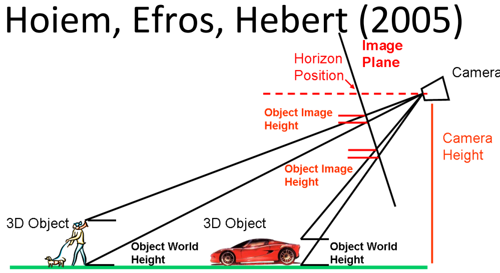
Kumar, Hebert (2005)



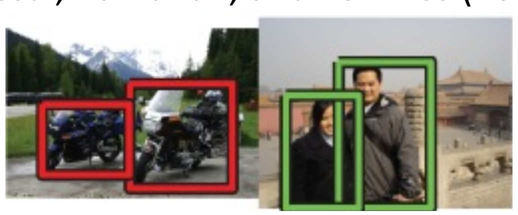
Sudderth, Torralba, Wilsky, Freeman (2005)



Hoiem, Efros, Hebert (2005)



Desai, Ramanan, and Fowlkes (2009)



	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbik
BB	.339	.381	.067	.099	.278	.229	.331	.146	.153	.119	.124	.066	.322	.366
context	.351	.402	.117	.114	.284	.251	.334	.188	.166	.114	.087	.078	.347	.395

DPM on PASCAL VOC [Felzenszwalb et al.]

Improvement on PASCAL <1.5%

What’s the problem in context?

Game: What is this object?

Please speak out when you recognize the object!

What is this object?

What is this object?

What is this object?

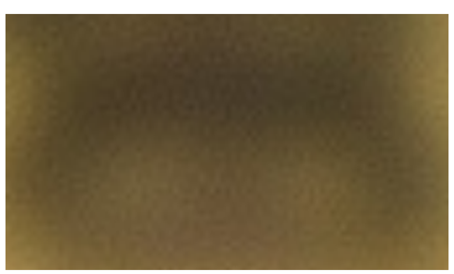
What is this object?

What is this object?

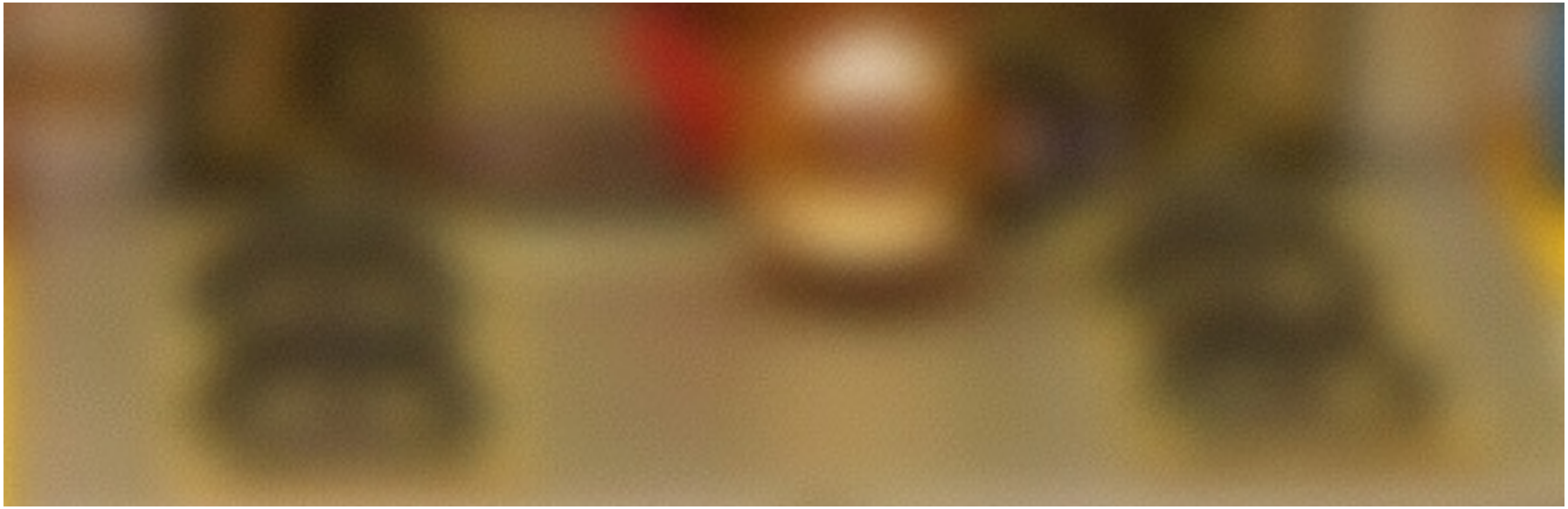
What is this object?

What is this object?

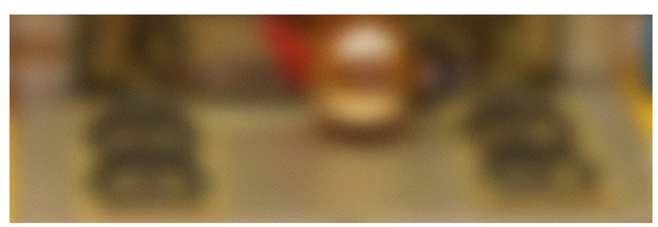
What is this object?



What is this object?

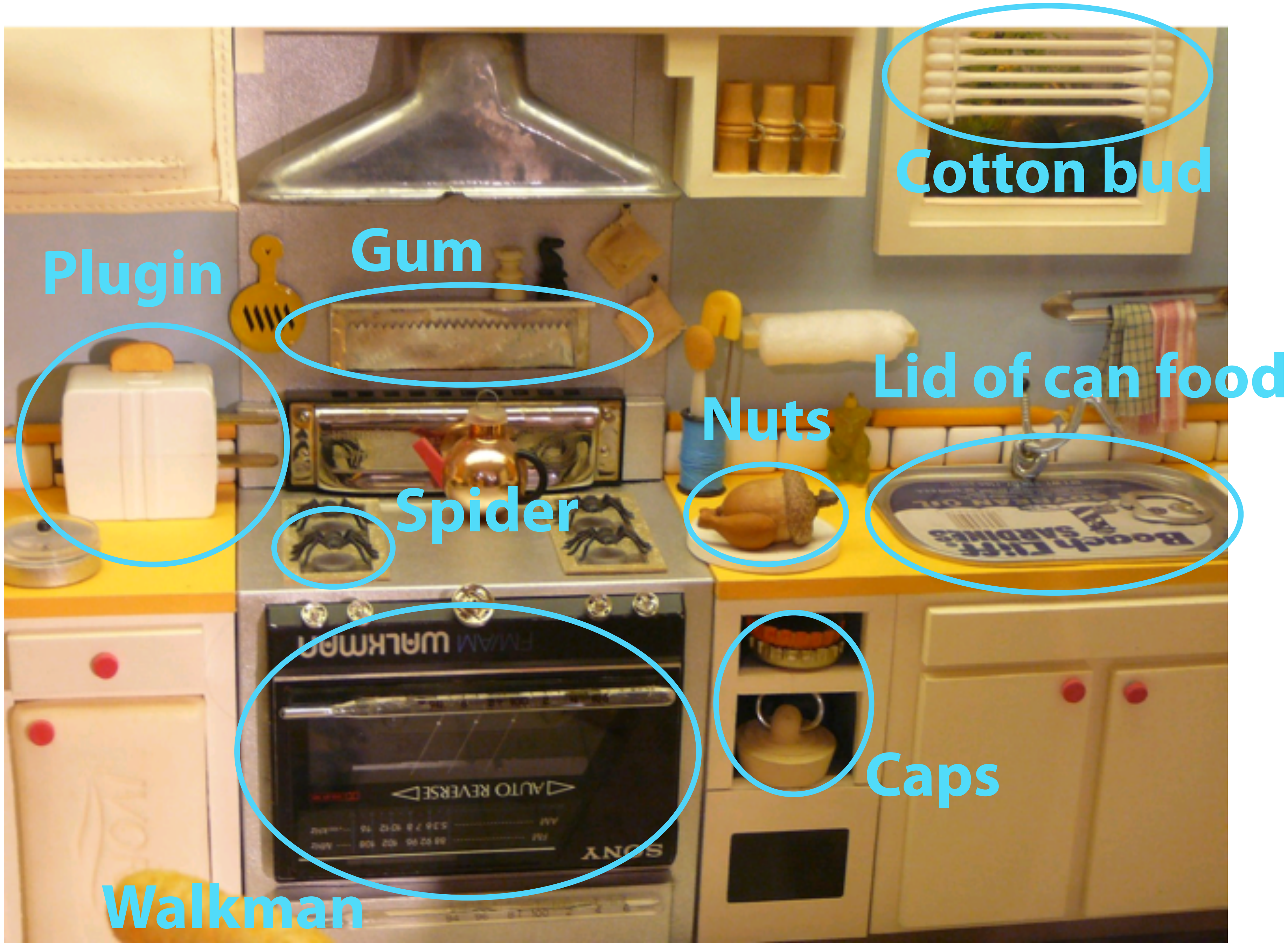


What is this object?



What is this object?

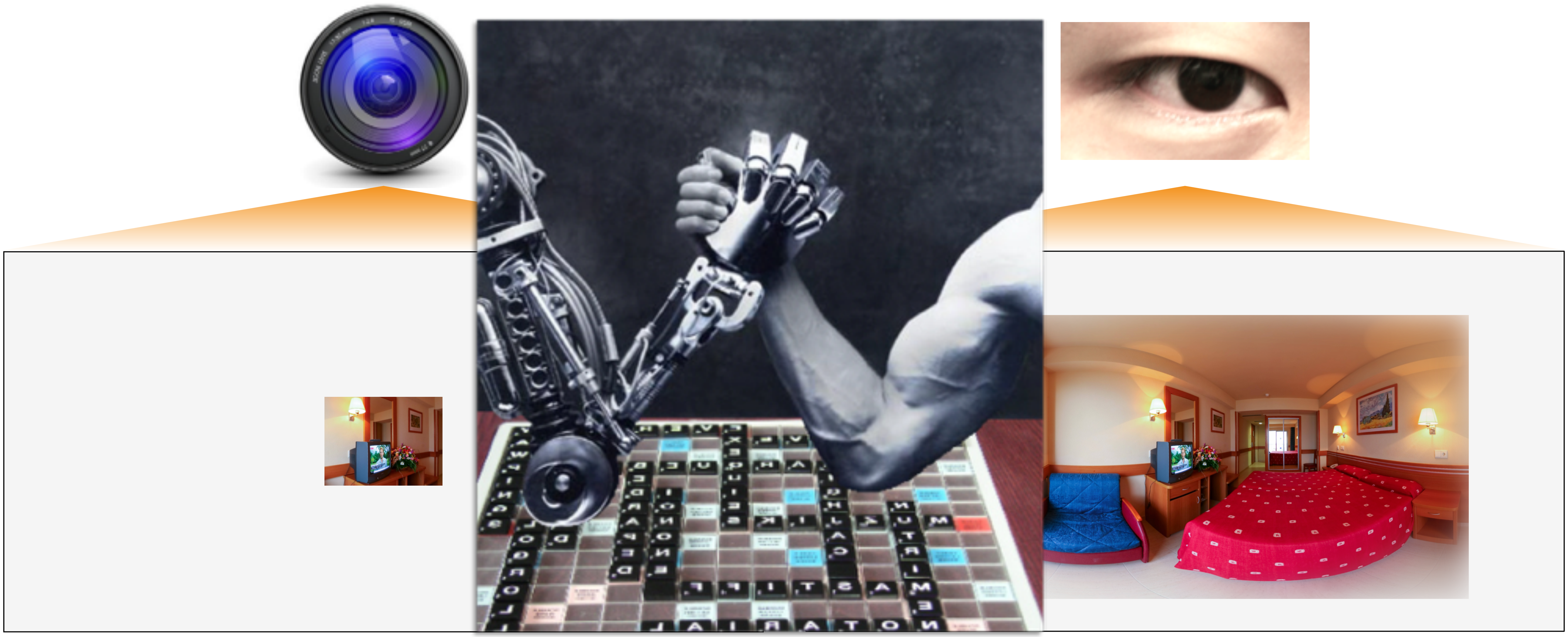
What is this object?



Look-Alikes by Joan Steiner

Small FOV, you will miss a lot!

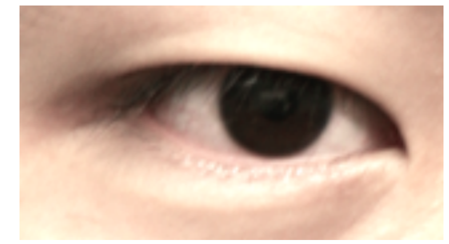
How does large FOV help?



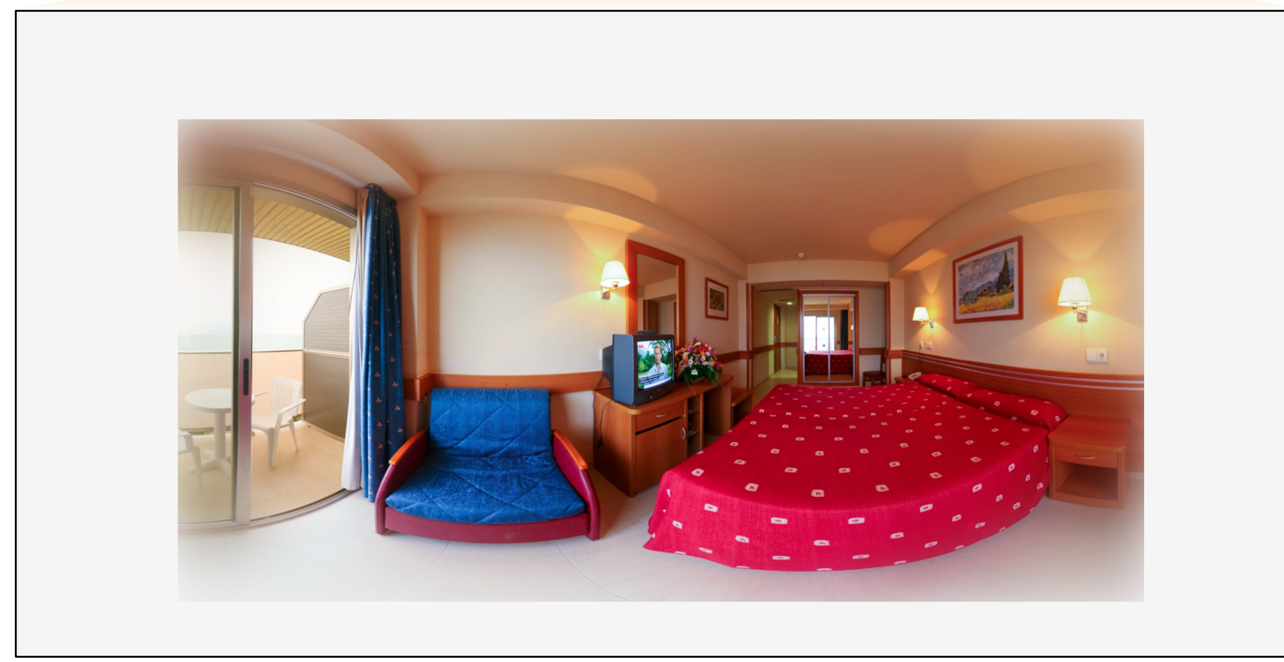
What a camera sees: 54º What your eyes see: 270º

1. Is there a bed in the room? → Unpredictable visibility
2. Relation between bed and TV? → Less interaction

How does large FOV help?



What a camera sees: ~~54°~~ 360°

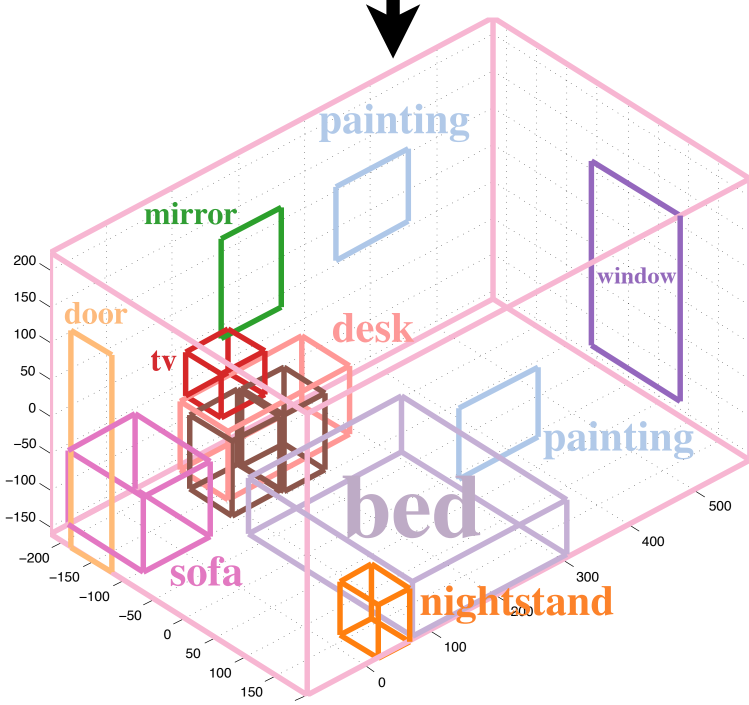


What your eyes see: 270°

PanoContext

Input: Panorama

Output: 2D projected result



Output: 3D model

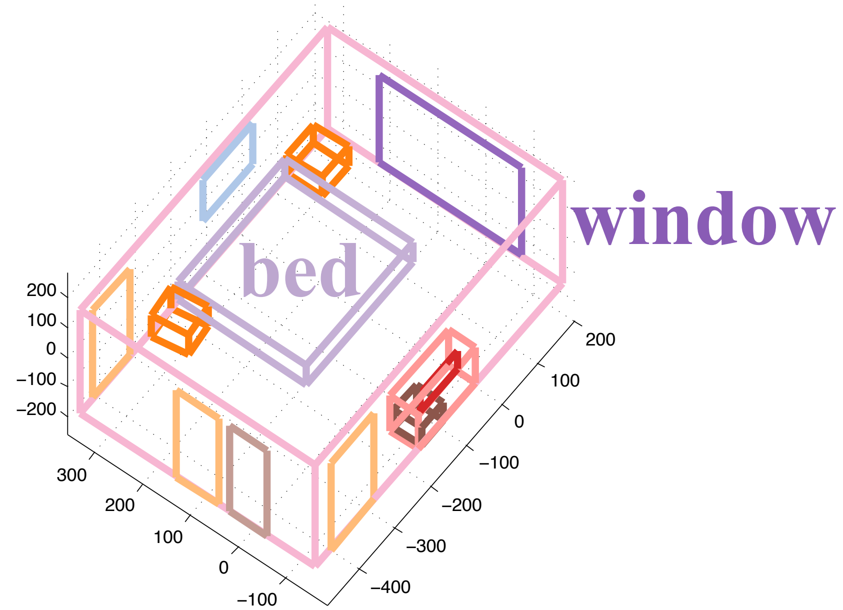
Output: 3D room exploration

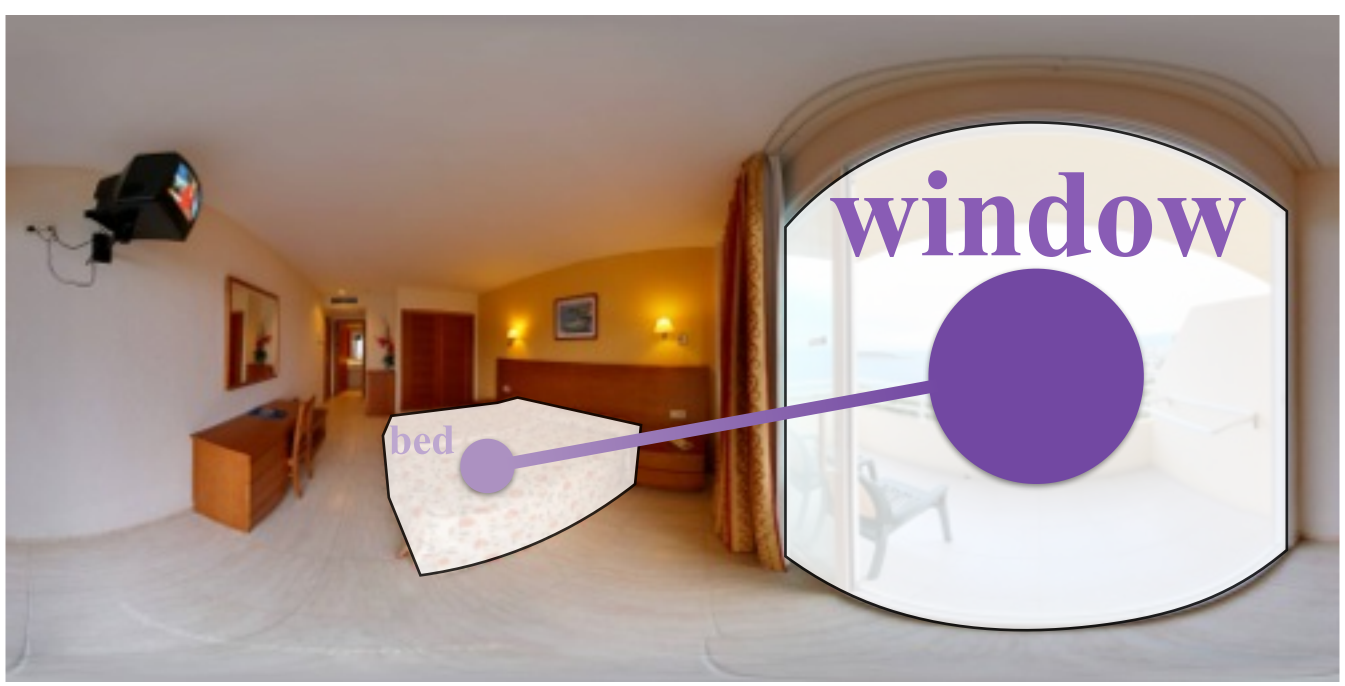
PanoContext

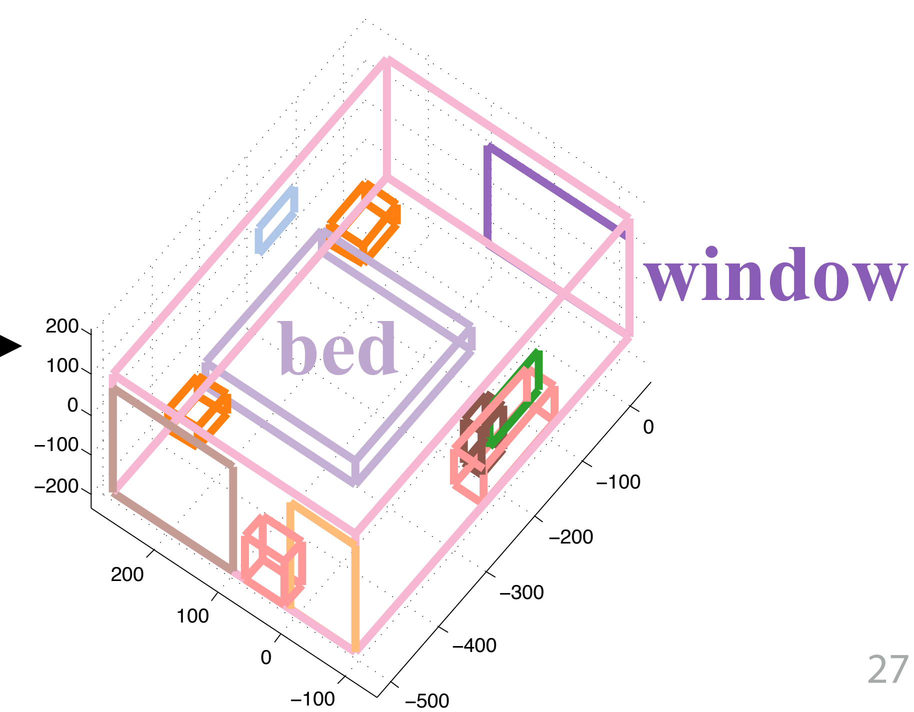
A 3D whole-room non-parametric context model

PanoContext

A 3D whole-room non-parametric context model







PanoContext

A 3D whole-room non-parametric context model

PanoContext

A 3D whole-room non-parametric context model

PanoContext

A 3D **whole-room** non-parametric **context model**

PanoContext

A 3D whole-room non-parametric context model

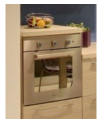
PanoContext

A 3D **whole-room** non-parametric **context model**

PanoContext

A 3D **whole-room** non-parametric **context model**

PanoContext

A 3D whole-room non-parametric context model

PanoContext

A **3D whole-room** non-parametric **context model**

place-centric

[Xiao et al. 2012, 2013]

view-centric

Majority of the Literature

PanoContext

A 3D whole-room non-parametric context model

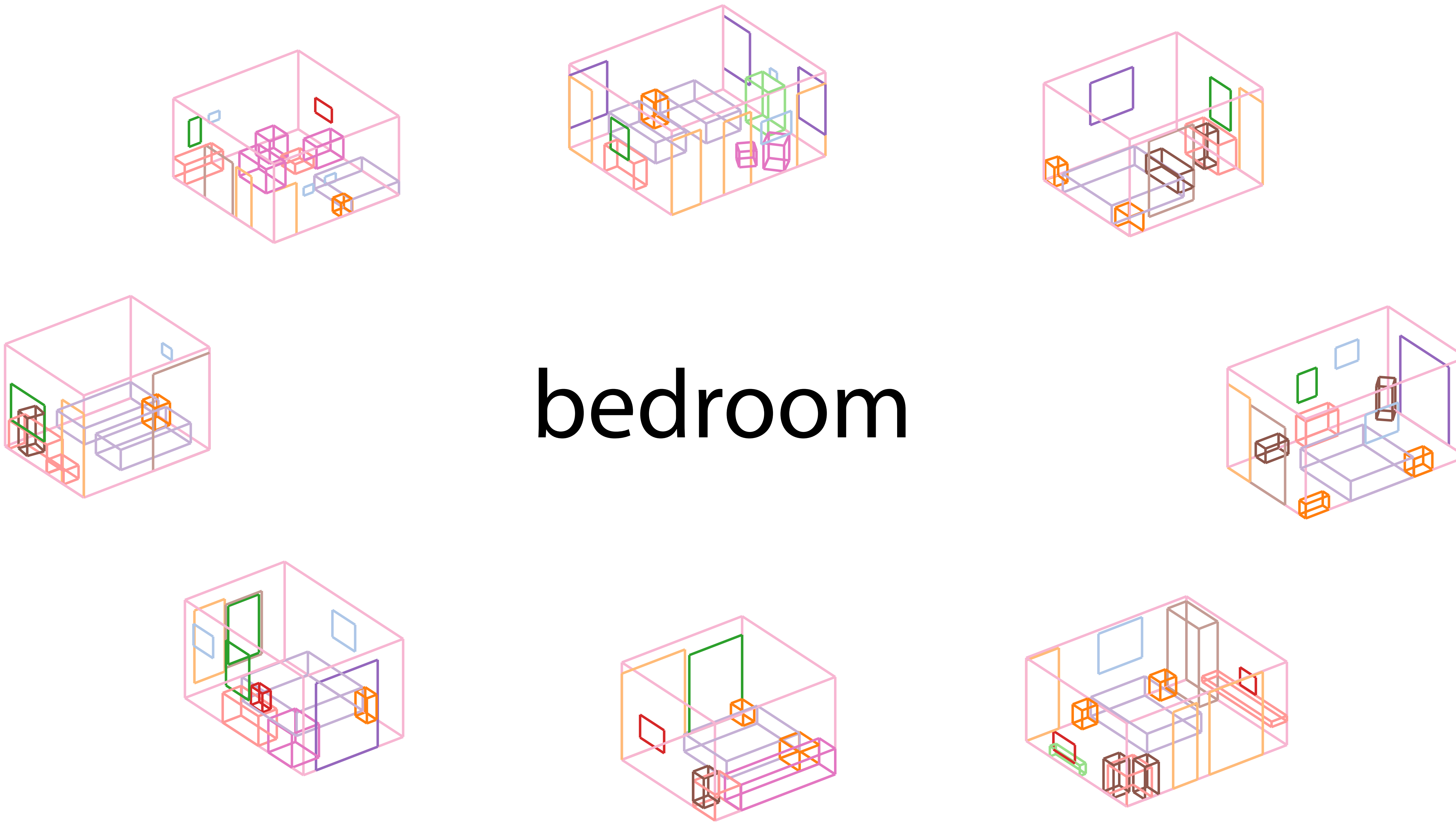
PanoContext

A 3D whole-room non-parametric context model

bedroom

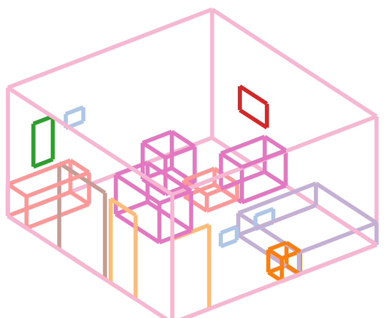
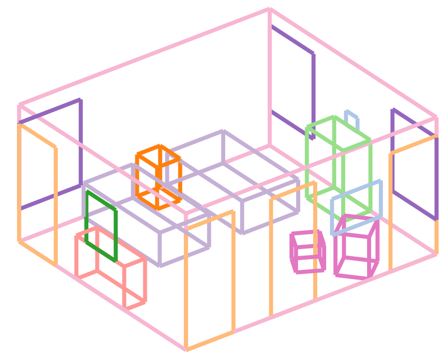
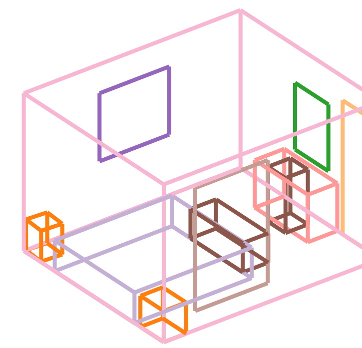
PanoContext

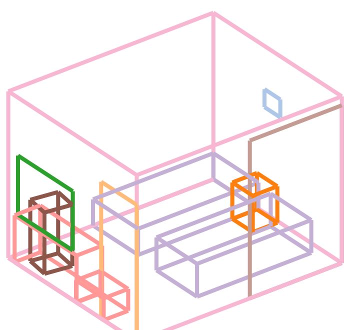
A 3D whole-room **non-parametric context model**



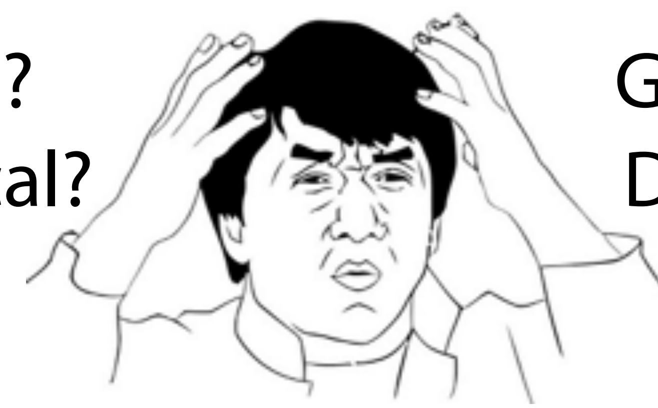
PanoContext

A 3D whole-room non-parametric context model

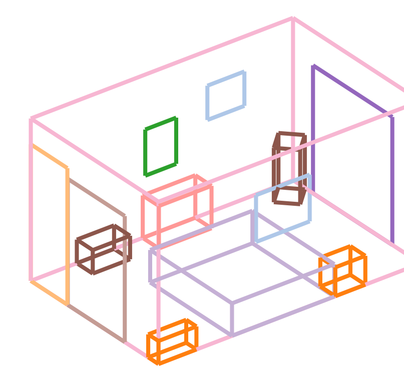




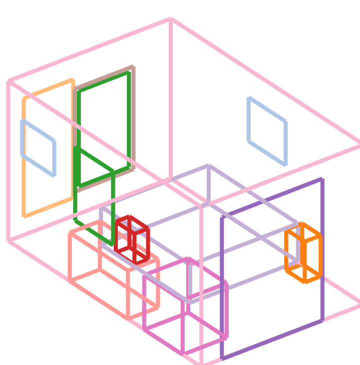
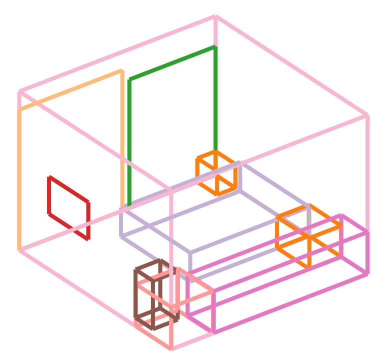
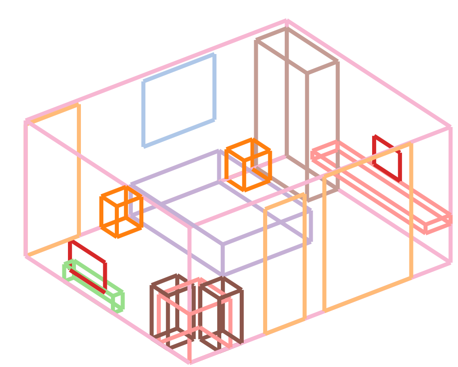


Pairwise?
Hierarchical?
….?



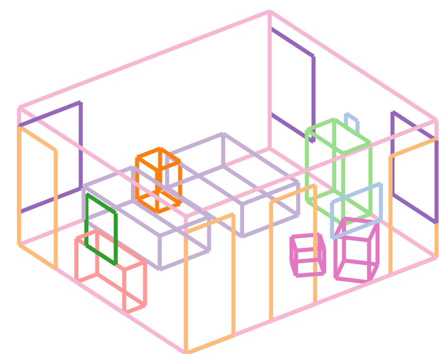
Gaussian?
Dirichlet?
….?

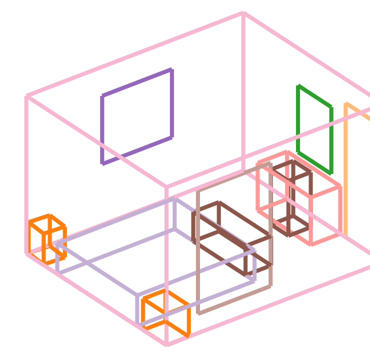


PanoContext

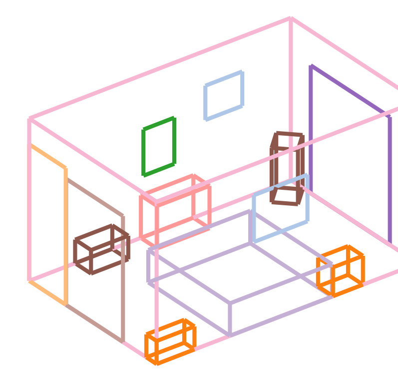
A 3D whole-room non-parametric context model

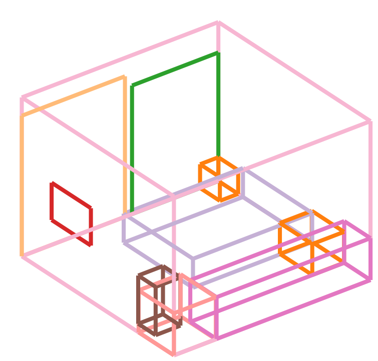


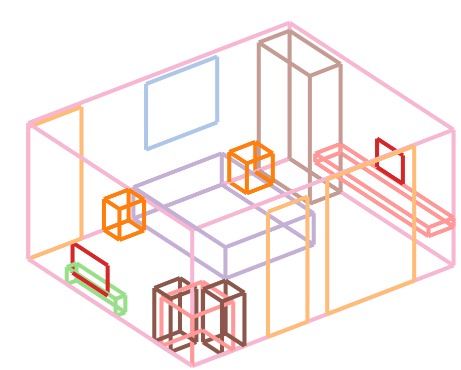


Ultimate solution for
all problems in the world:

Nearest Neighbor

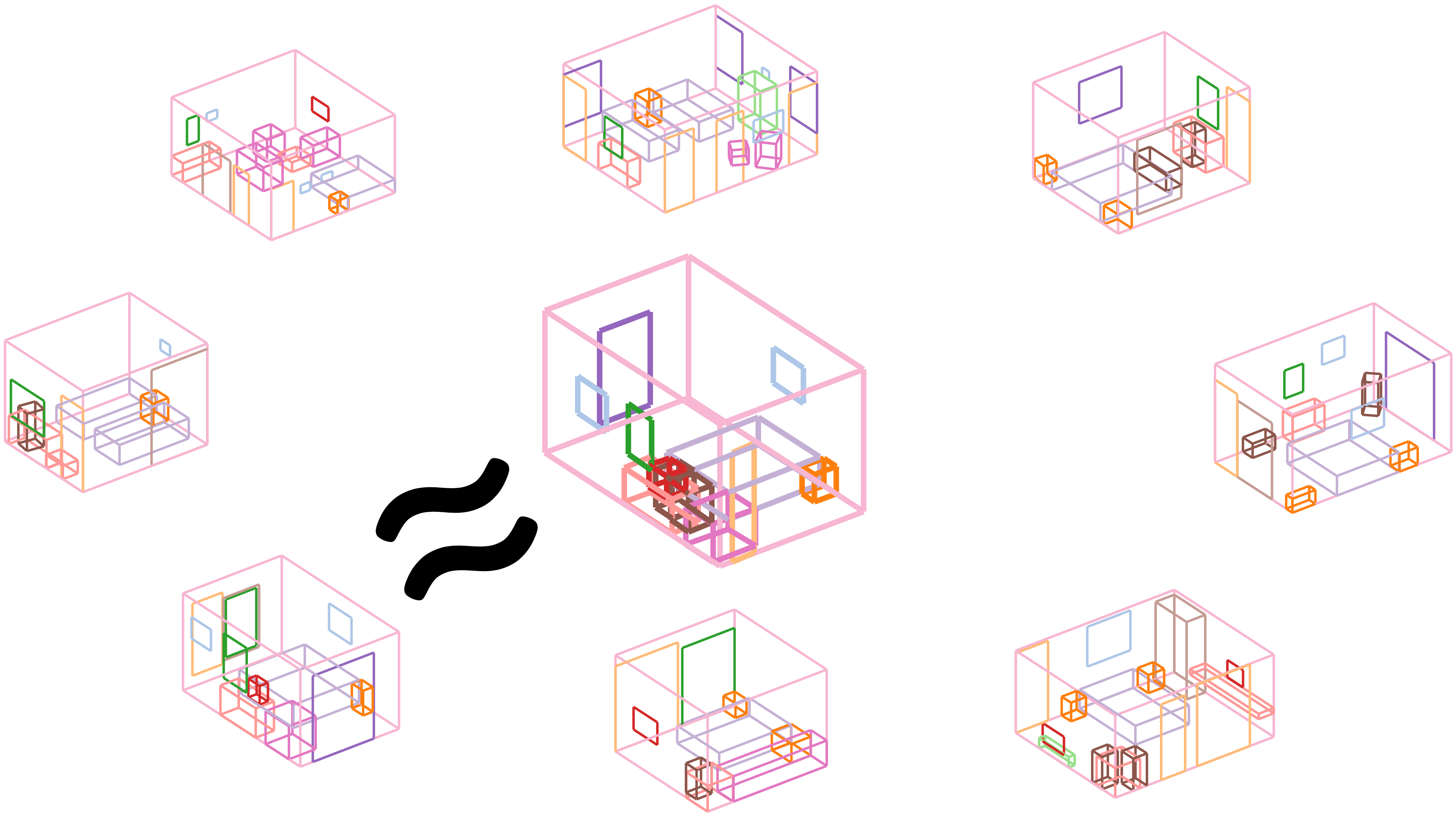






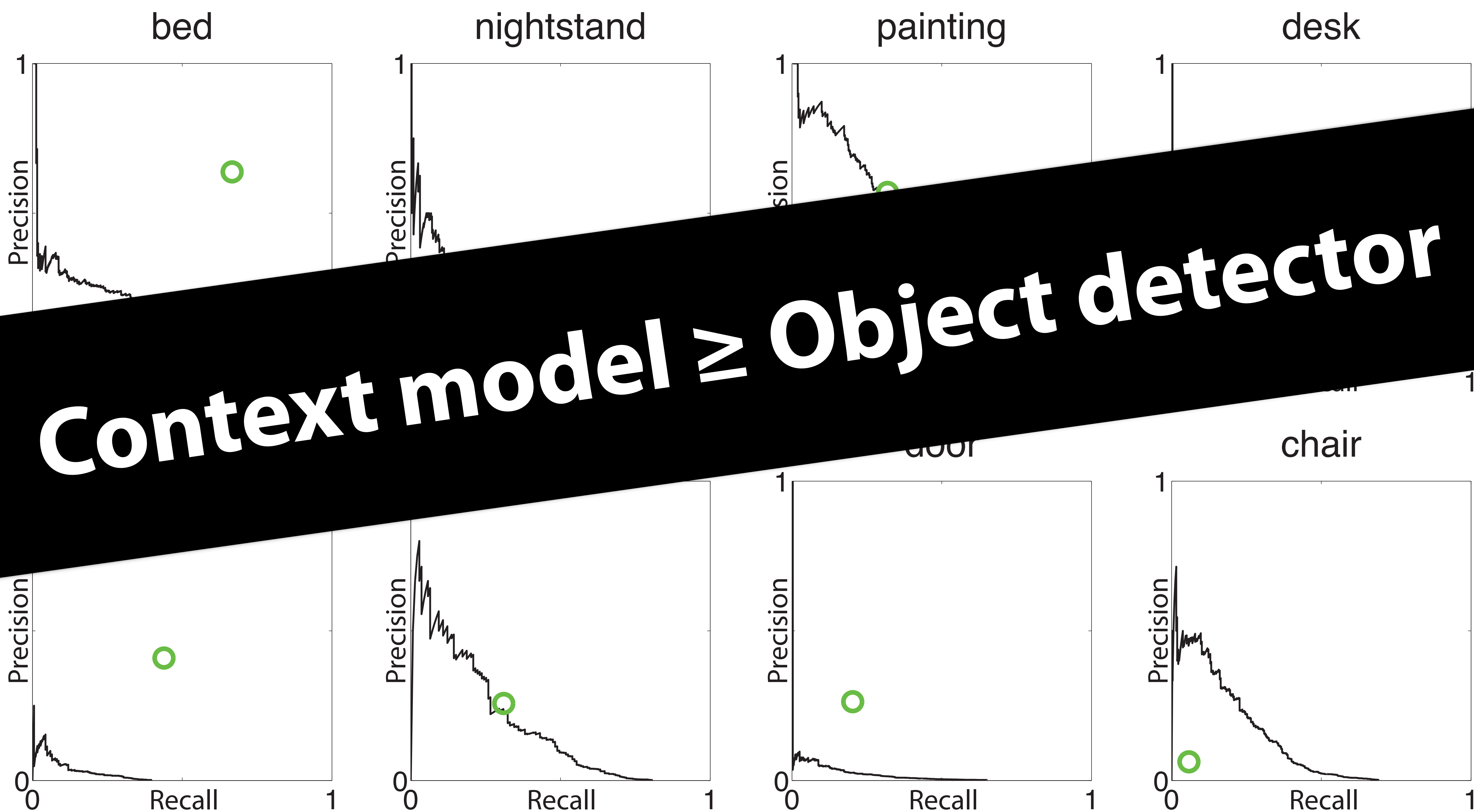
PanoContext

A 3D whole-room non-parametric context model



Our model: 3D whole-room context

DPM: 2D local image appearance

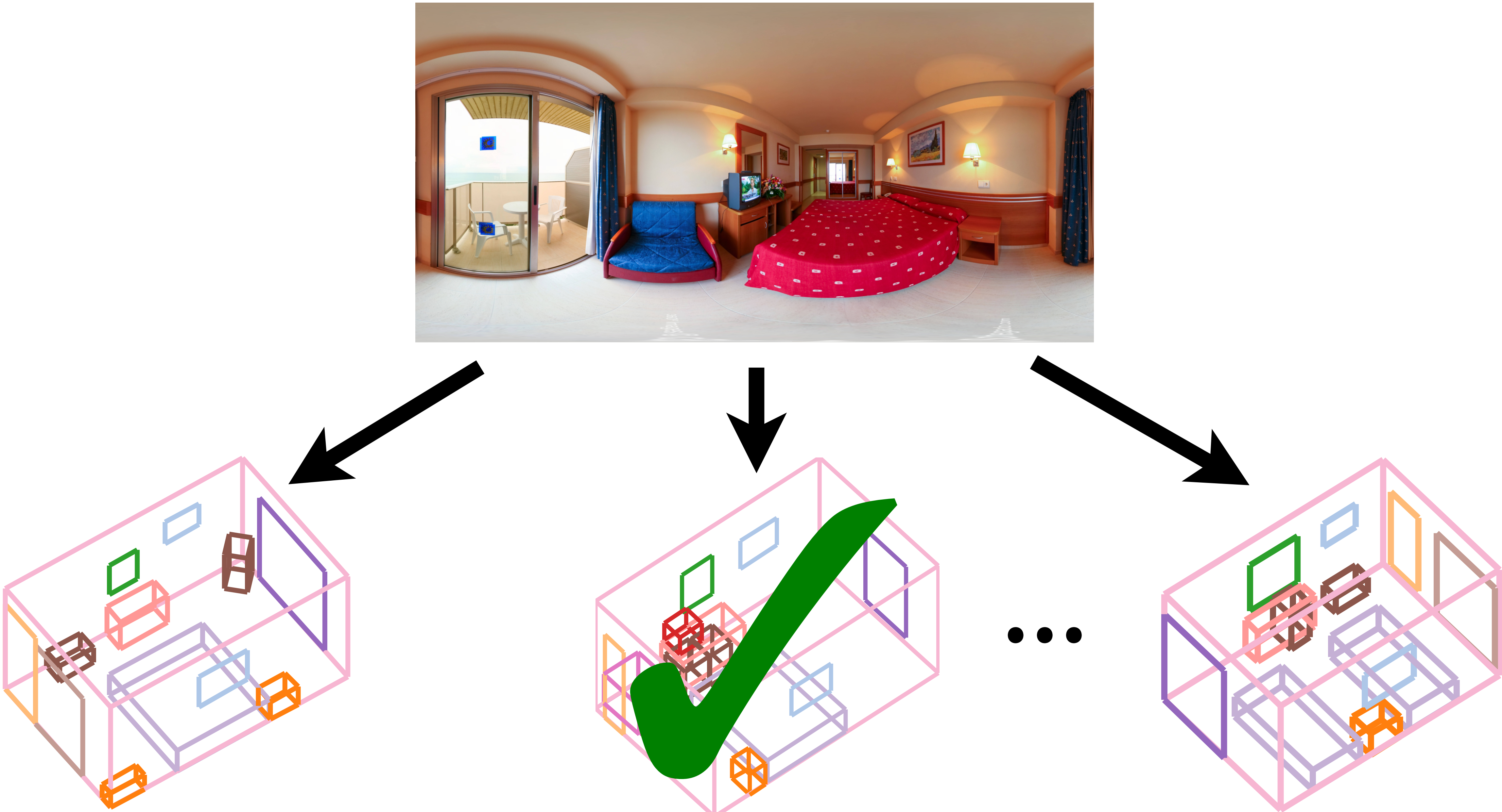


Context model ≥ Object detector

Algorithm

Algorithm

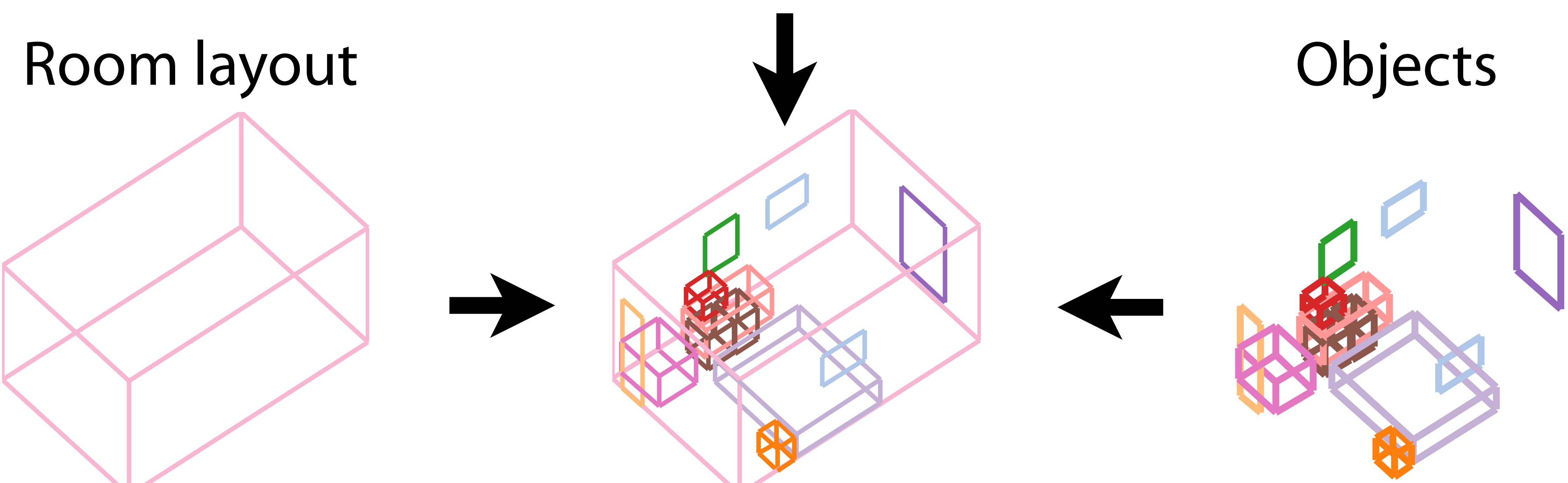
Step 1: Generate a pool of hypotheses
Step 2: Choose the best hypothesis



Algorithm

Step 1: Generate a pool of hypotheses

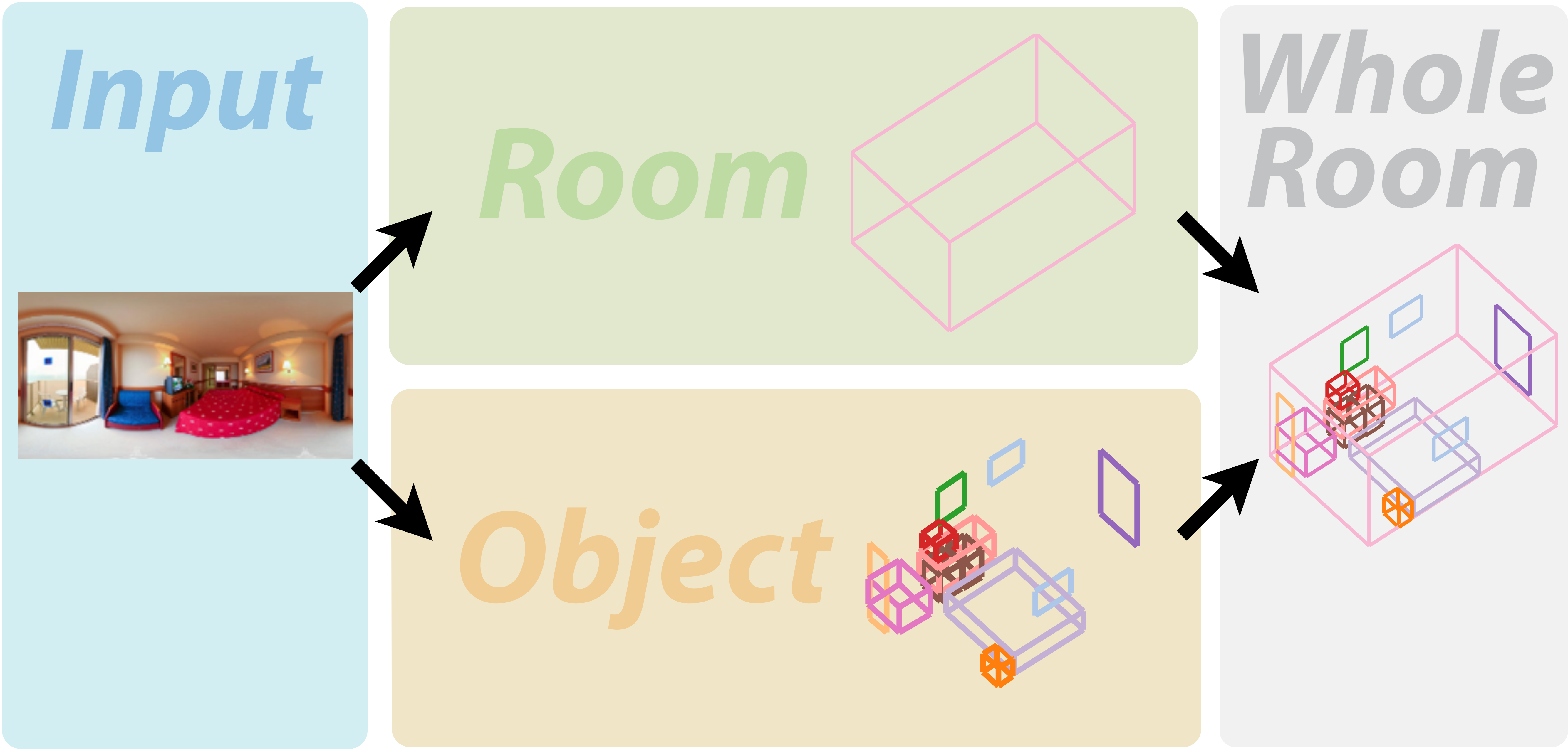
Step 2: Choose the best hypothesis



Algorithm

Step 1: Generate a pool of hypotheses

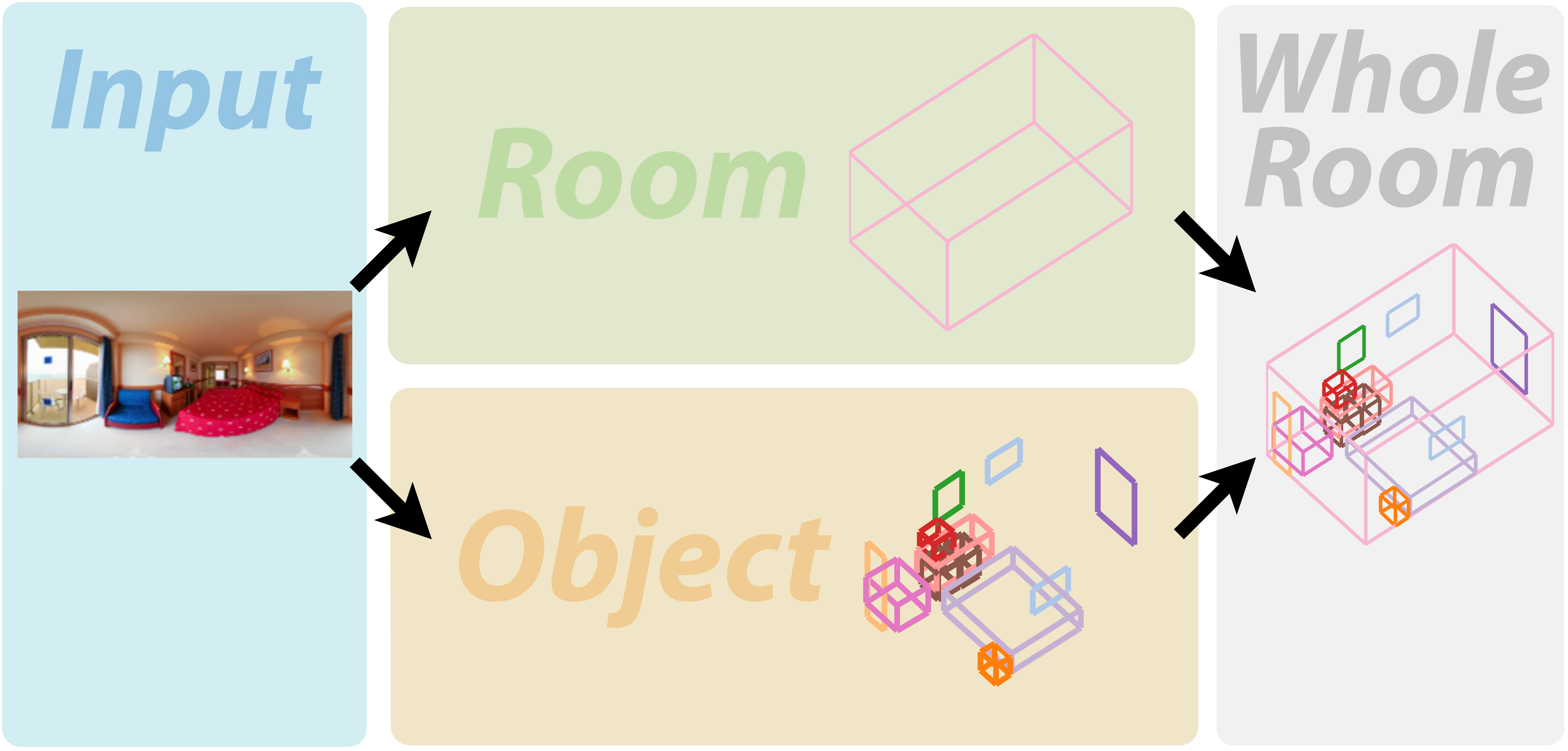
Step 2: Choose the best hypothesis



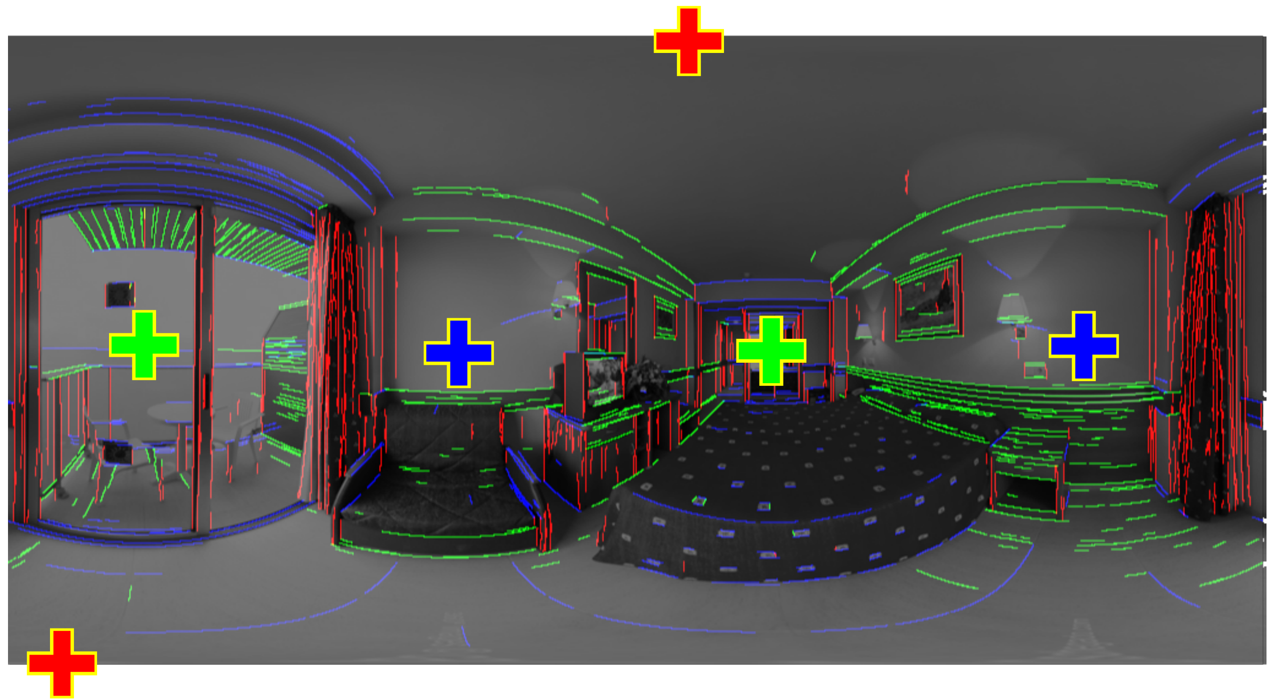
Algorithm

Step 1: Generate a pool of hypotheses

Step 2: Choose the best hypothesis



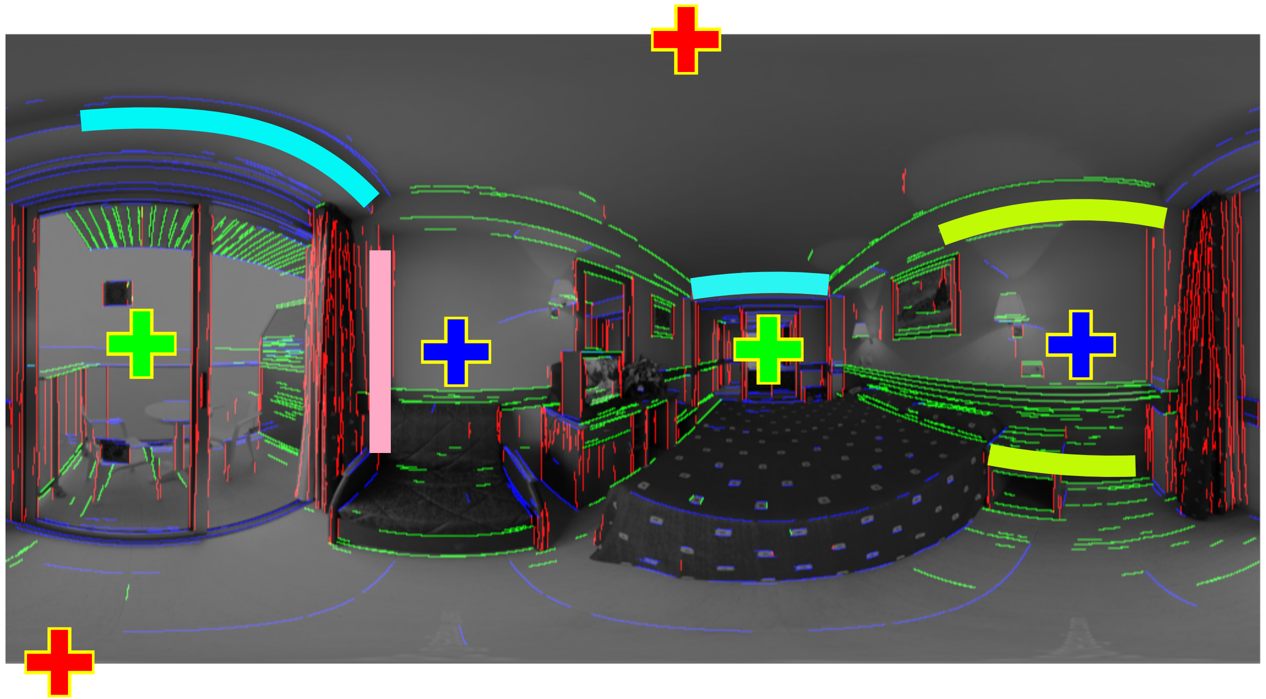
Room layout hypothesis



Line segments

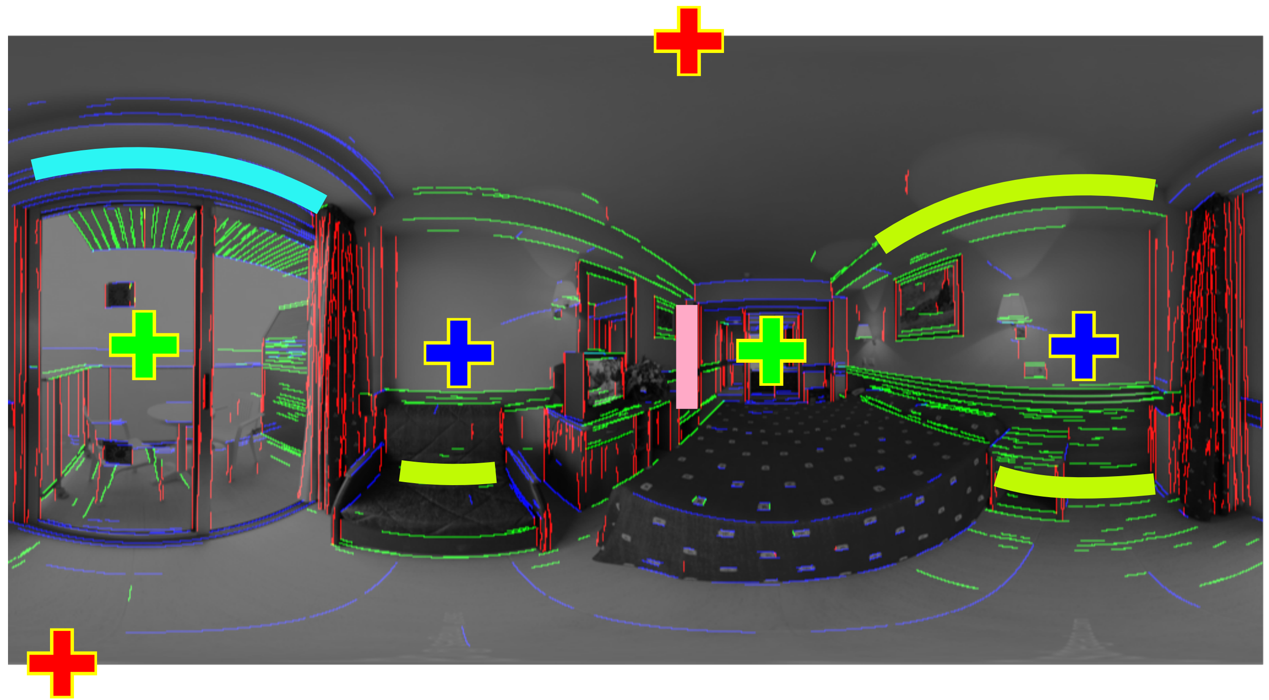
Vanishing points

Room layout hypothesis



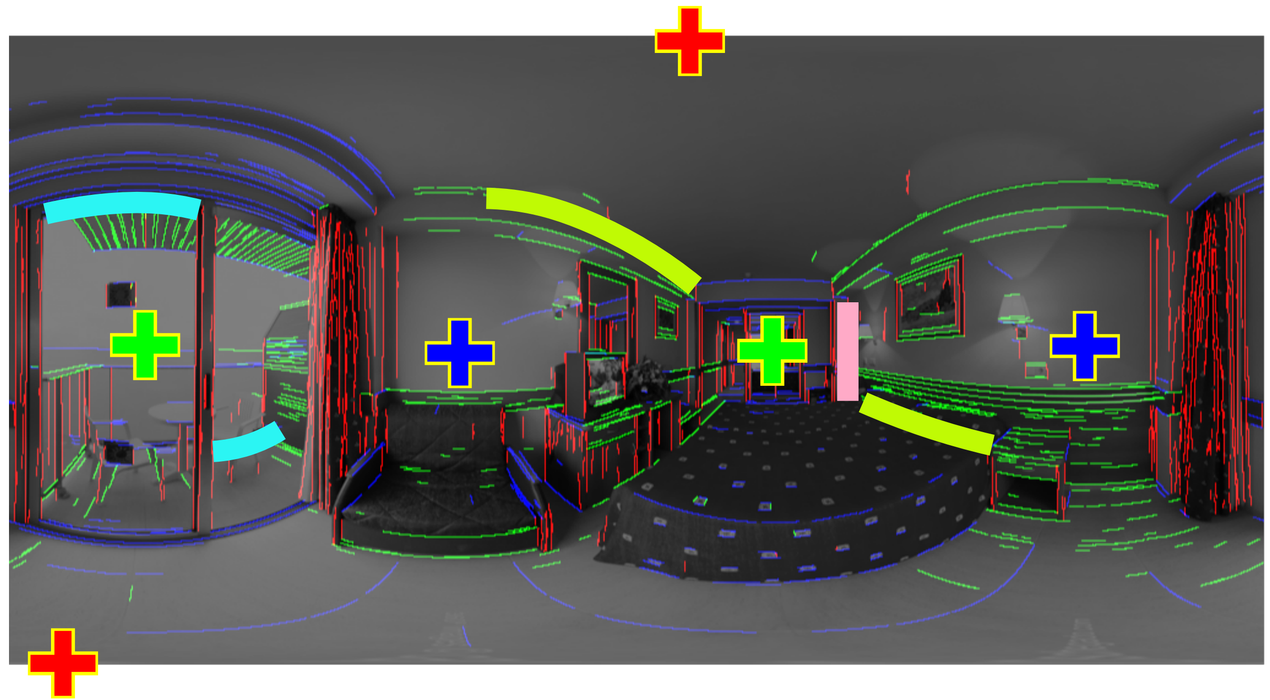
Sample 5 line segments to generate a room layout

Room layout hypothesis



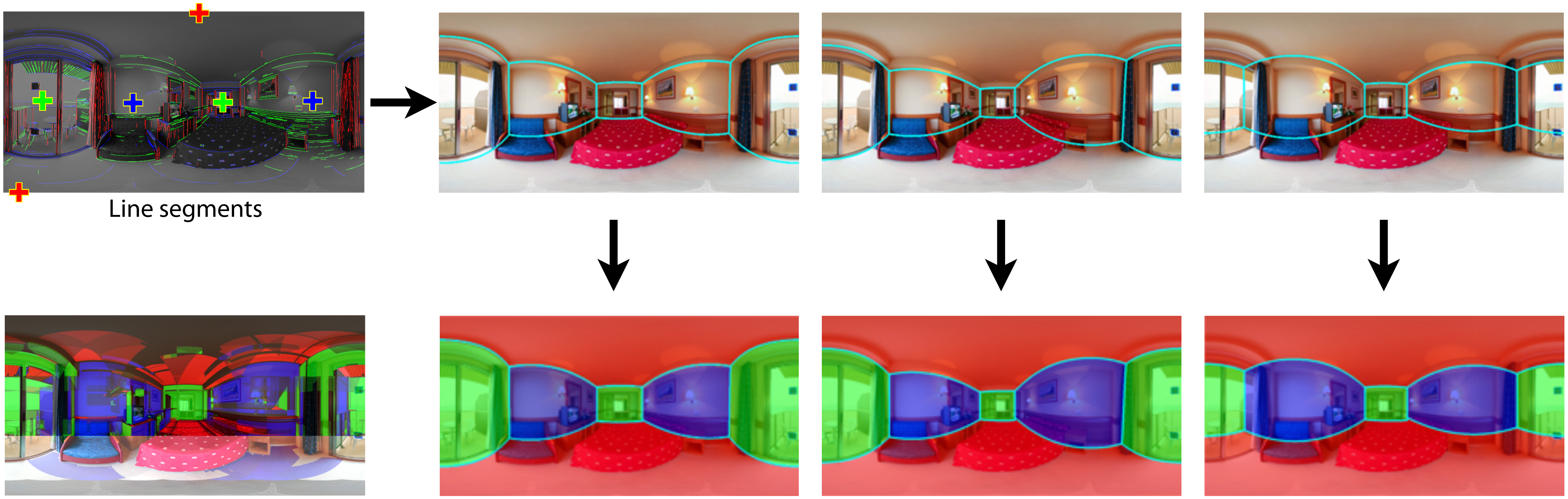
Sample 5 line segments to generate a room layout

Room layout hypothesis



Sample 5 line segments to generate a room layout

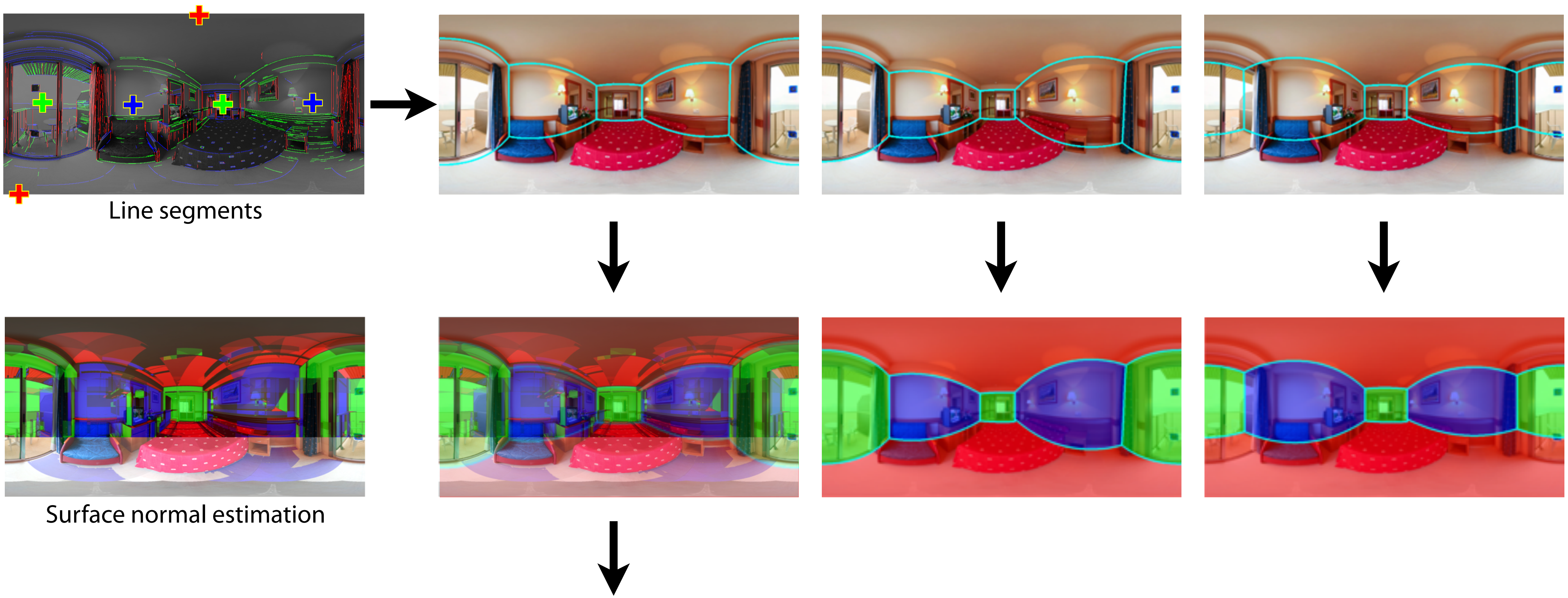
Room layout hypothesis



Line segments

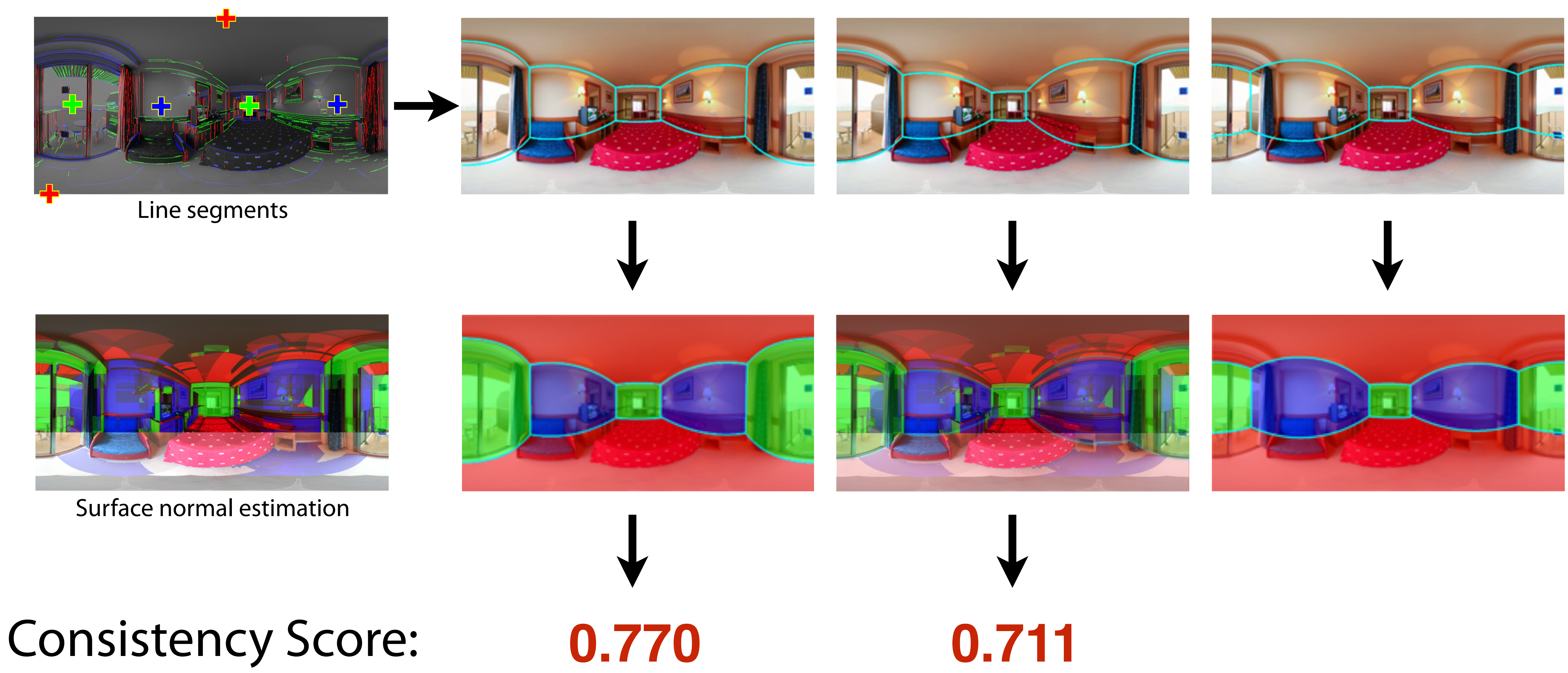
Surface normal estimation

Room layout hypothesis

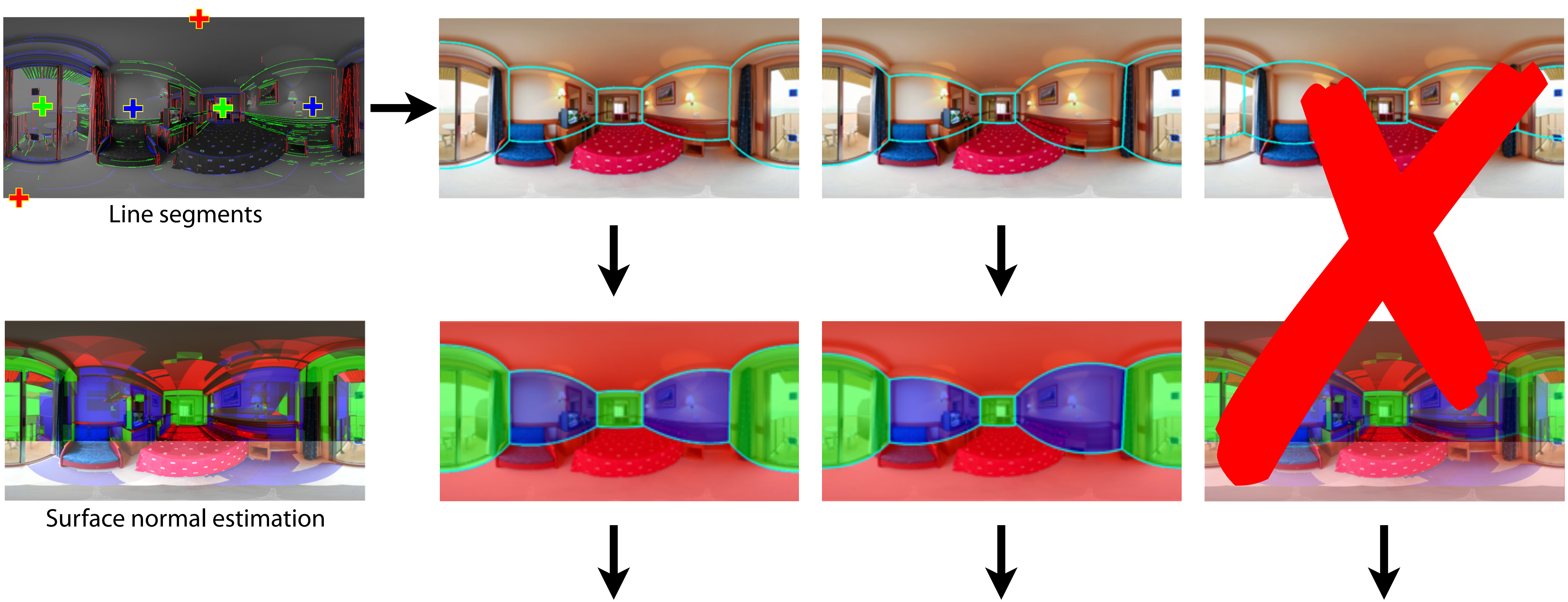


Consistency Score: **0.770**

Room layout hypothesis

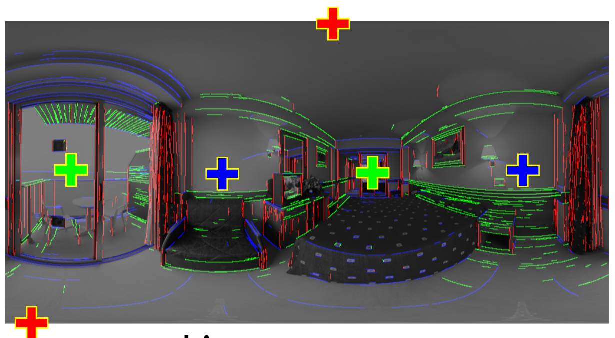


Room layout hypothesis

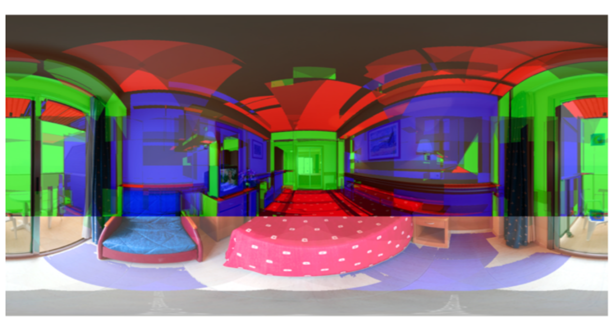


Consistency Score: **0.770** **0.711** **0.504**

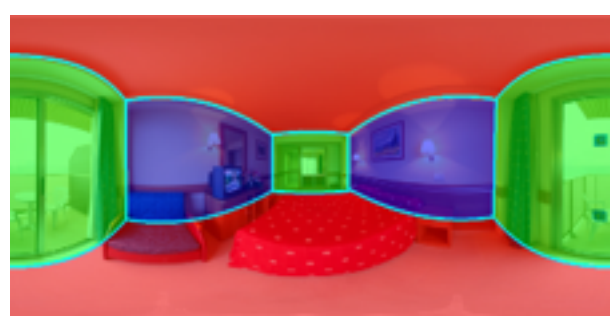
Room layout hypothesis

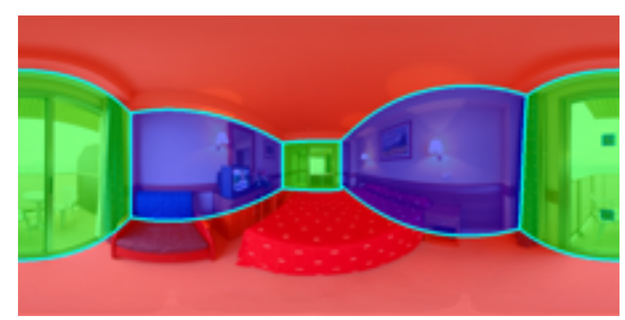


Line segments



Surface normal estimation



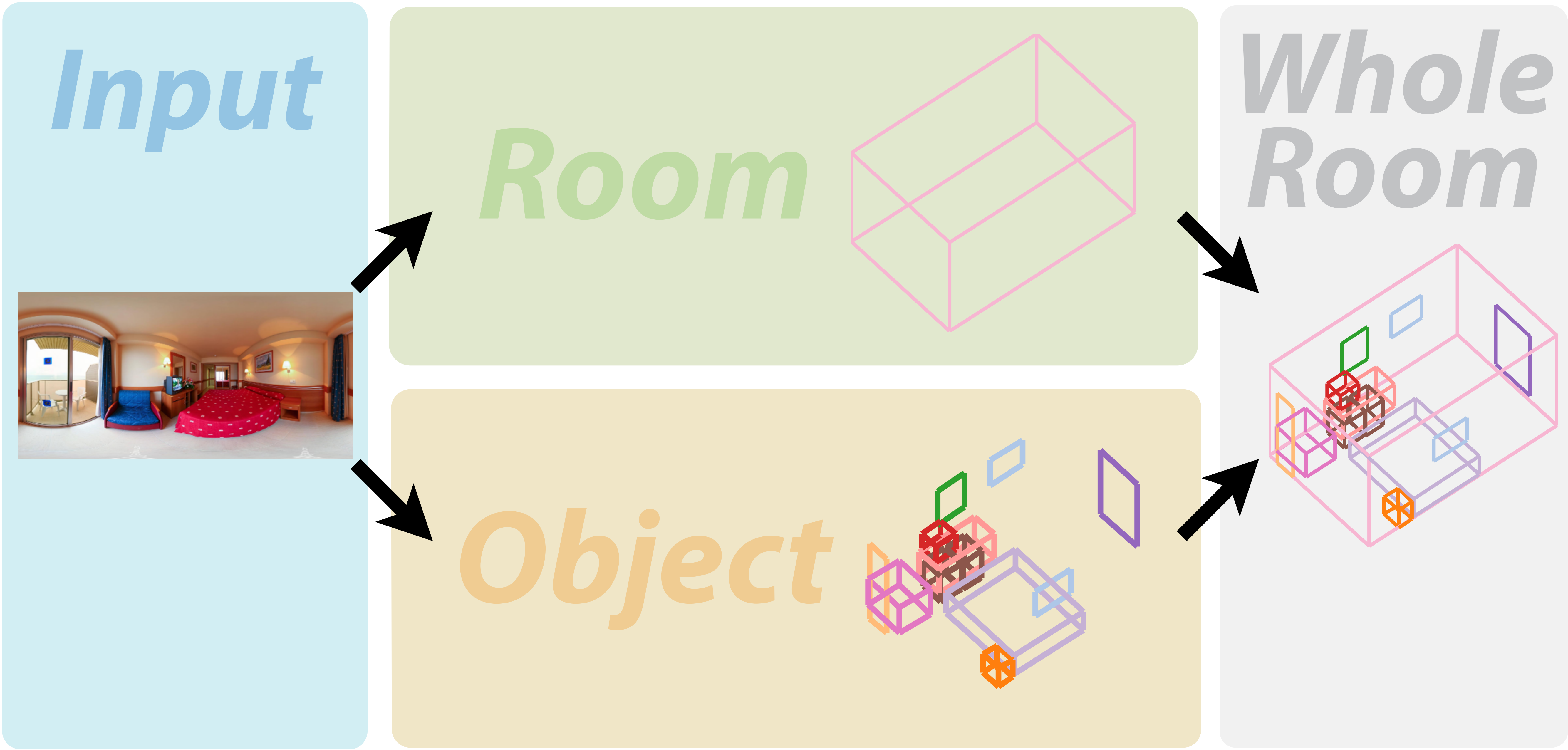


Consistency Score: **0.770** **0.711**

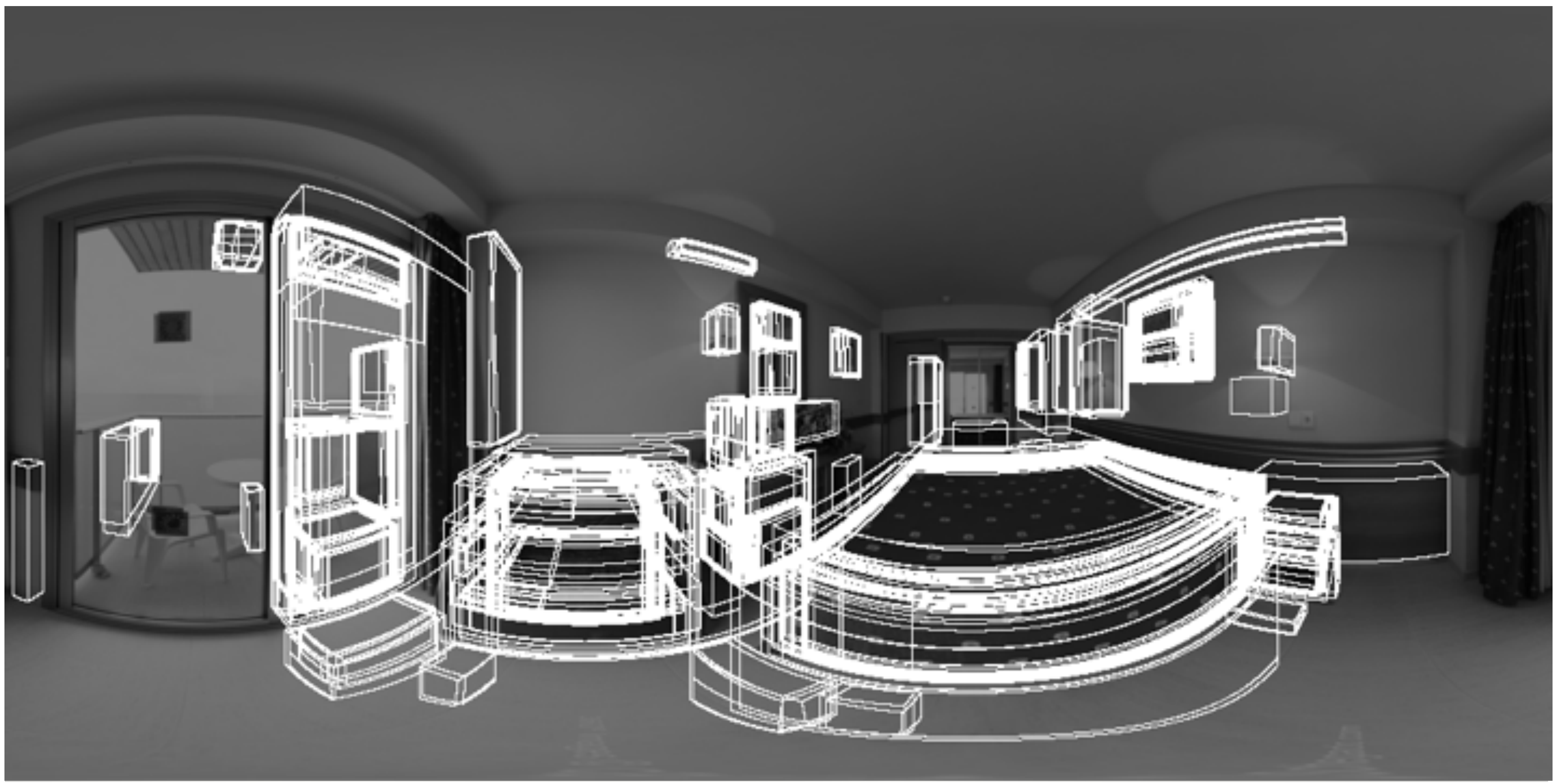
Algorithm

Step 1: Generate a pool of hypotheses

Step 2: Choose the best hypothesis



Cuboid detection

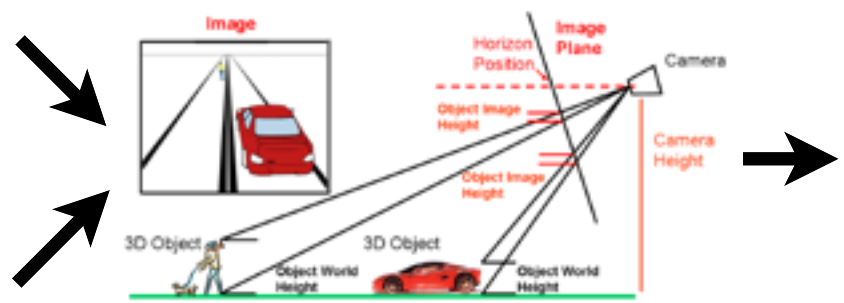


Fitted cuboids

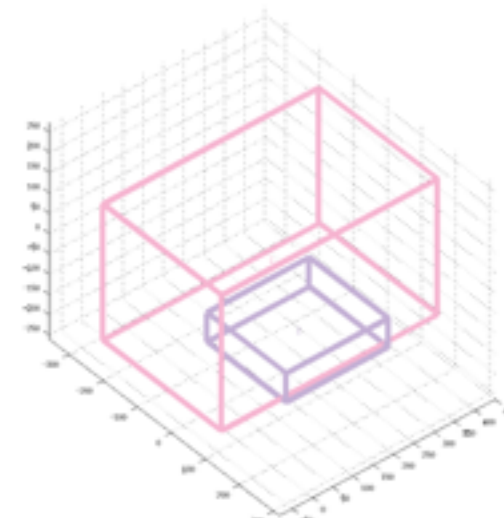
From 2D to 3D

Room

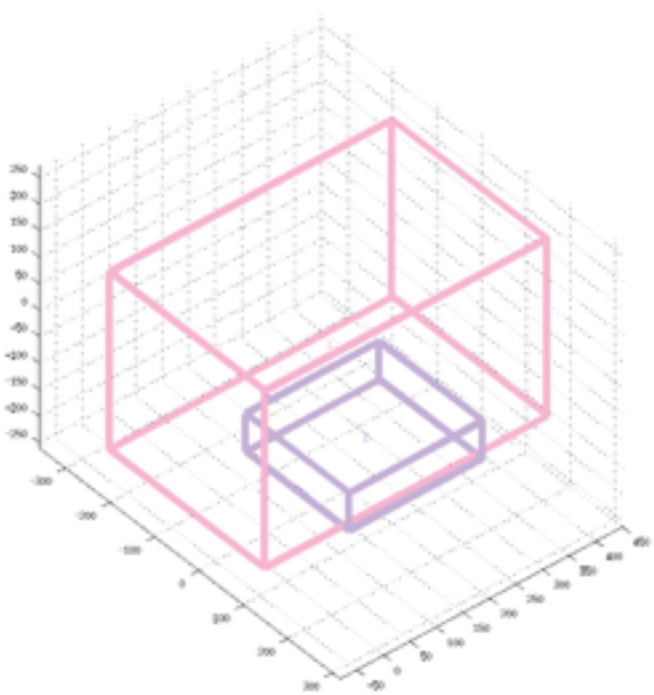
Object



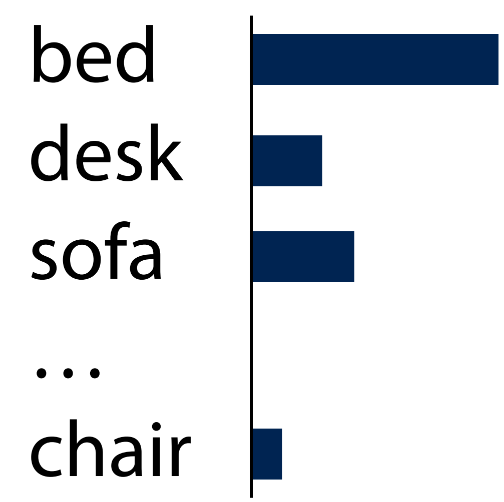
Single view geometry
[Hoiem et al. 2006]



Semantic classification

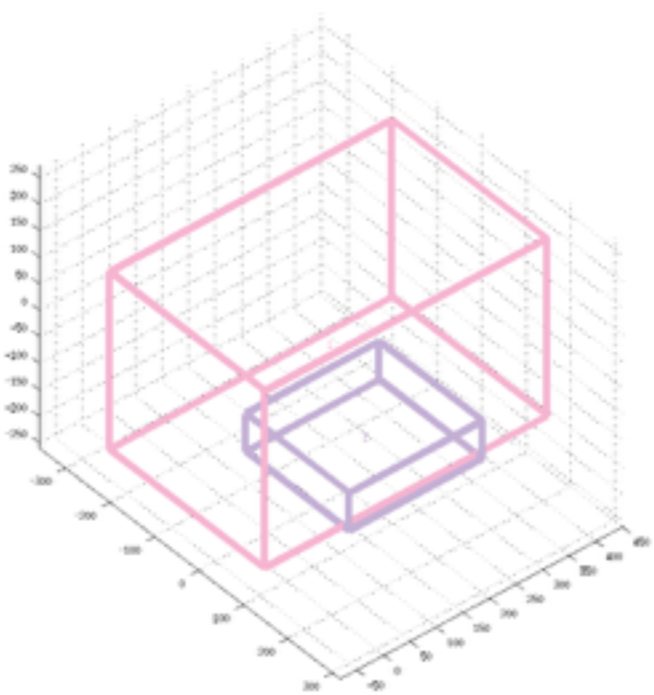


Object Category

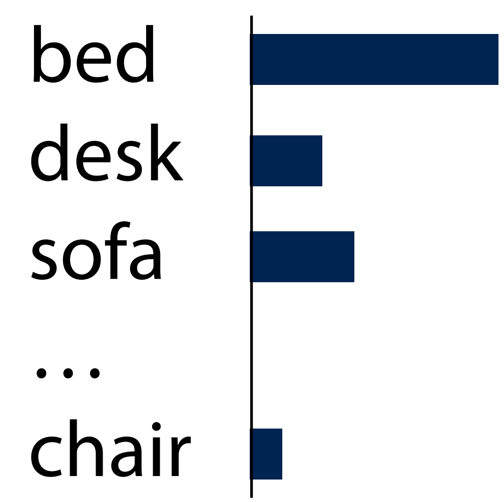


Use 3D position & size for semantic classification

Semantic classification

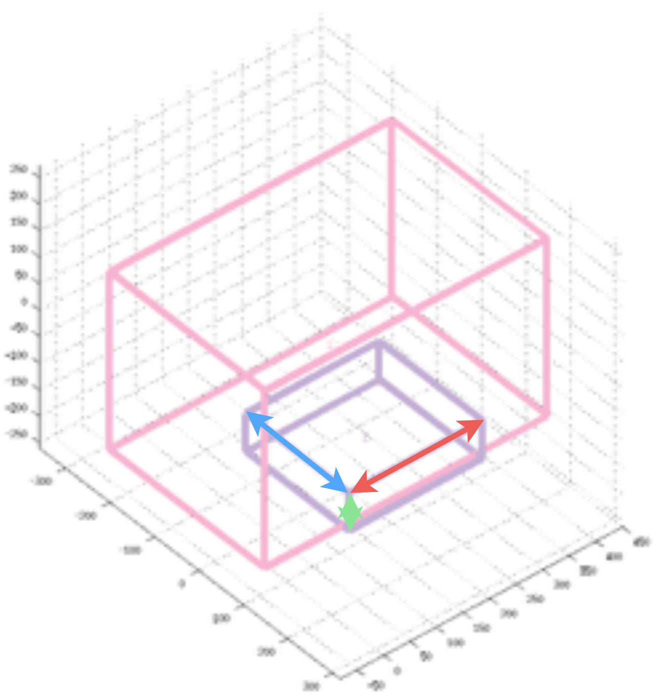


Feature **Classifier** **Object Category**



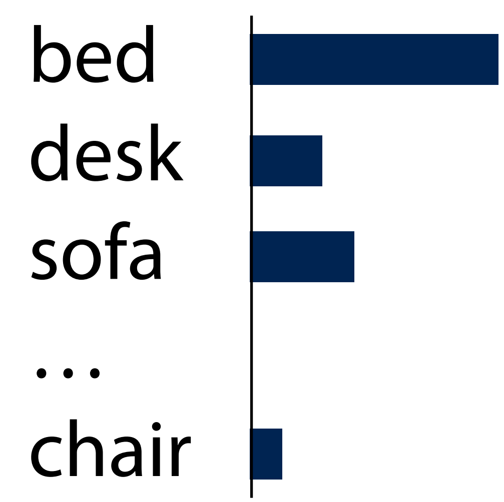
Use 3D position & size for semantic classification

Semantic classification



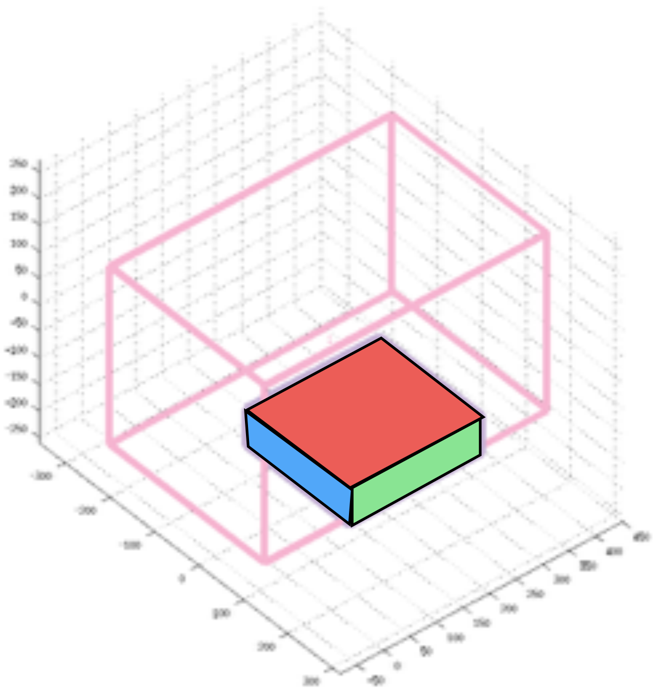
3D Cuboid Feature **Classifier** **Object Category**

- Size



Use 3D position & size for semantic classification

Semantic classification

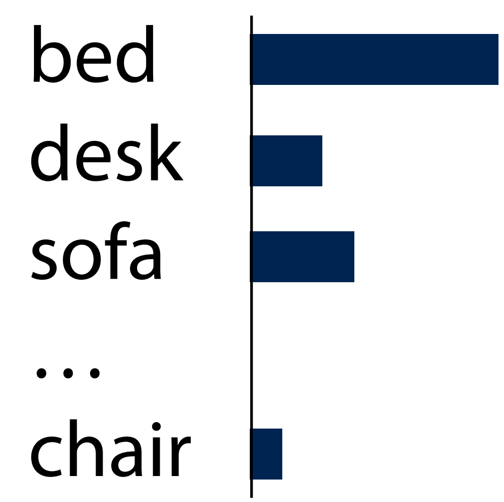


3D Cuboid Feature

- Size
- Aspect ratio & Area

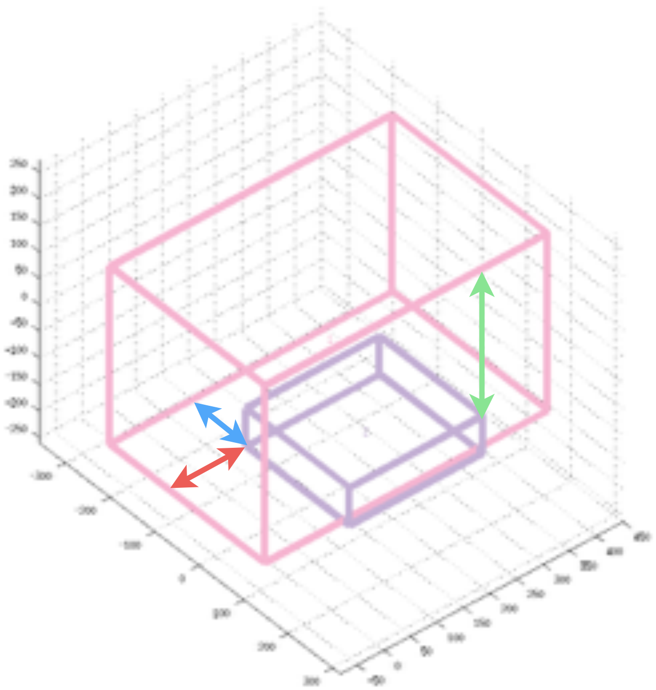
Classifier

Object Category



Use 3D position & size for semantic classification

Semantic classification



3D Cuboid Feature

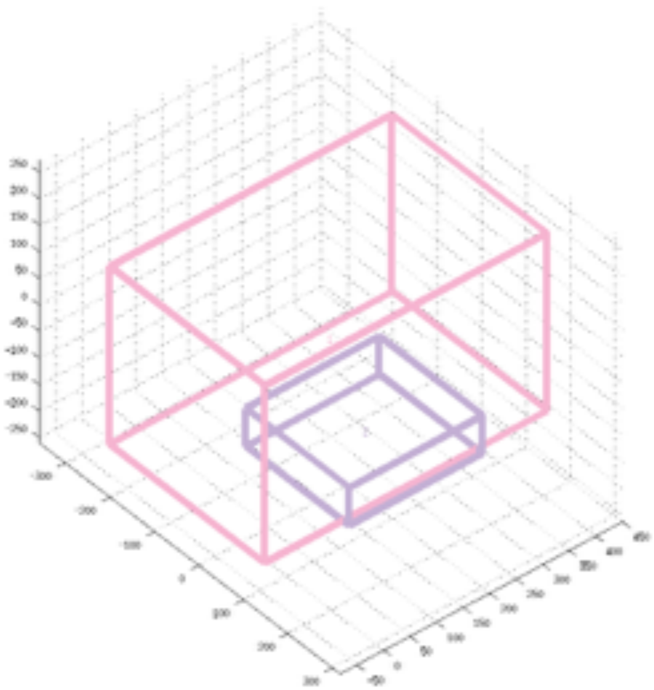
- Size
- Aspect ratio & Area
- Distance to walls

Classifier

Object Category

Use 3D position & size for semantic classification

Semantic classification



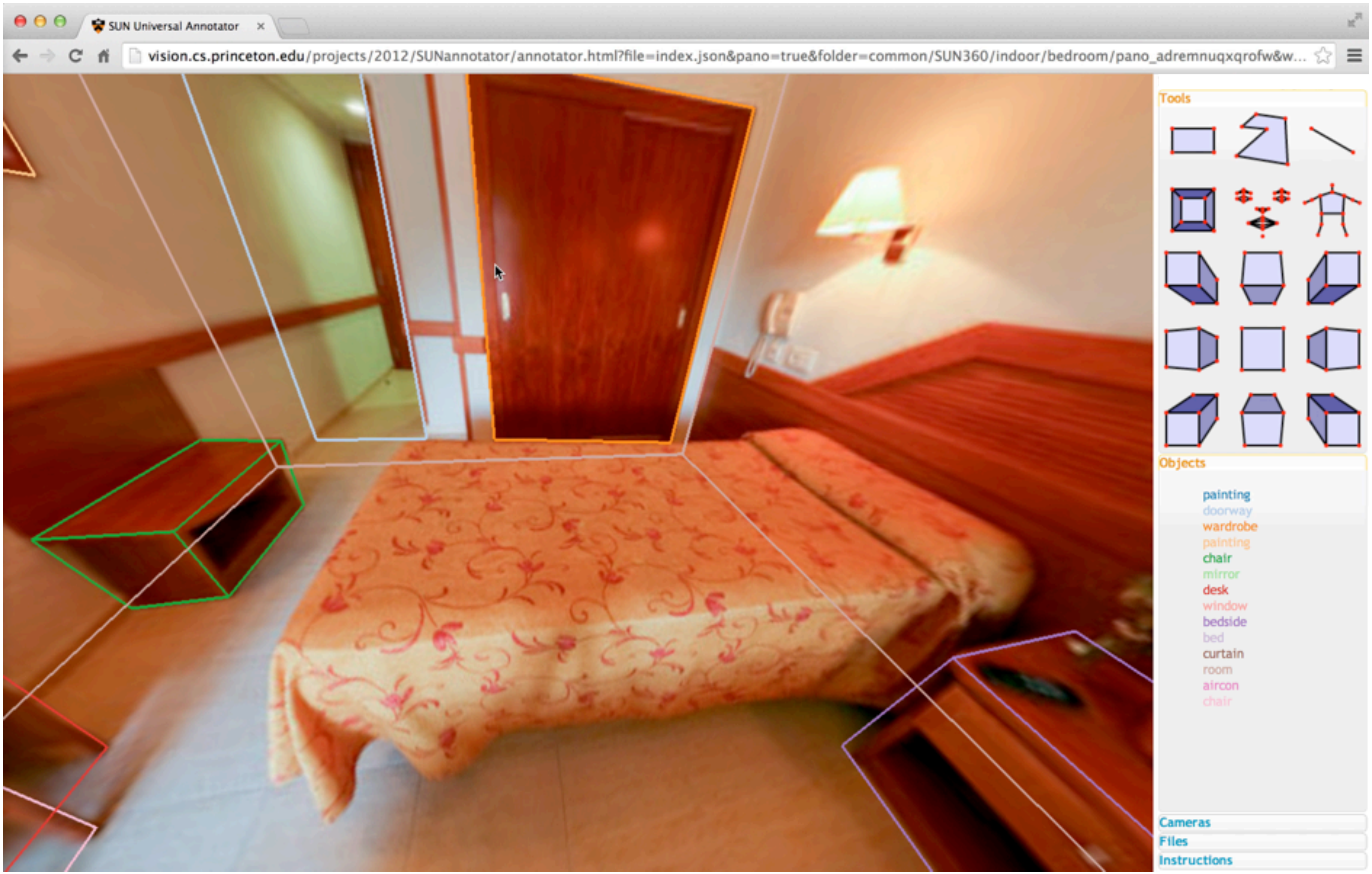
Data???

3D Cuboid Feature	Classifier	Object Category

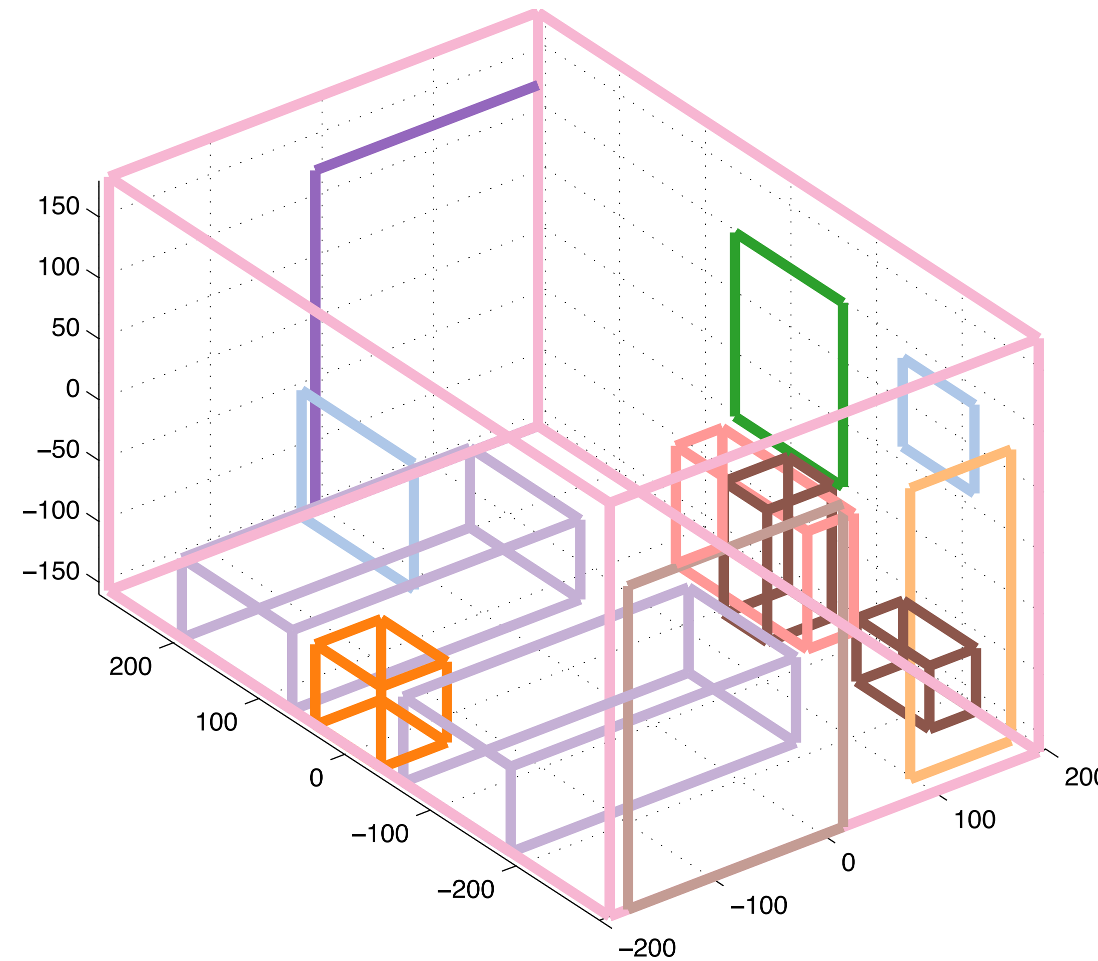
- Size
- Aspect ratio & Area
- Distance to walls

Use 3D position & size for semantic classification

2/3D annotation on panorama

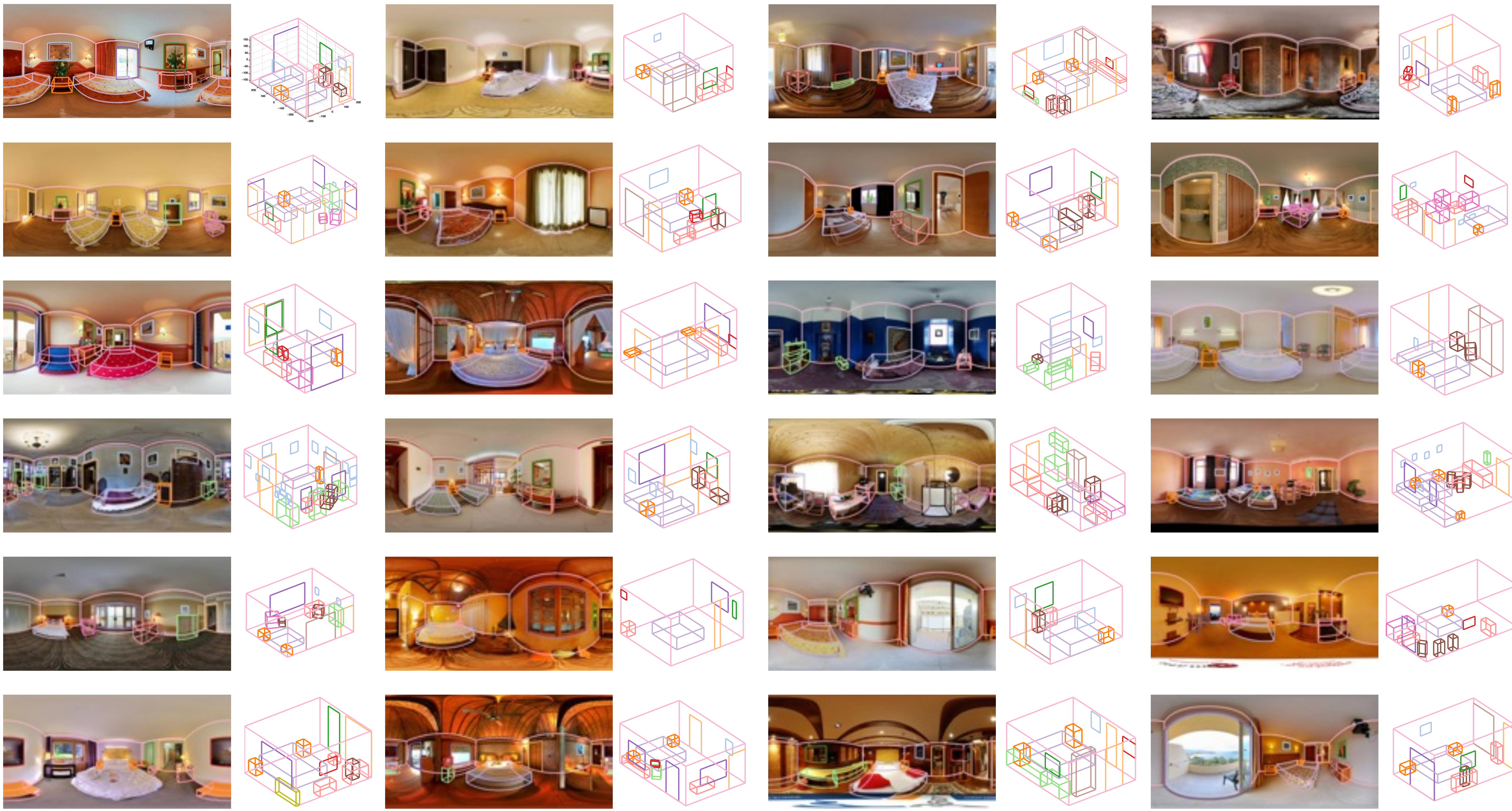


2/3D annotation on panorama



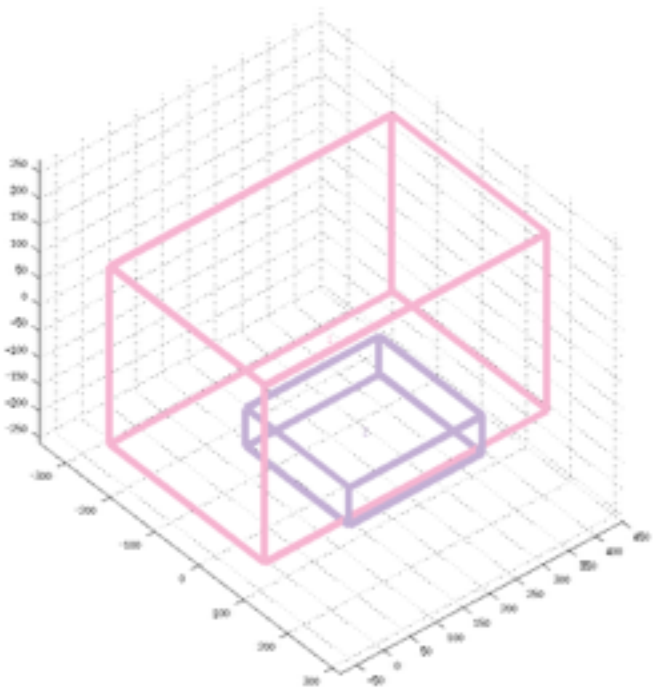
room bed window door nightstand desk sofa chair coffee table
painting mirror cabinet wardrobe dining table tv stand end table tv

Annotated panorama dataset



bedroom: 539   living room: 448   bathroom: 317   counter: 140

Semantic classification



2/3D annotated panorama

3D Cuboid Feature

- Size
- Aspect ratio & Area
- Distance to walls

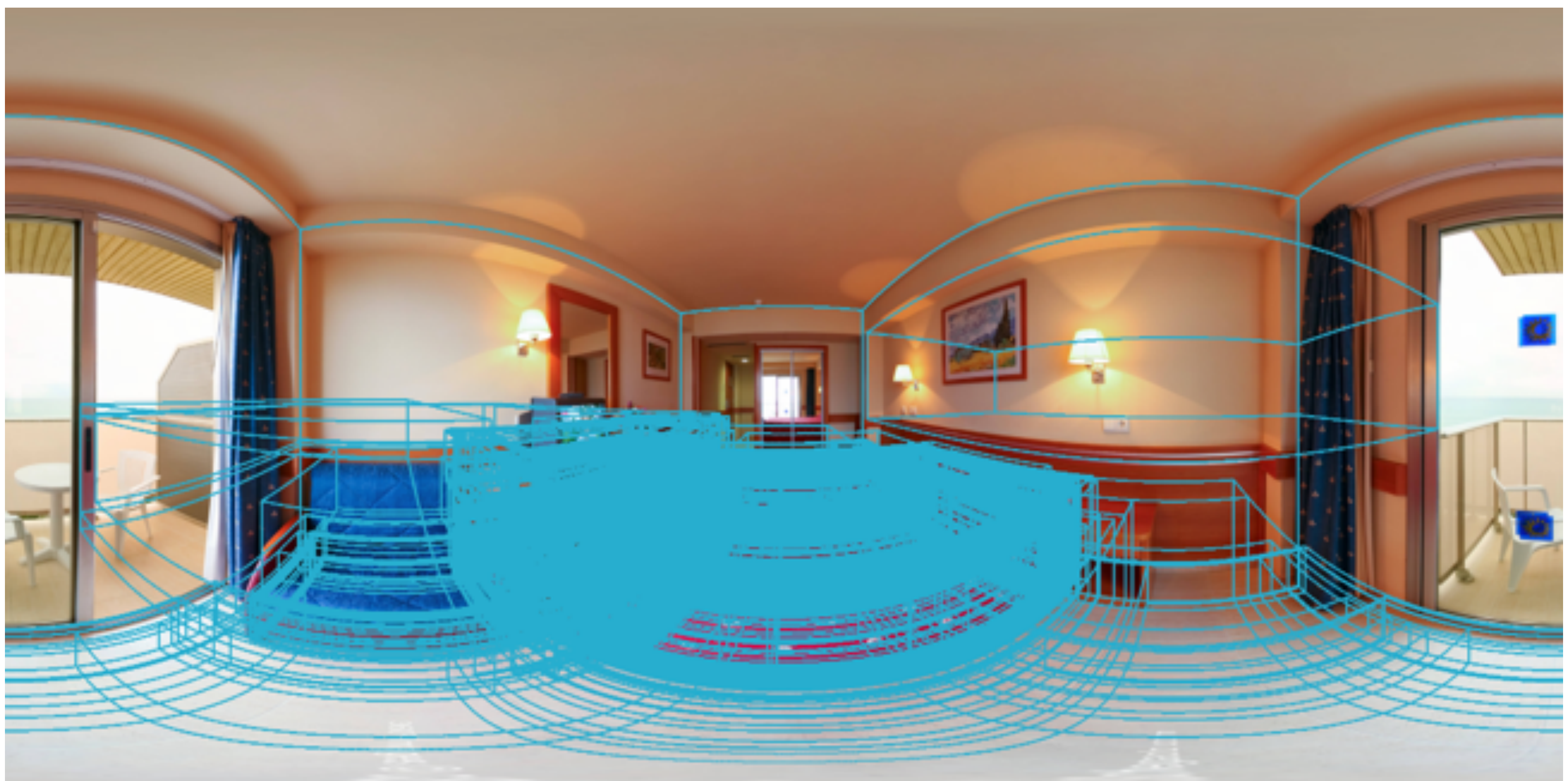
Classifier

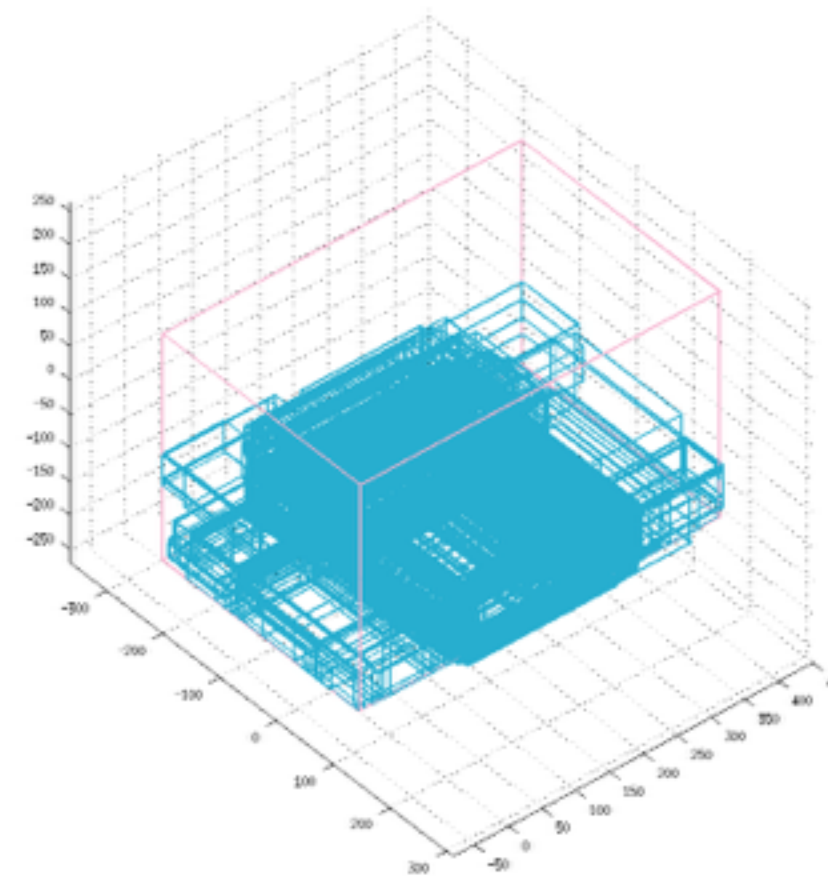
Object Category

Use 3D position & size for semantic classification

Hypotheses for some categories

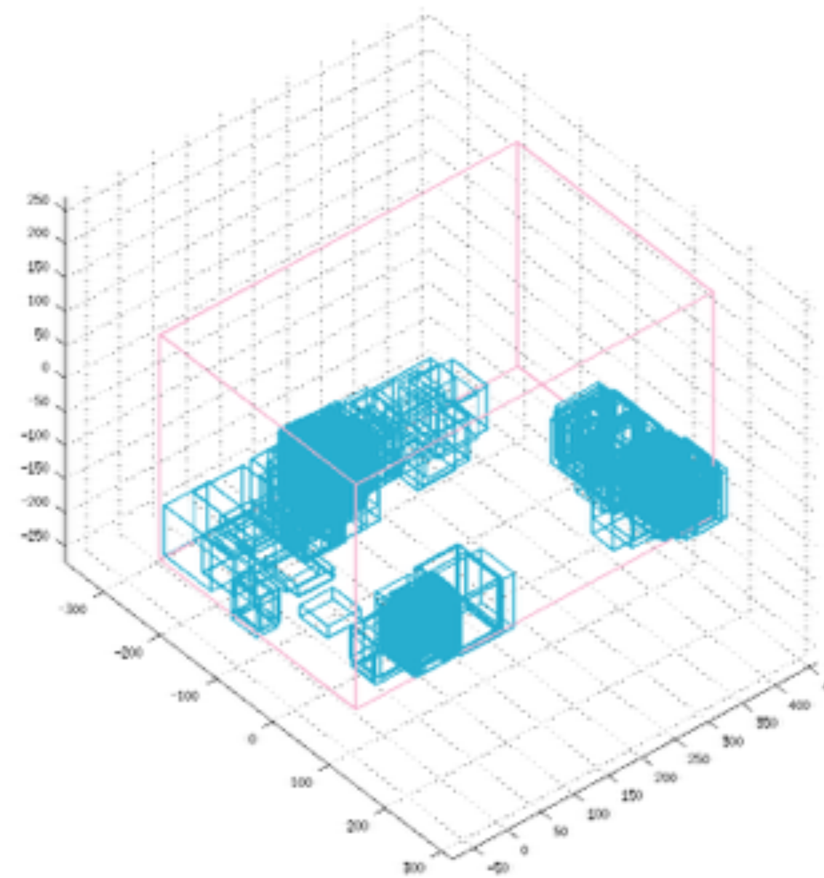
bed





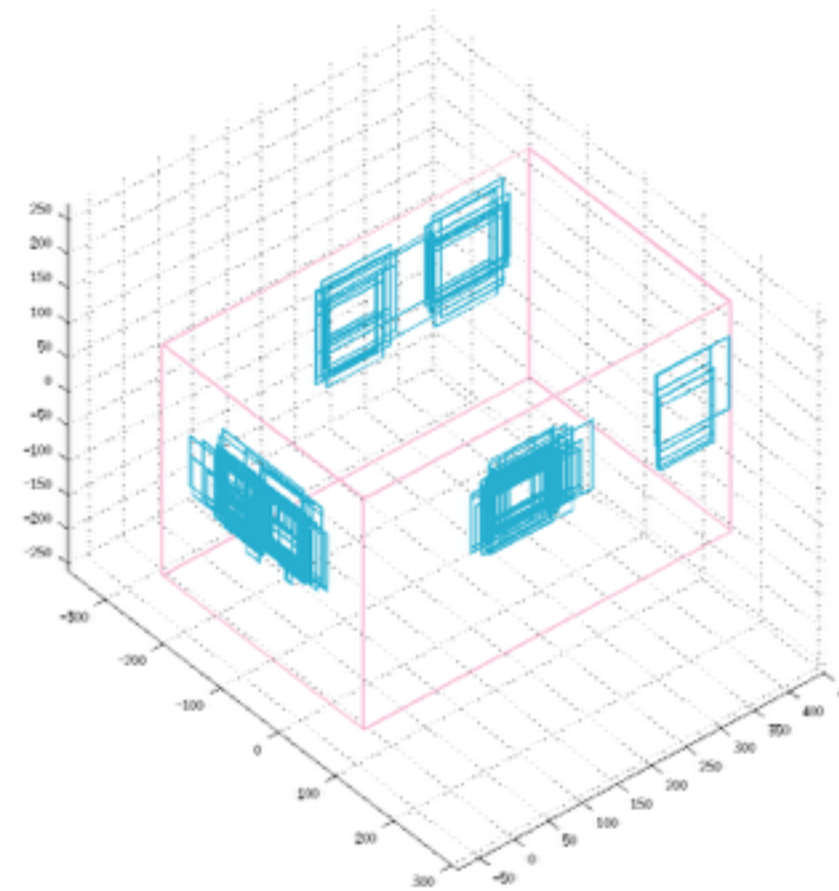
Hypotheses for some categories

nightstand



Hypotheses for some categories

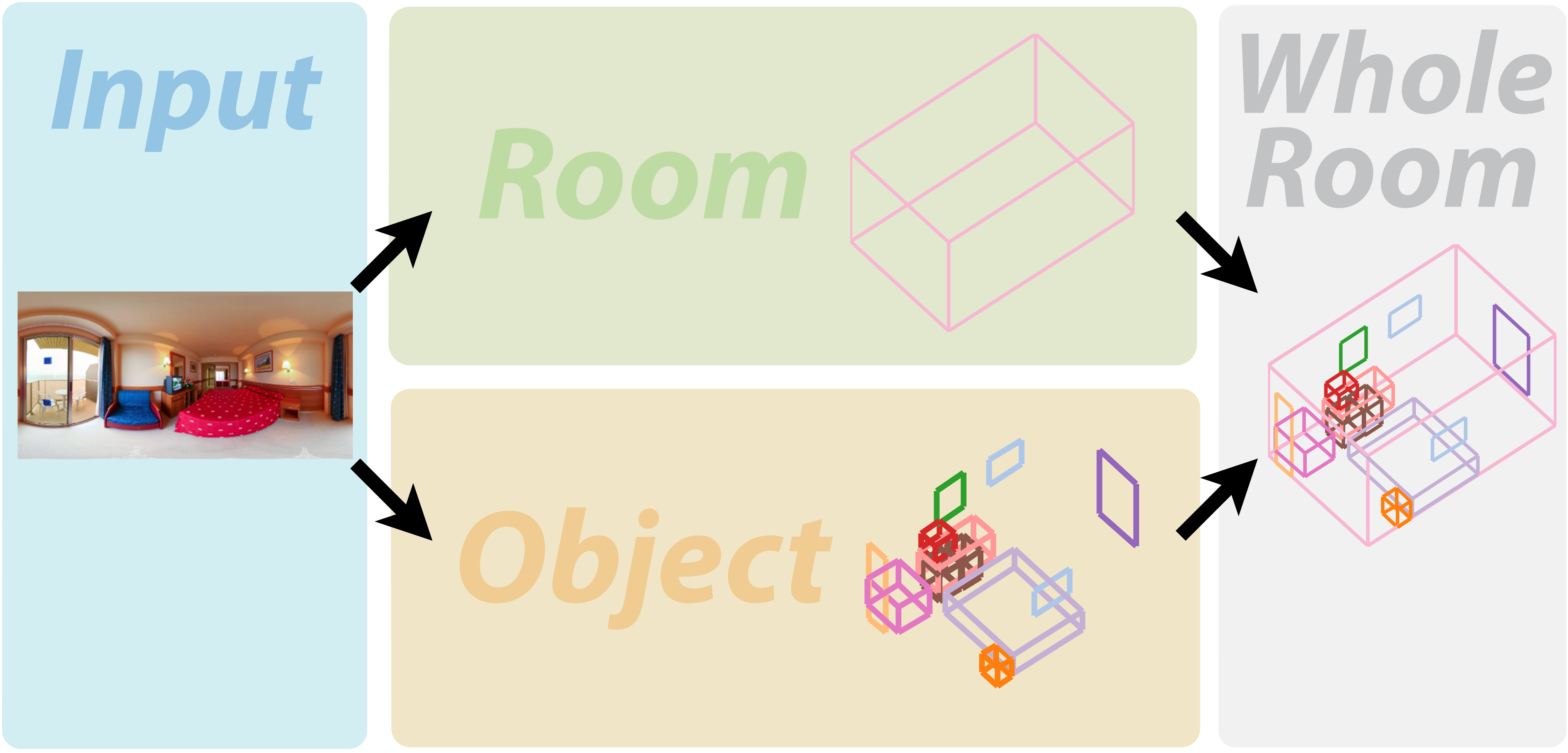
painting



Algorithm

Step 1: Generate a pool of hypotheses

Step 2: Choose the best hypothesis



Data-driven sampling

Data-driven sampling

Decide number of object based on prior distribution:

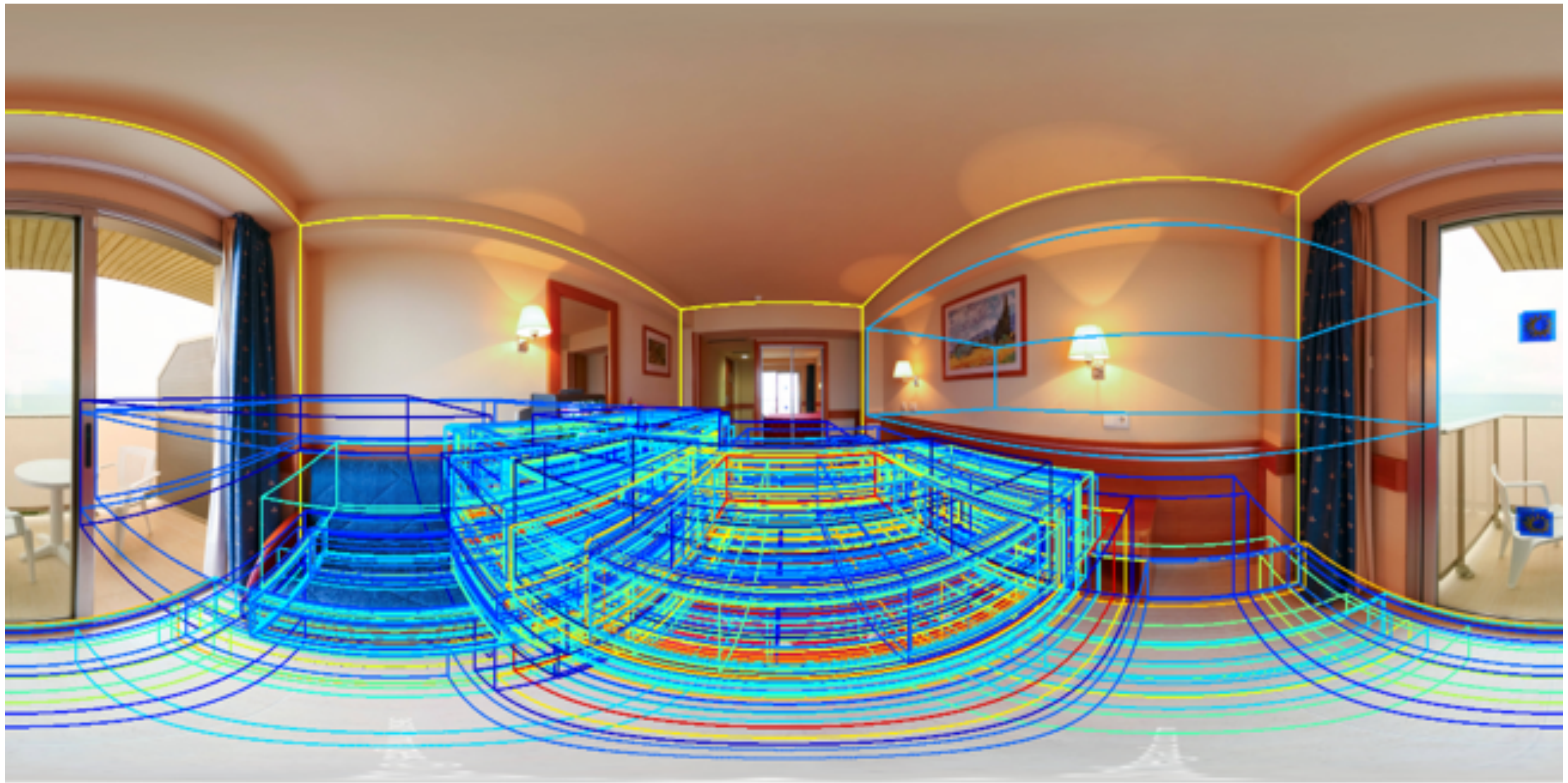
painting	2
bed	1
desk	1
nightsta	1
mirror	1
sofa	1
tv	1
window	1

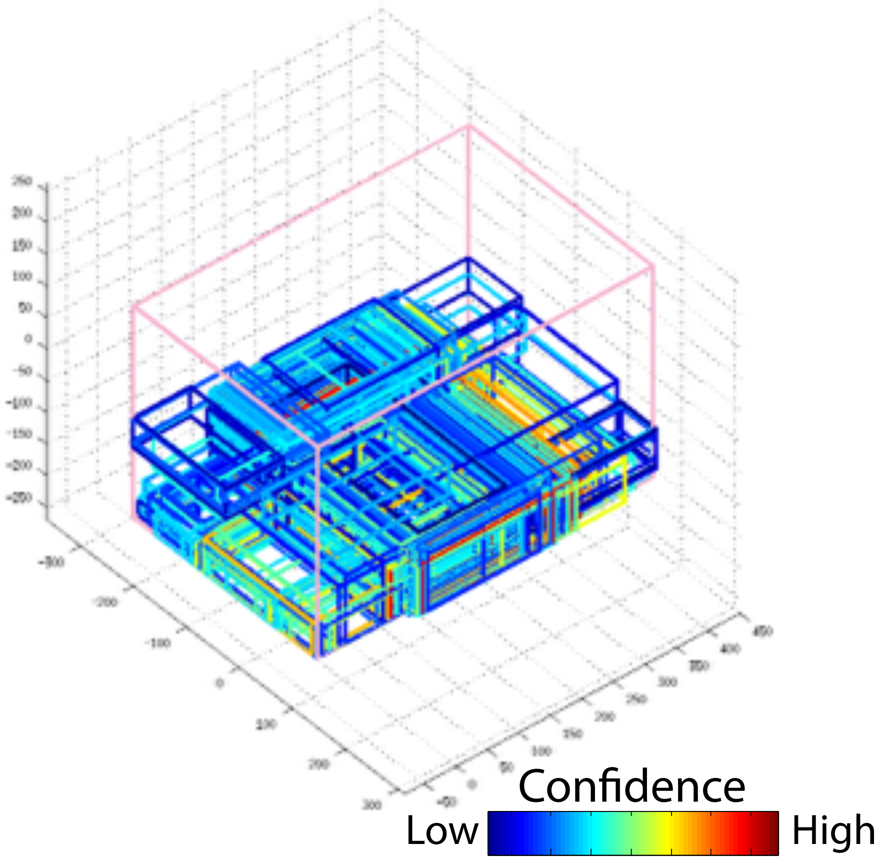
Decide object sampling sequence based on bottom up scores:

bed
nightstand
painting
desk
window
painting
tv
sofa
mirror

Data-driven sampling

Sample a bed in empty room first…

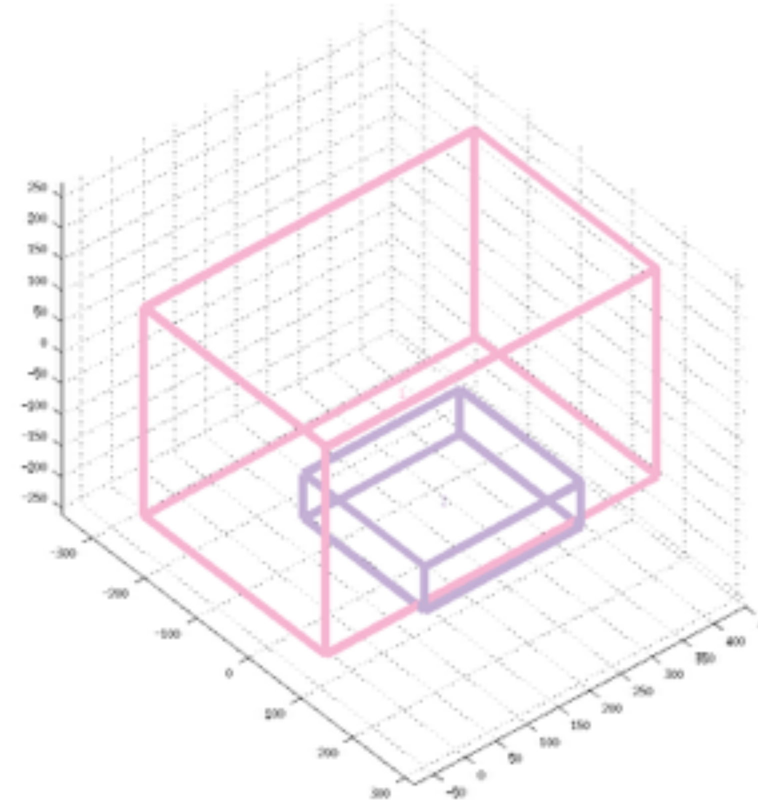




Bottom-up score as bed

Data-driven sampling

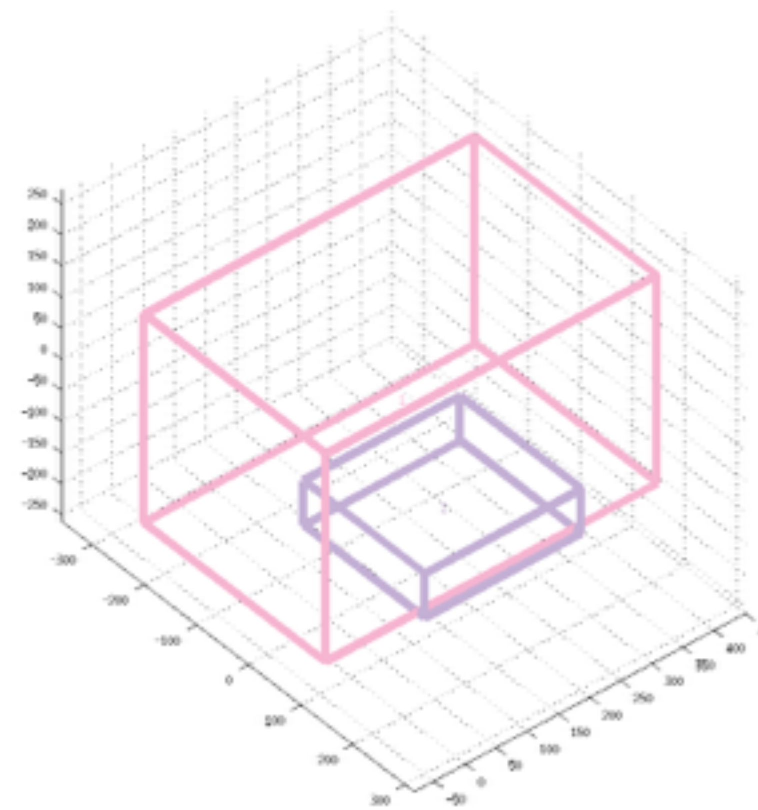
Sample a bed in empty room first…



Randomly select one according to priority

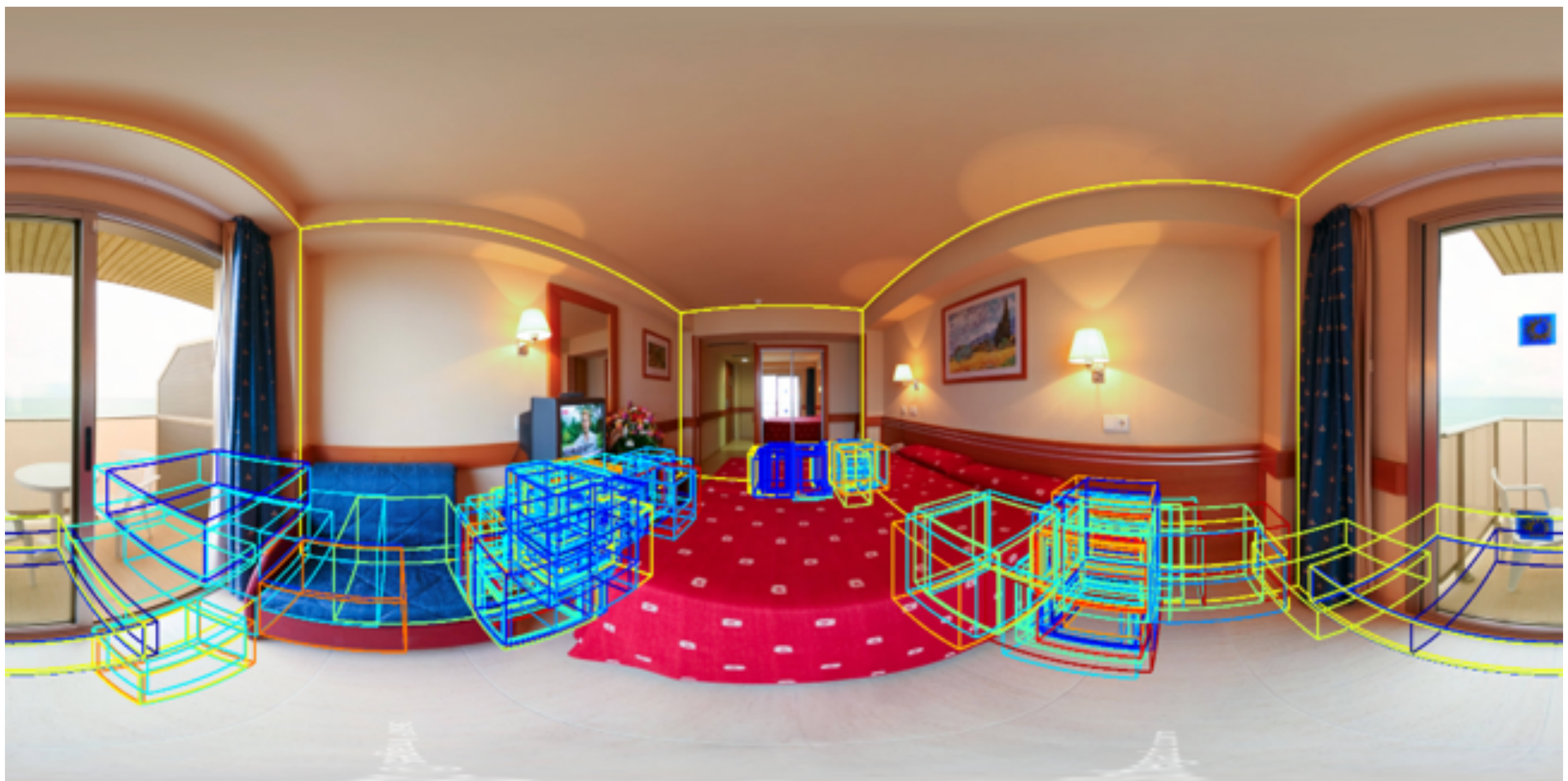
Data-driven sampling

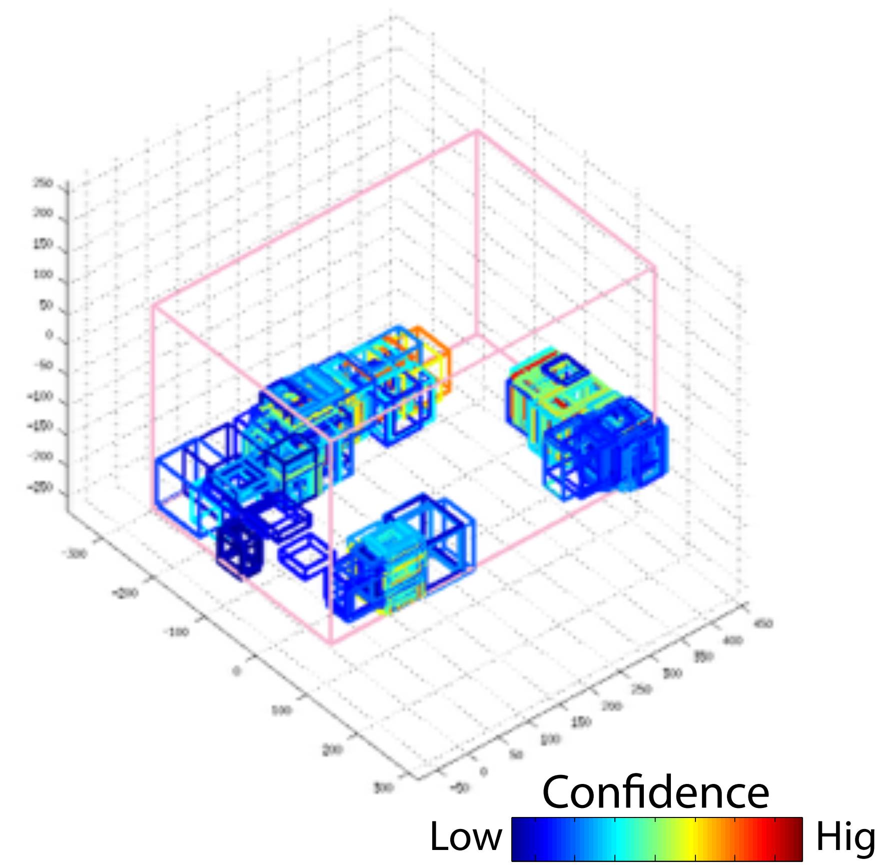
Then, sample a nightstand given a bed



Data-driven sampling

Then, sample a nightstand given a bed

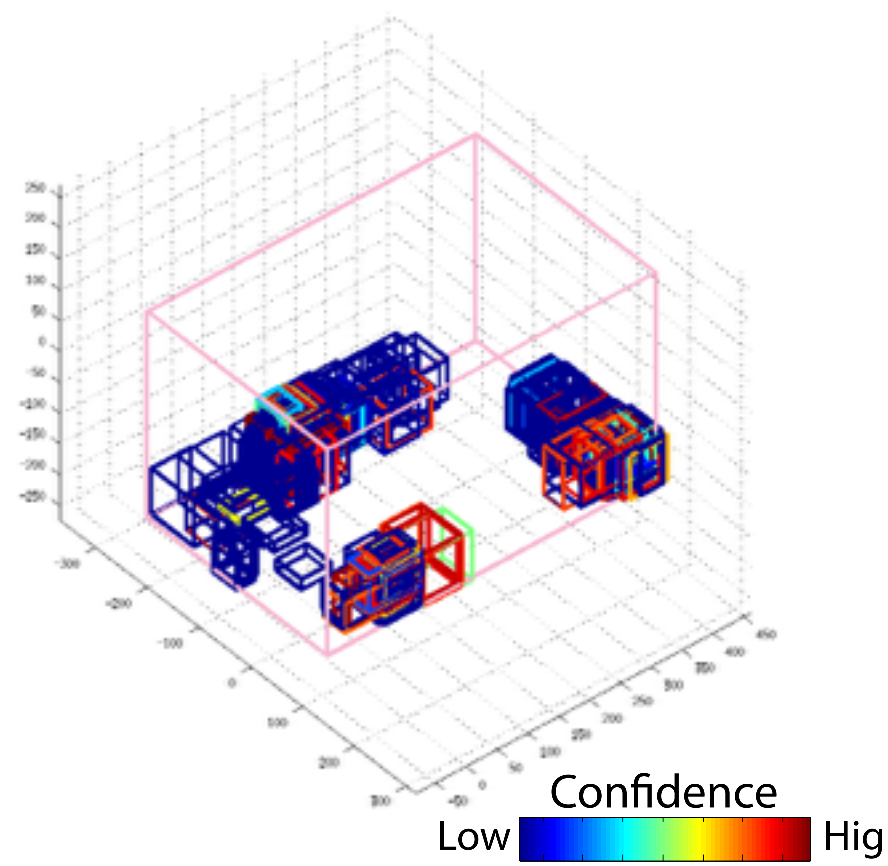




Bottom-up score

Data-driven sampling

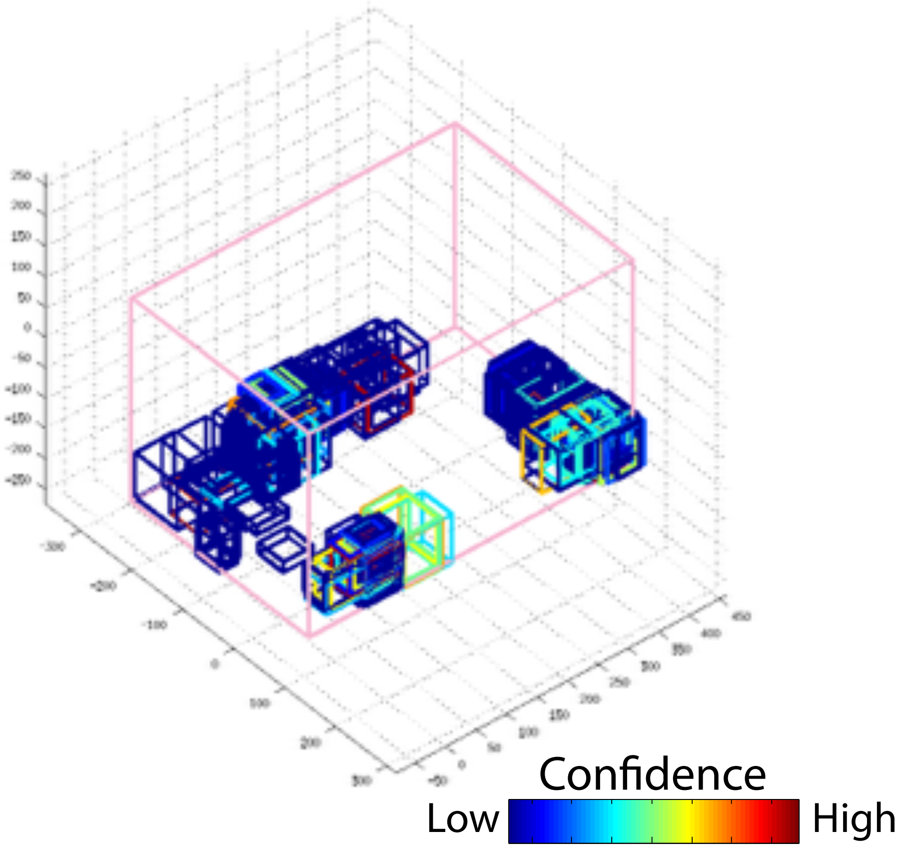
Then, sample a nightstand given a bed



Bottom-up score+Pairwise context

Data-driven sampling

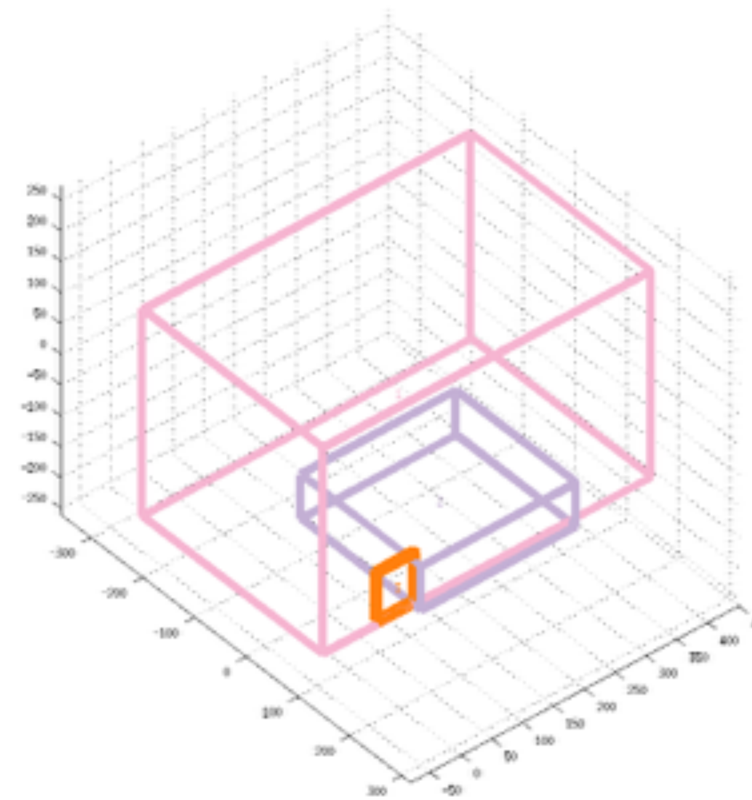
Then, sample a nightstand given a bed



Bottom-up score + Pairwise context → **Merged sampling prior**

Data-driven sampling

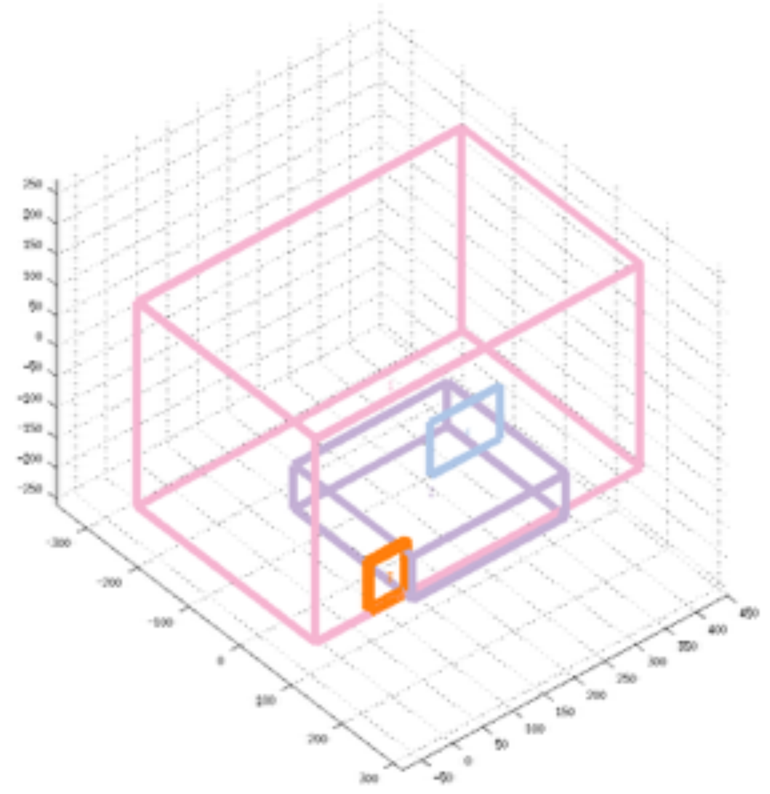
Then, sample a nightstand given a bed



Randomly select one according to the merged priority

Data-driven sampling

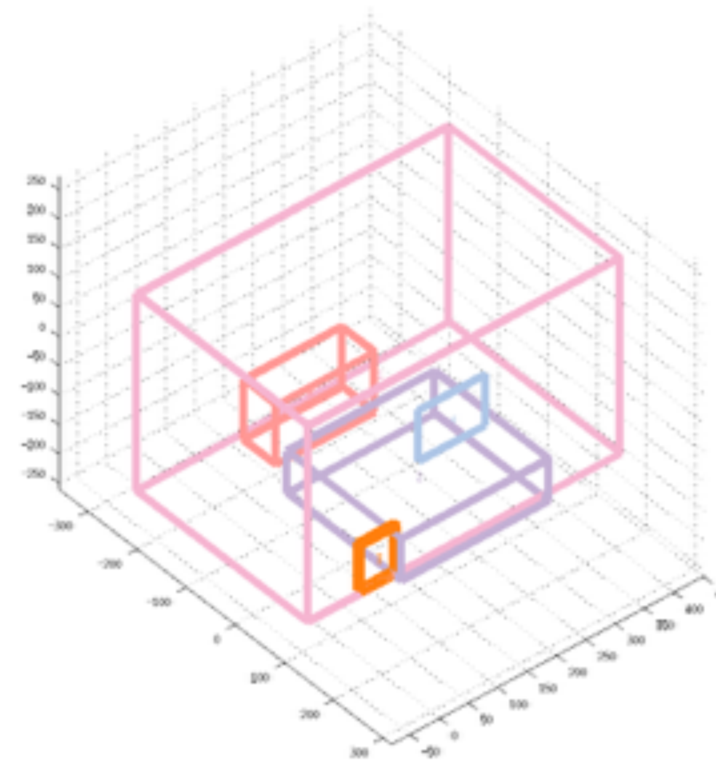
Keep on sampling until finishing the list…



List: bed, nightstand, painting, desk, window, painting, TV, sofa, mirror

Data-driven sampling

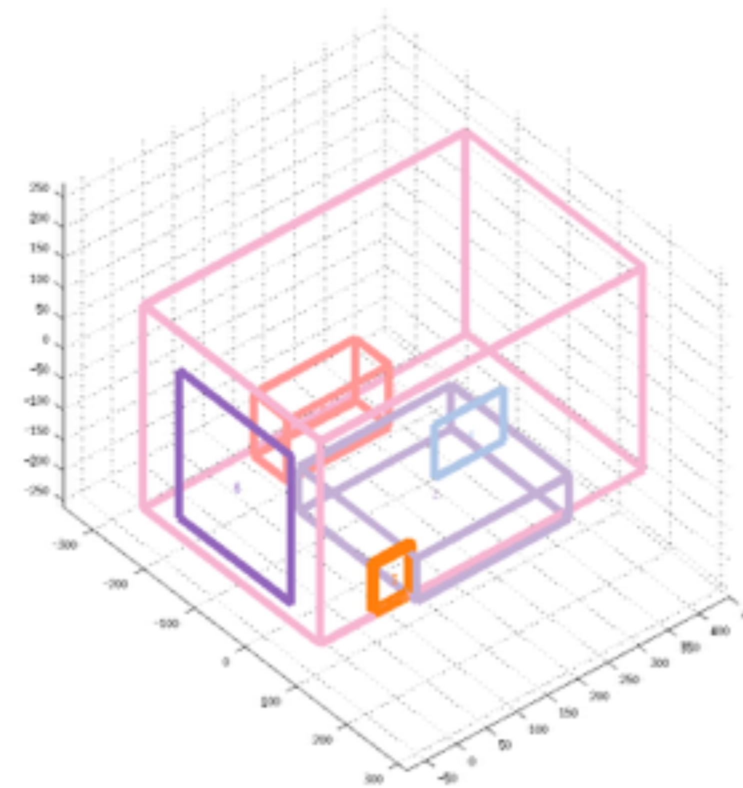
Keep on sampling until finishing the list…



List: bed, nightstand, painting, desk, window, painting, TV, sofa, mirror

Data-driven sampling

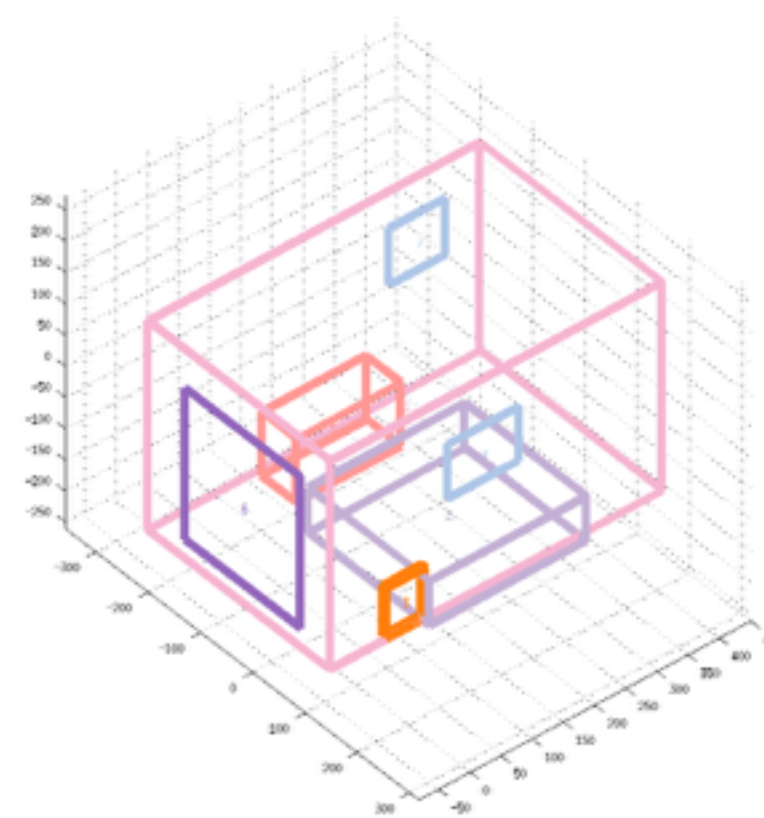
Keep on sampling until finishing the list…



List: bed, nightstand, painting, desk, window, painting, TV, sofa, mirror

Data-driven sampling

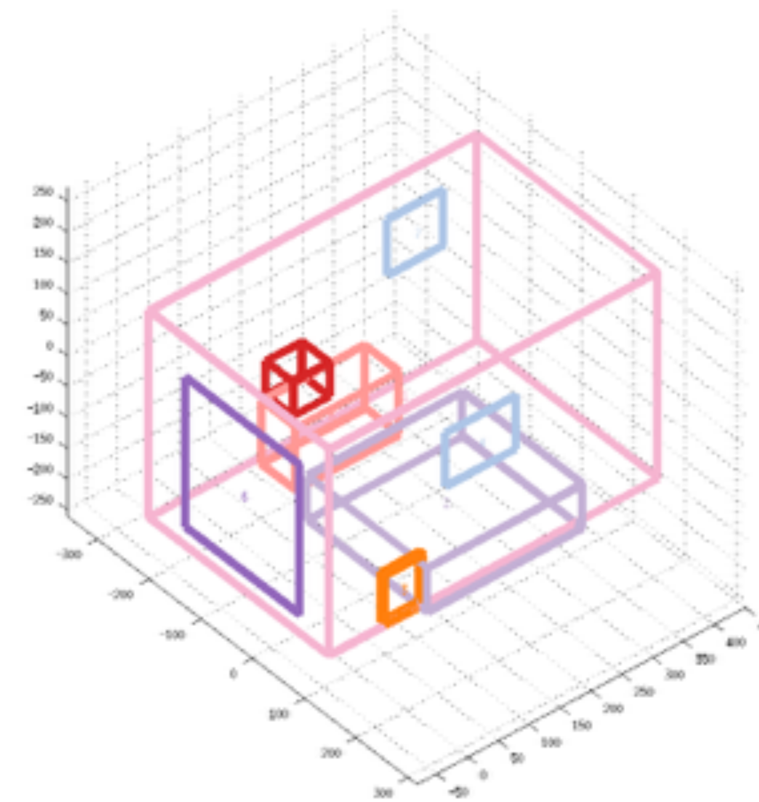
Keep on sampling until finishing the list…



List: bed, nightstand, painting, desk, window, painting, TV, sofa, mirror

Data-driven sampling

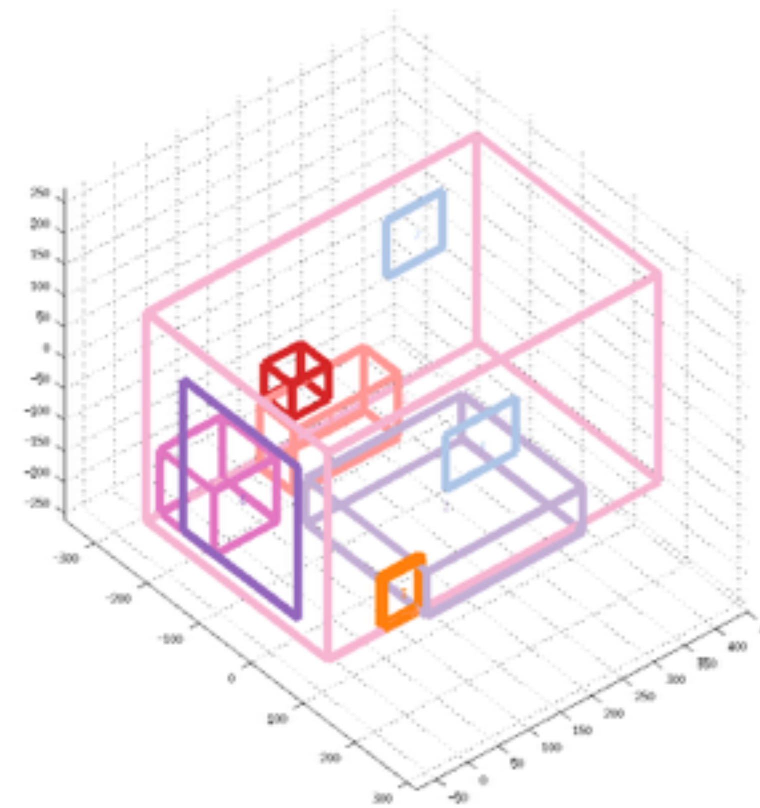
Keep on sampling until finishing the list…



List: bed, nightstand, painting, desk, window, painting, TV, sofa, mirror

Data-driven sampling

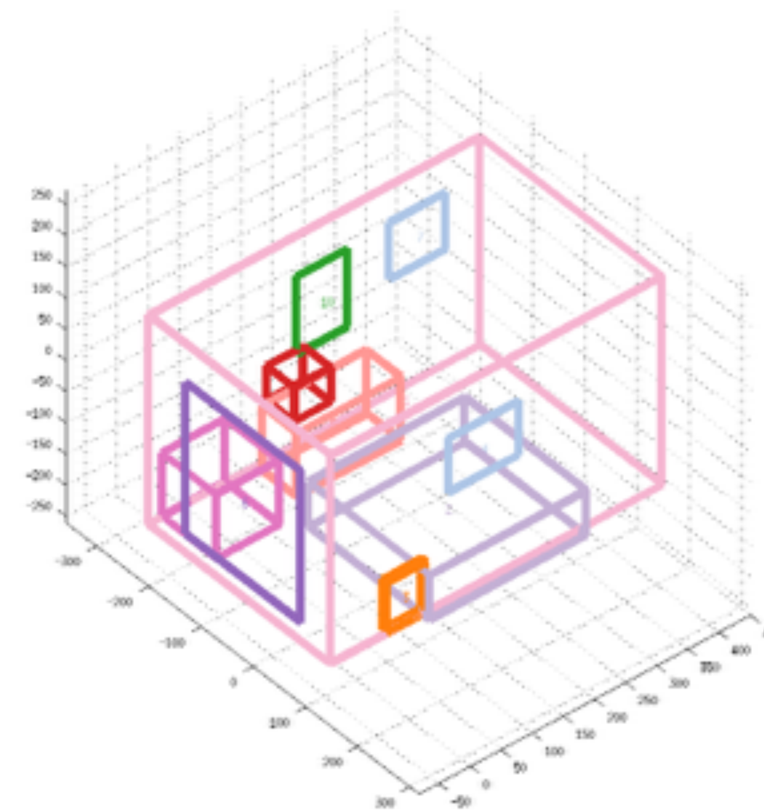
Keep on sampling until finishing the list…



List: bed, nightstand, painting, desk, window, painting, TV, sofa, mirror

Data-driven sampling

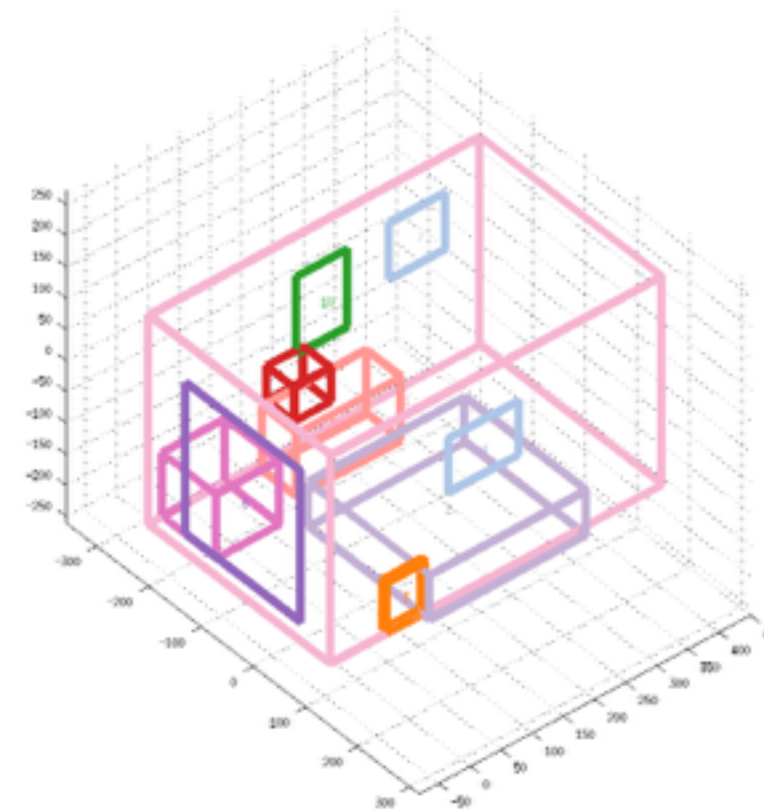
Keep on sampling until finishing the list…



List: bed, nightstand, painting, desk, window, painting, TV, sofa, mirror

Data-driven sampling

Keep on sampling until finishing the list…

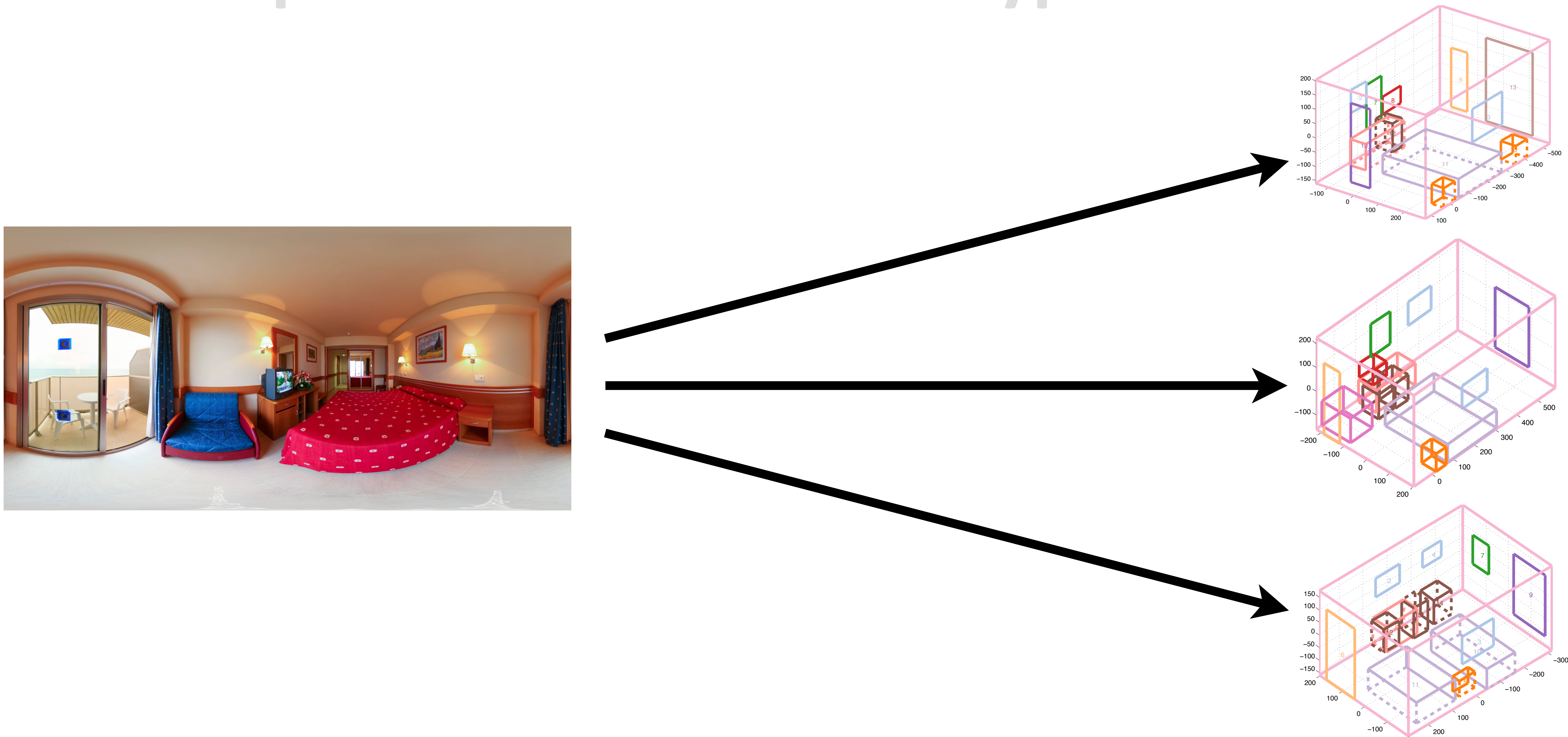


Whole-room sampling is finished.

Algorithm

Step 1: Generate a pool of hypotheses

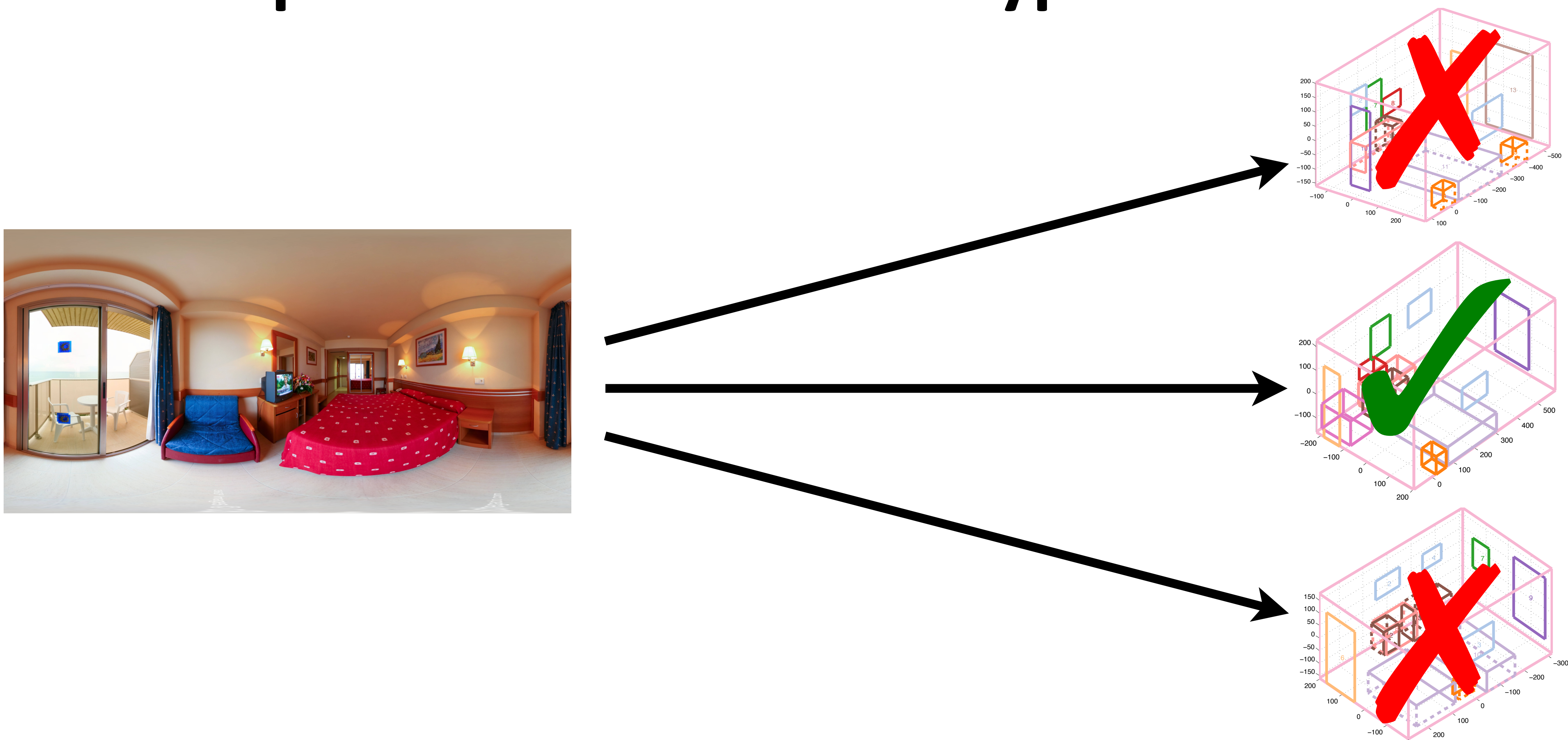
Step 2: Choose the best hypothesis



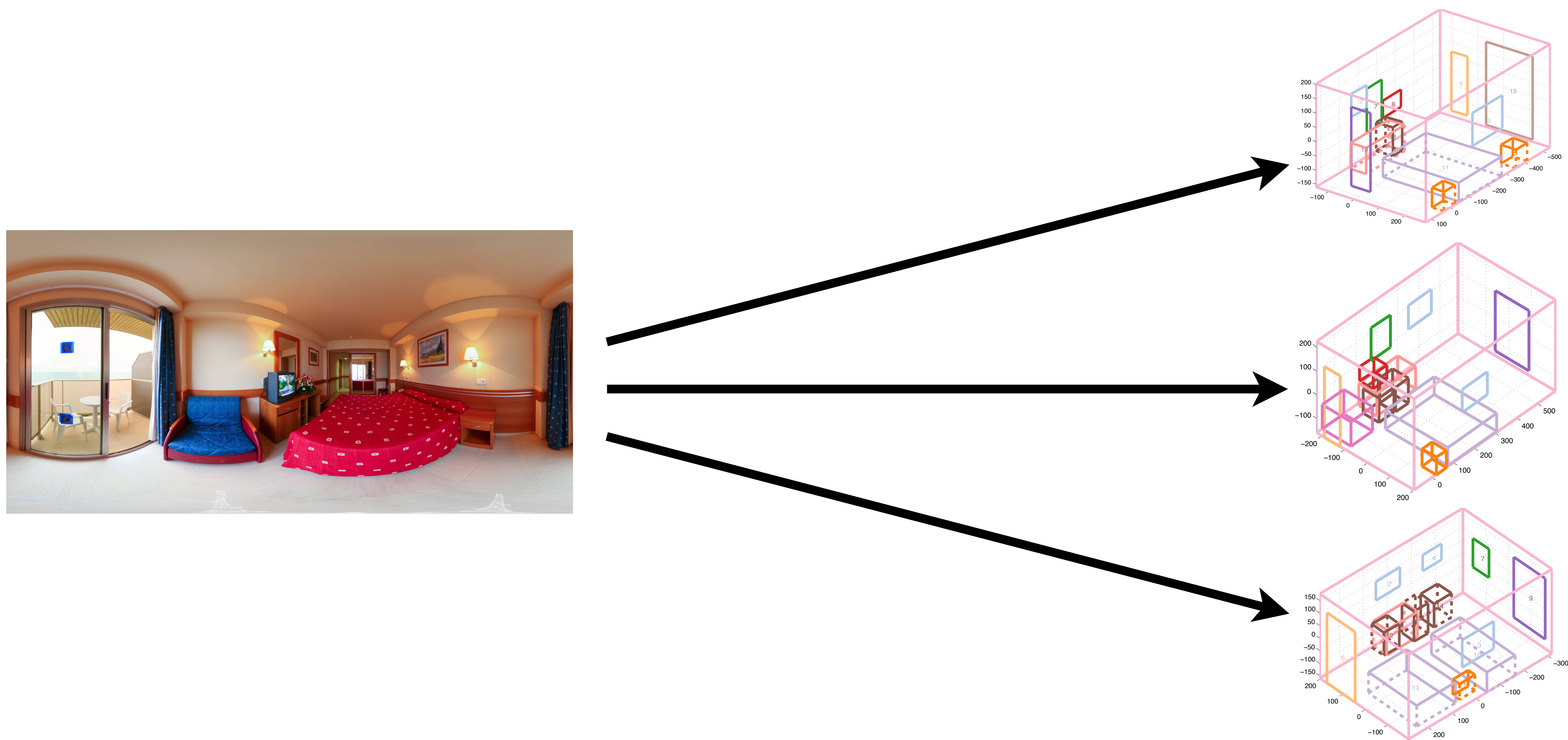
Algorithm

Step 1: Generate a pool of hypotheses

Step 2: Choose the best hypothesis

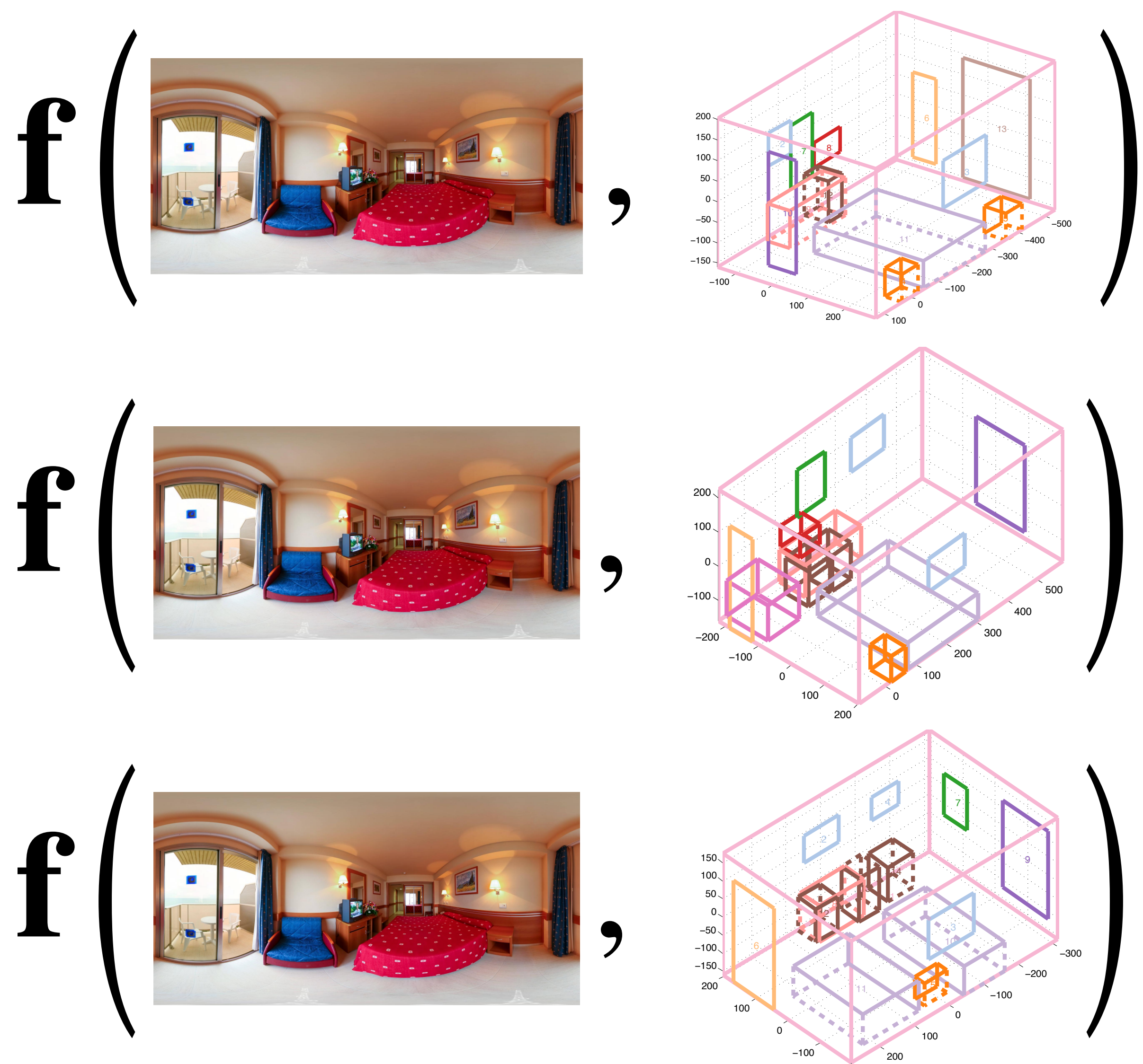


Holistic ranking



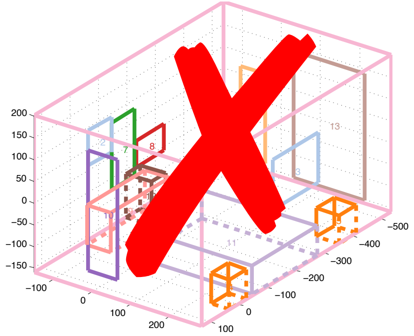
Holistic ranking

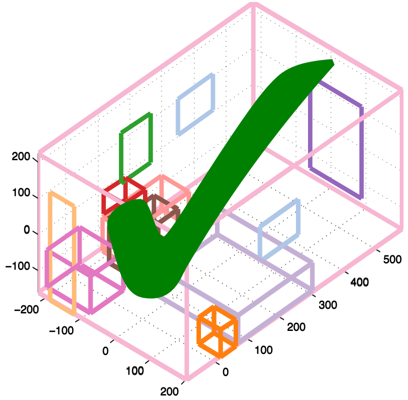
Compute holistic feature for whole-room hypotheses

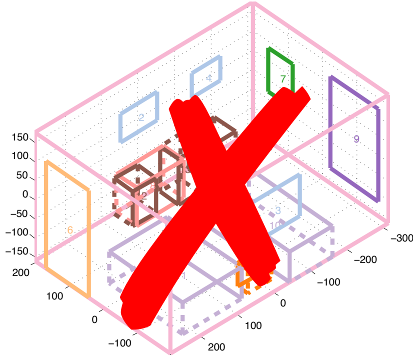


Holistic ranking

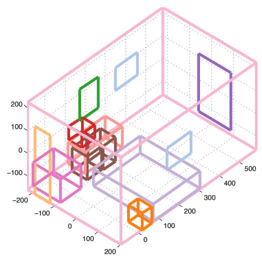
Learn a linear SVM for scoring and take the best

$\mathbf{w}^T \mathbf{f}\Big($  ,  $\Big)$

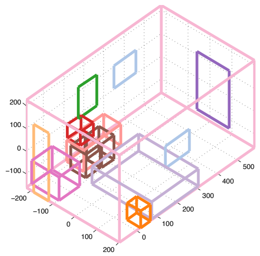
$\mathbf{w}^T \mathbf{f}\Big($  ,  $\Big)$

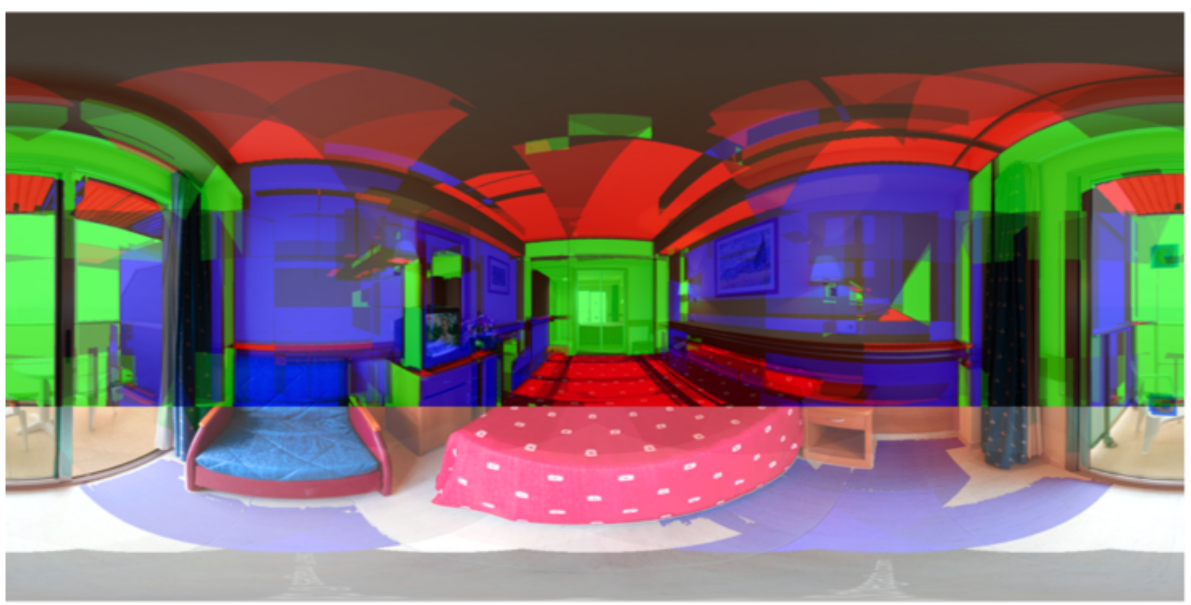
$\mathbf{w}^T \mathbf{f}\Big($  ,  $\Big)$

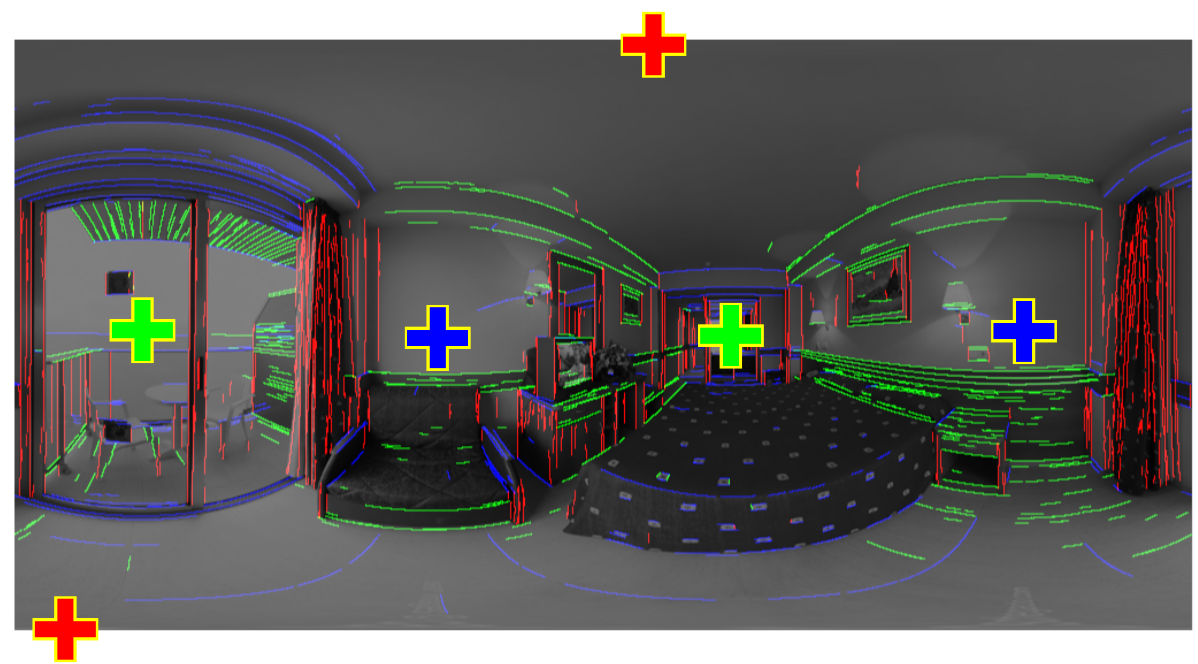
Holistic feature

$\mathbf{f}\Big($  ,  $\Big)$ = bottom-up feature + top-down feature

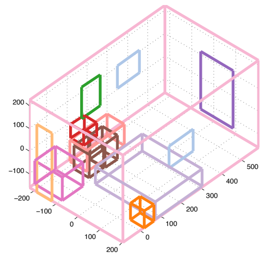
Holistic feature

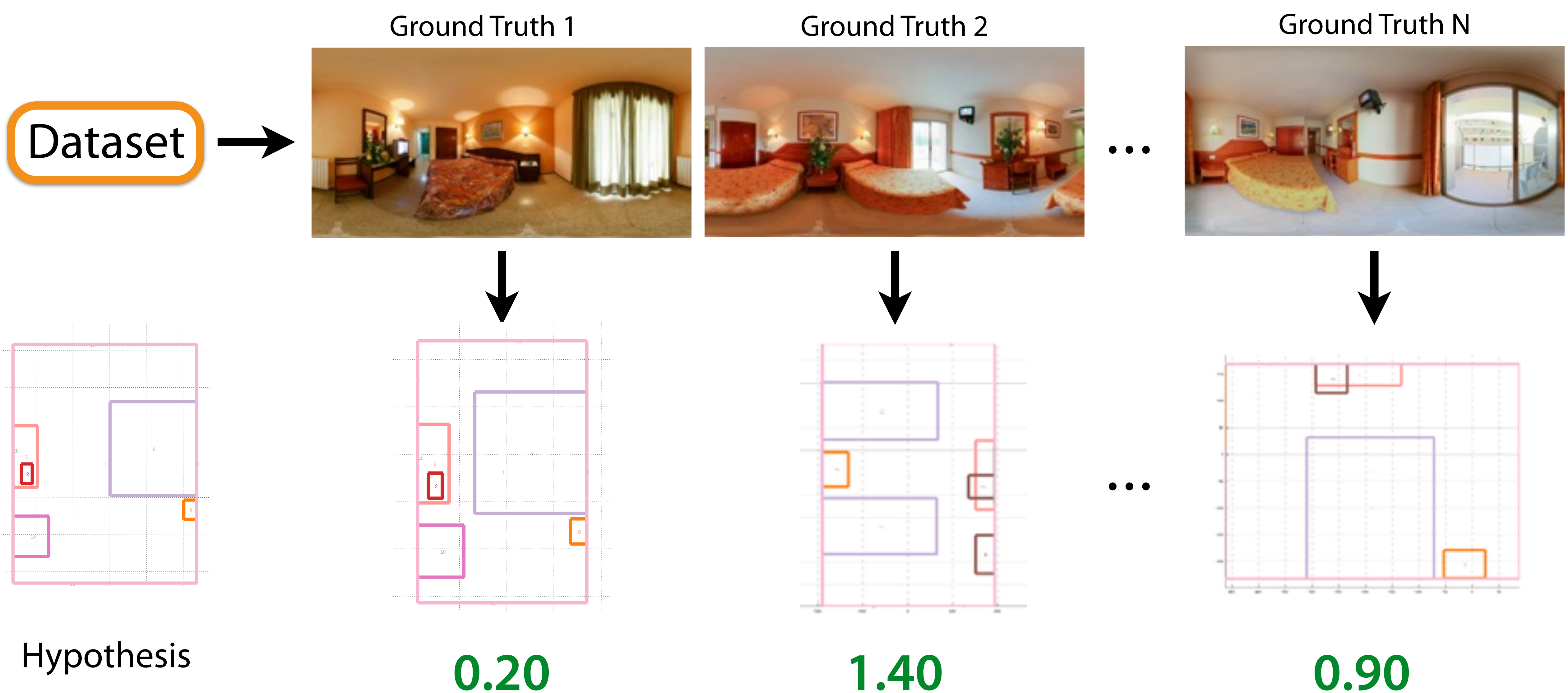
f ( , ) = **bottom-up feature** + top-down feature



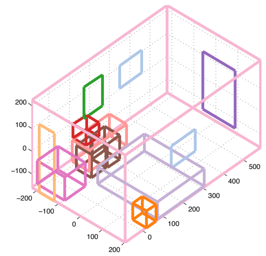


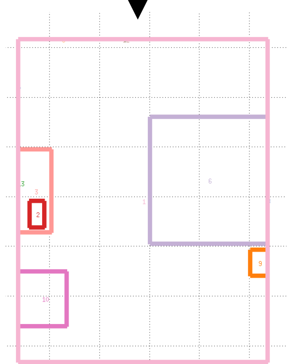
Holistic feature

f( , ) = bottom-up feature + **top-down feature**



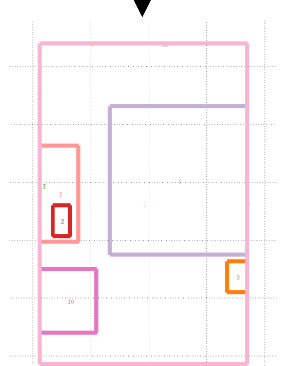
Holistic feature

f ( , ) = bottom-up feature +**top-down feature**



Hypothesis

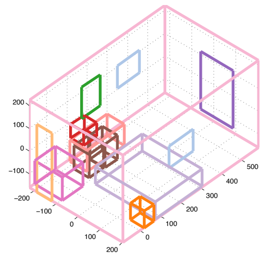
Dataset

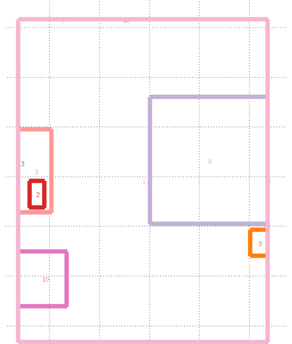


A ground truth room

Transform ground truth to get more good rooms

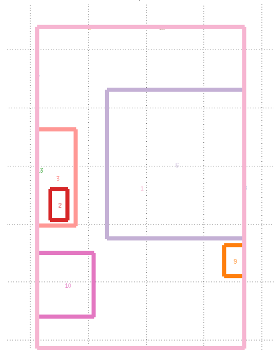
Holistic feature

f ( , ) = bottom-up feature + **top-down feature**



Hypothesis

Dataset



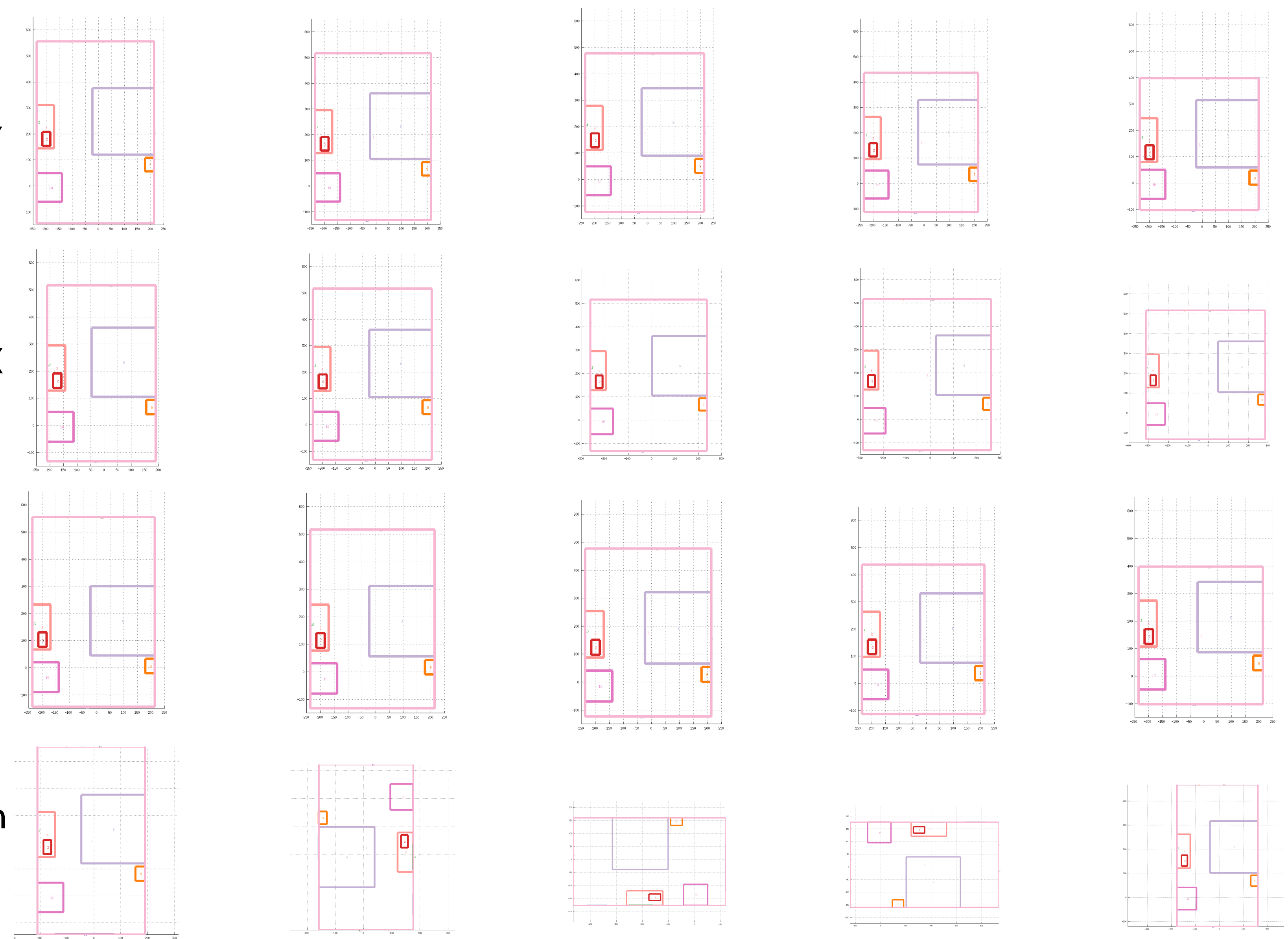
A ground truth room

Resize Y

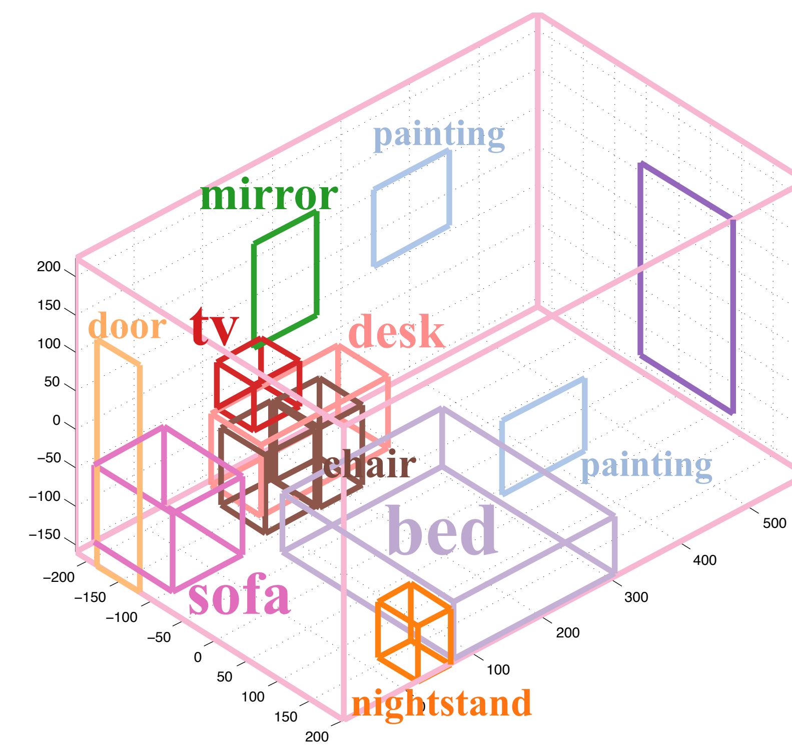
Resize X

Fix dist. to wall

Rotation & Scale

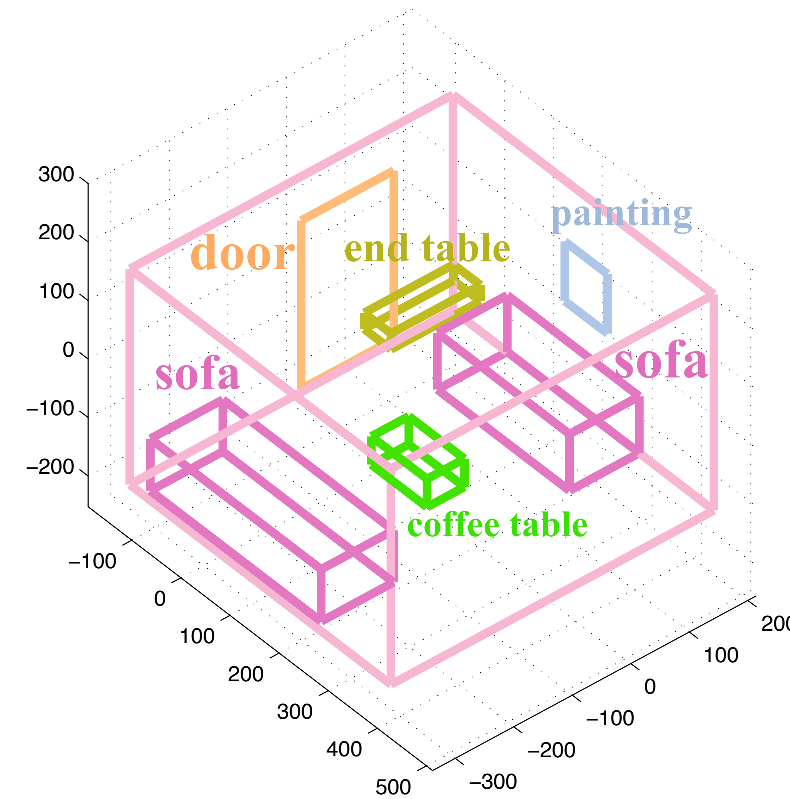


Final output

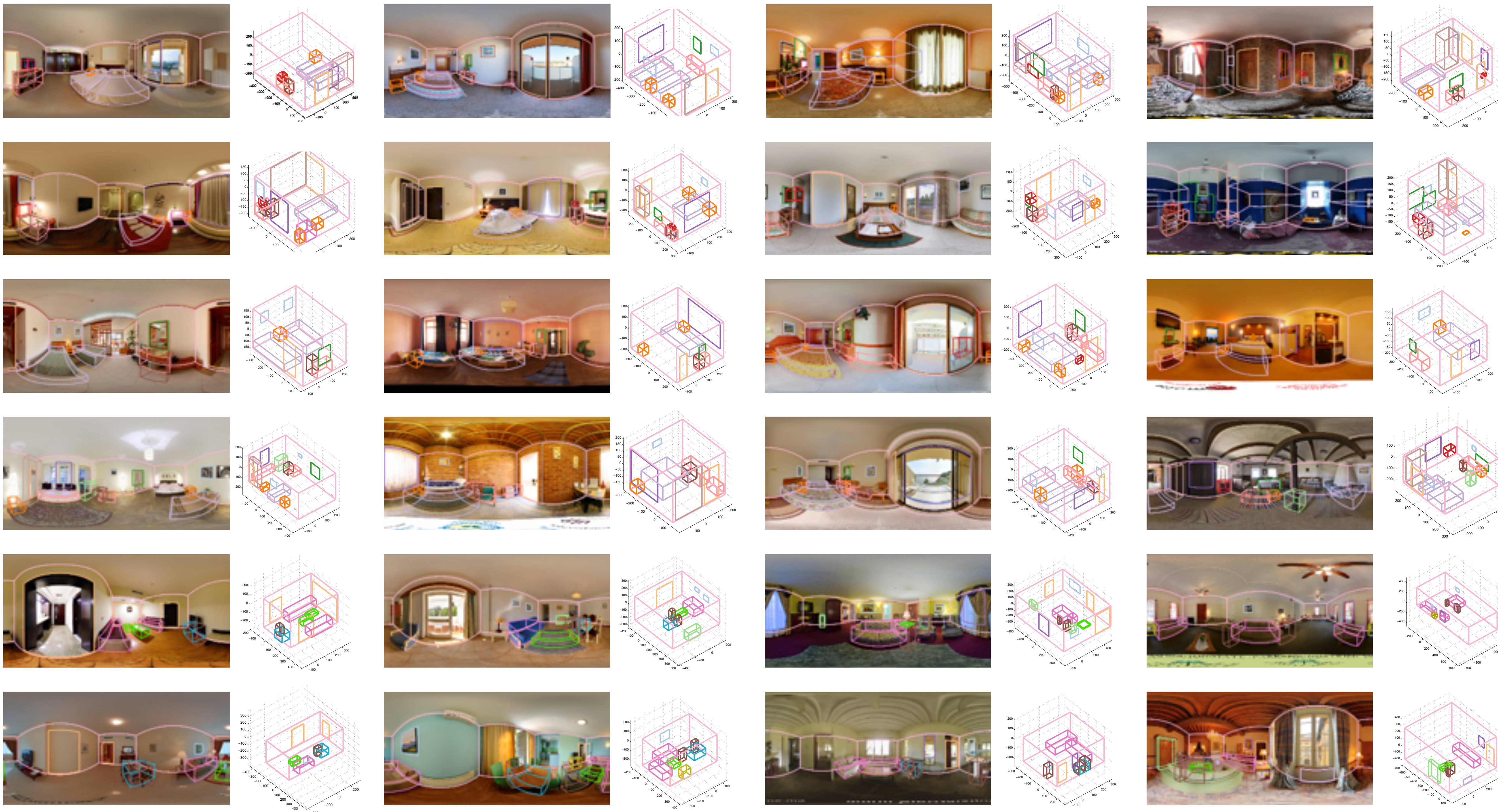


Living room


painting
door
end table
sofa
sofa
sofa
coffee table


door
end table
painting
sofa
sofa
coffee table
300
200
100
0
−100
−200
−100
0
100
200
300
400
500
200
100
0
−100
−200
−300

Automatic Recognition Results

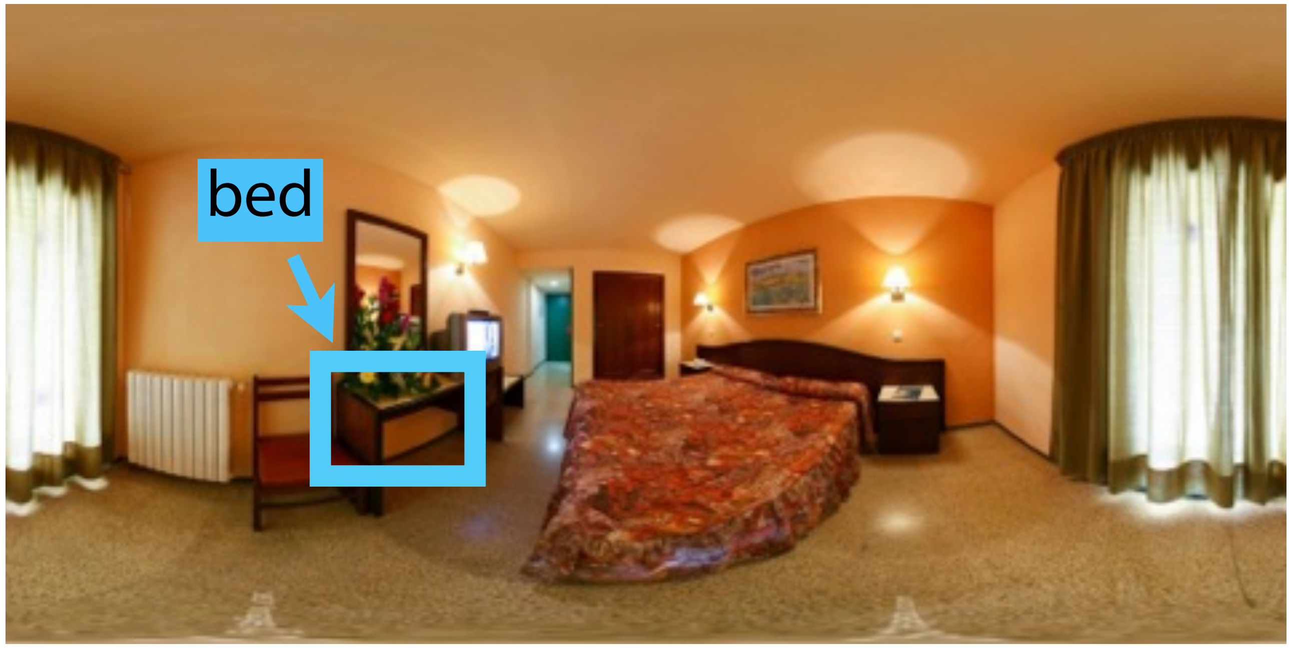


Hundreds of results are available in our website.

Analysis

How does 3D context help?

- Helps to decide 3D sizes of objects

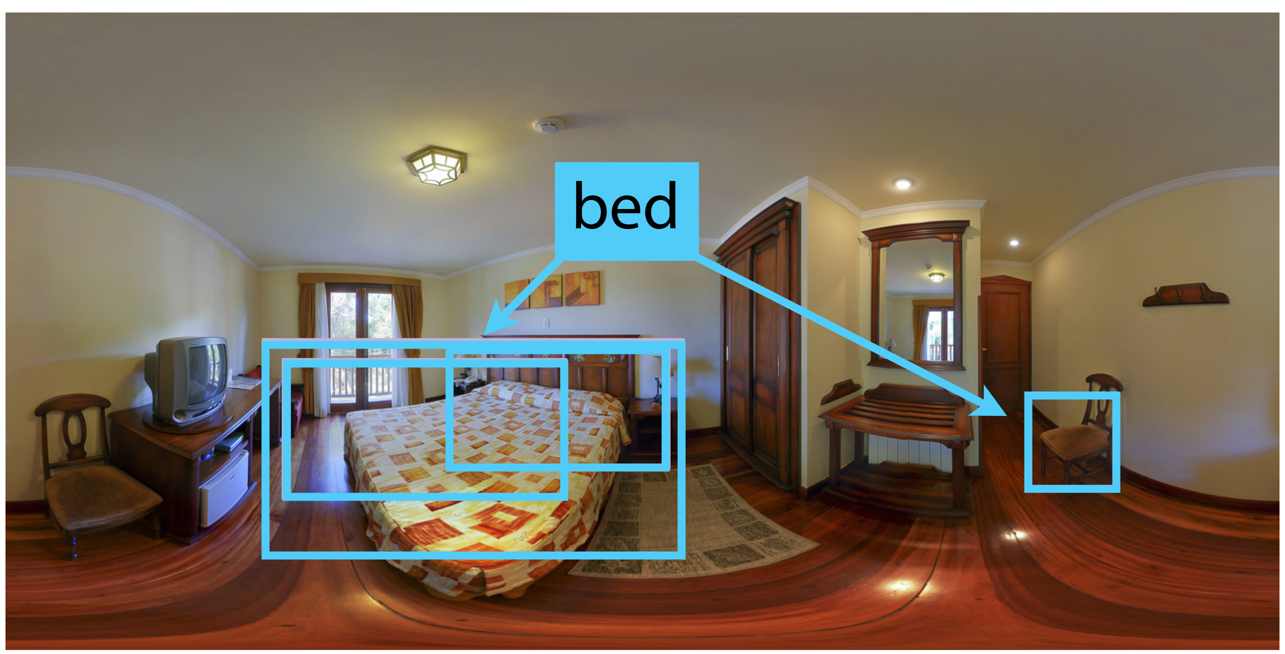


DPM: Wrong relative size

PanoContext

How does 3D context help?

- Helps to decide sizes of objects
- Helps to decide number of objects

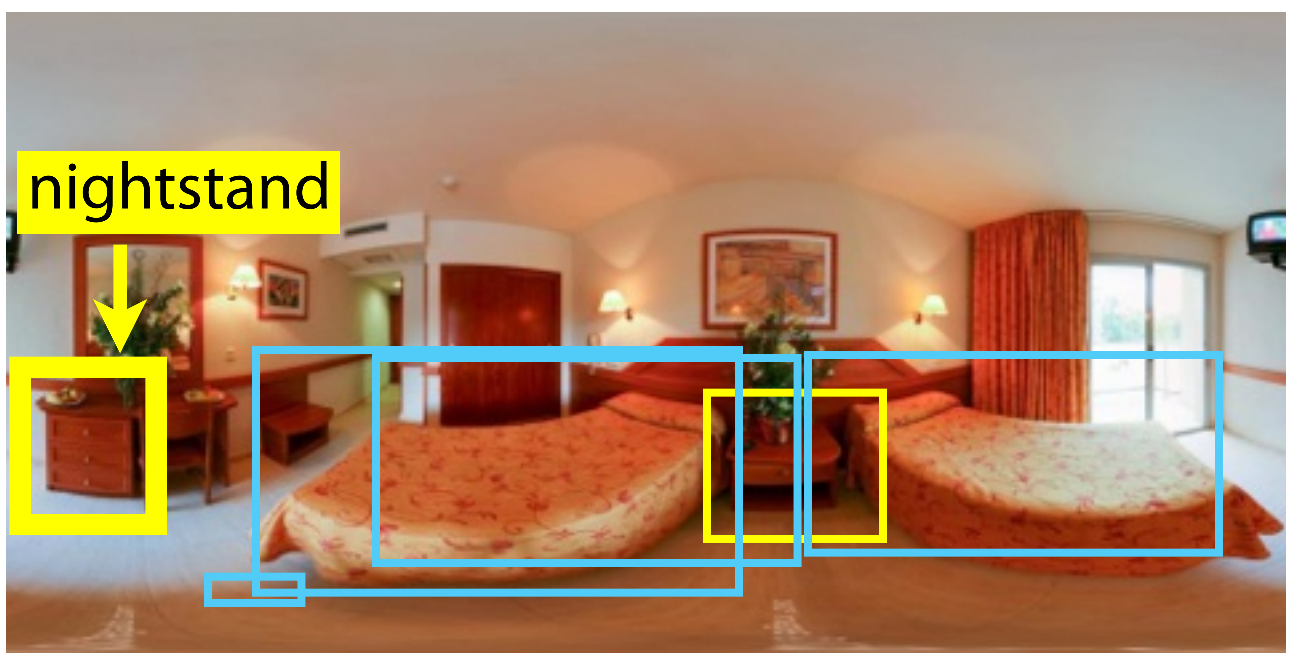


DPM: Wrong number of objects

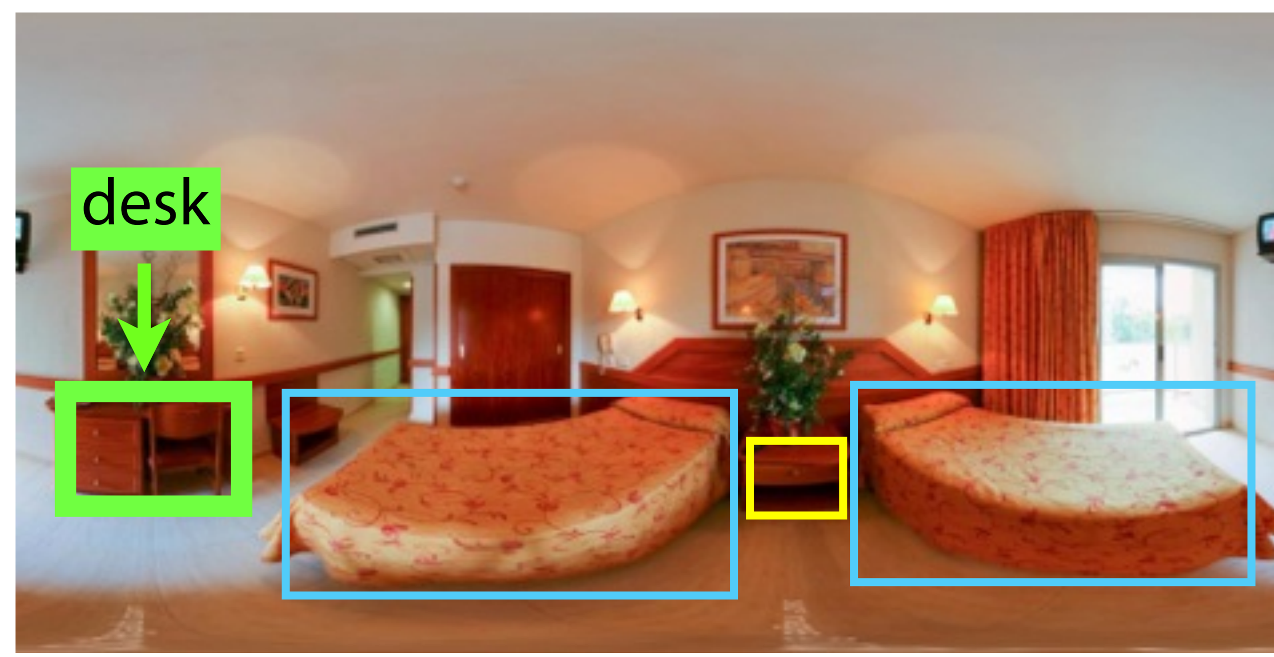
Our detection

How does 3D context help?

- Helps to decide sizes of objects
- Helps to decide number of objects
- Helps to constrain relative position



DPM: Wrong relative position



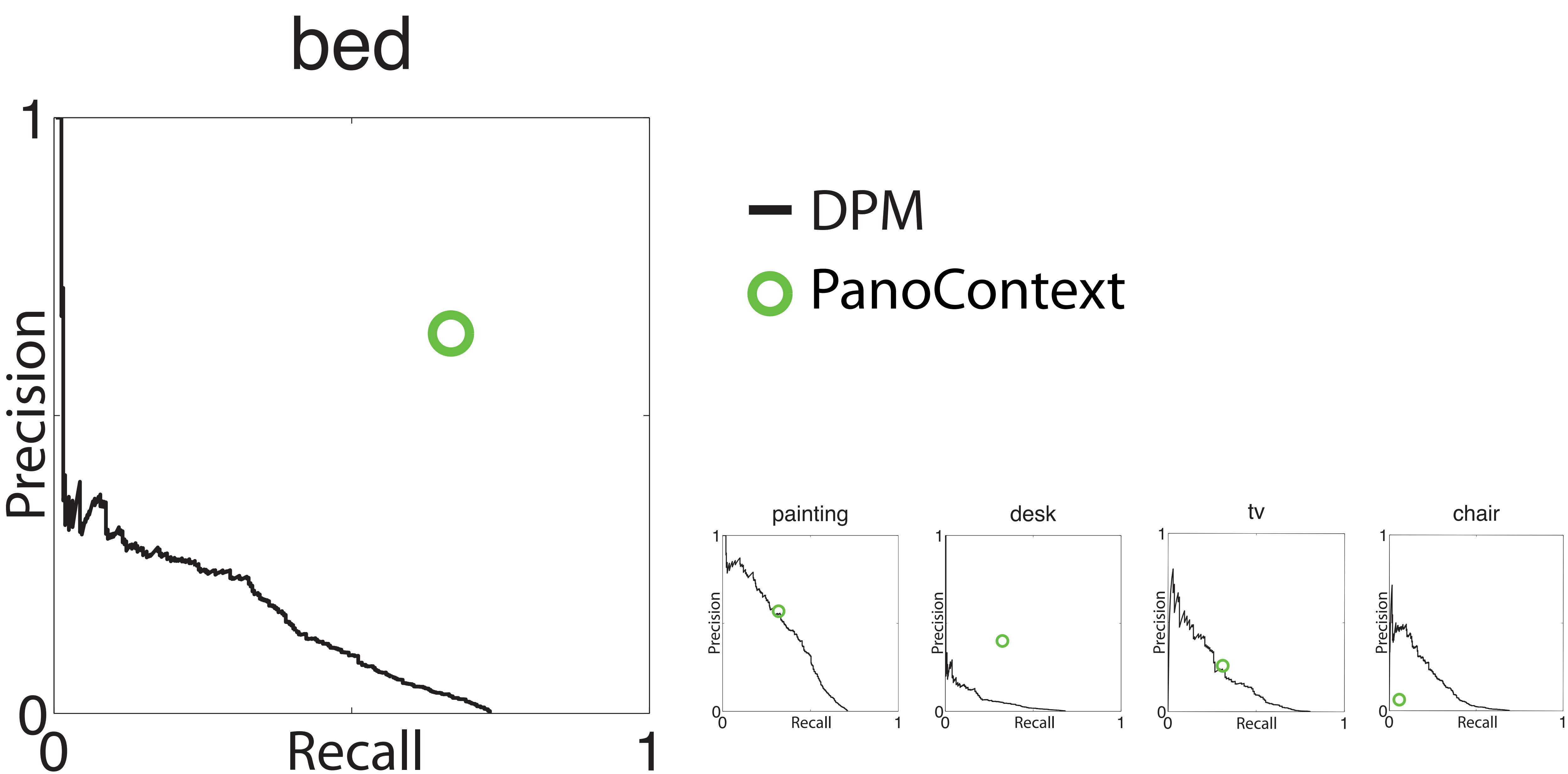
Our detection

Context vs. Appearance

- Context is as powerful as local appearance for object detection

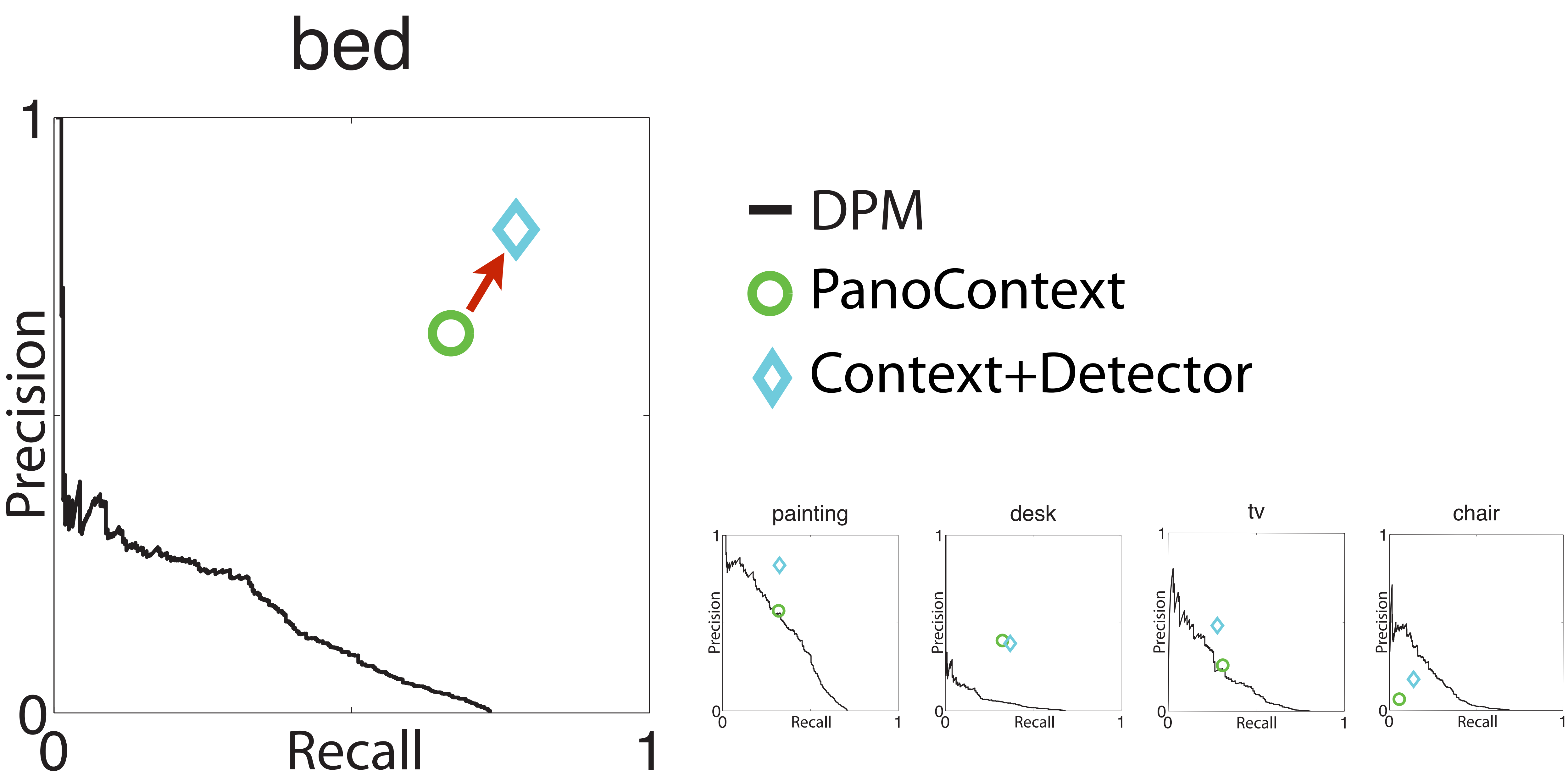
Context vs. Appearance

- Context is as powerful as local appearance for object detection

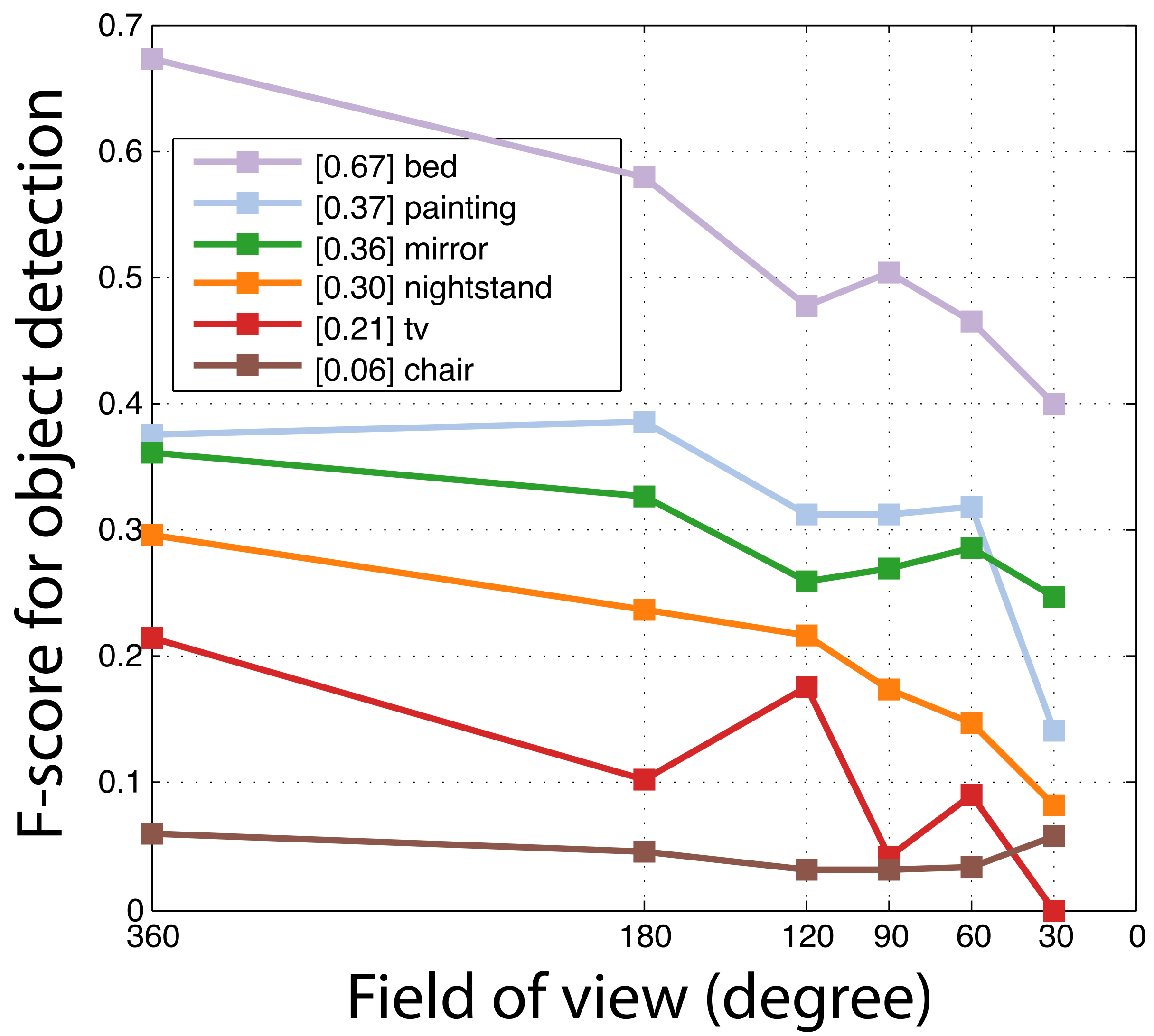


Context vs. Appearance

- Context is as powerful as local appearance for object detection
- Context is complementary with local appearance

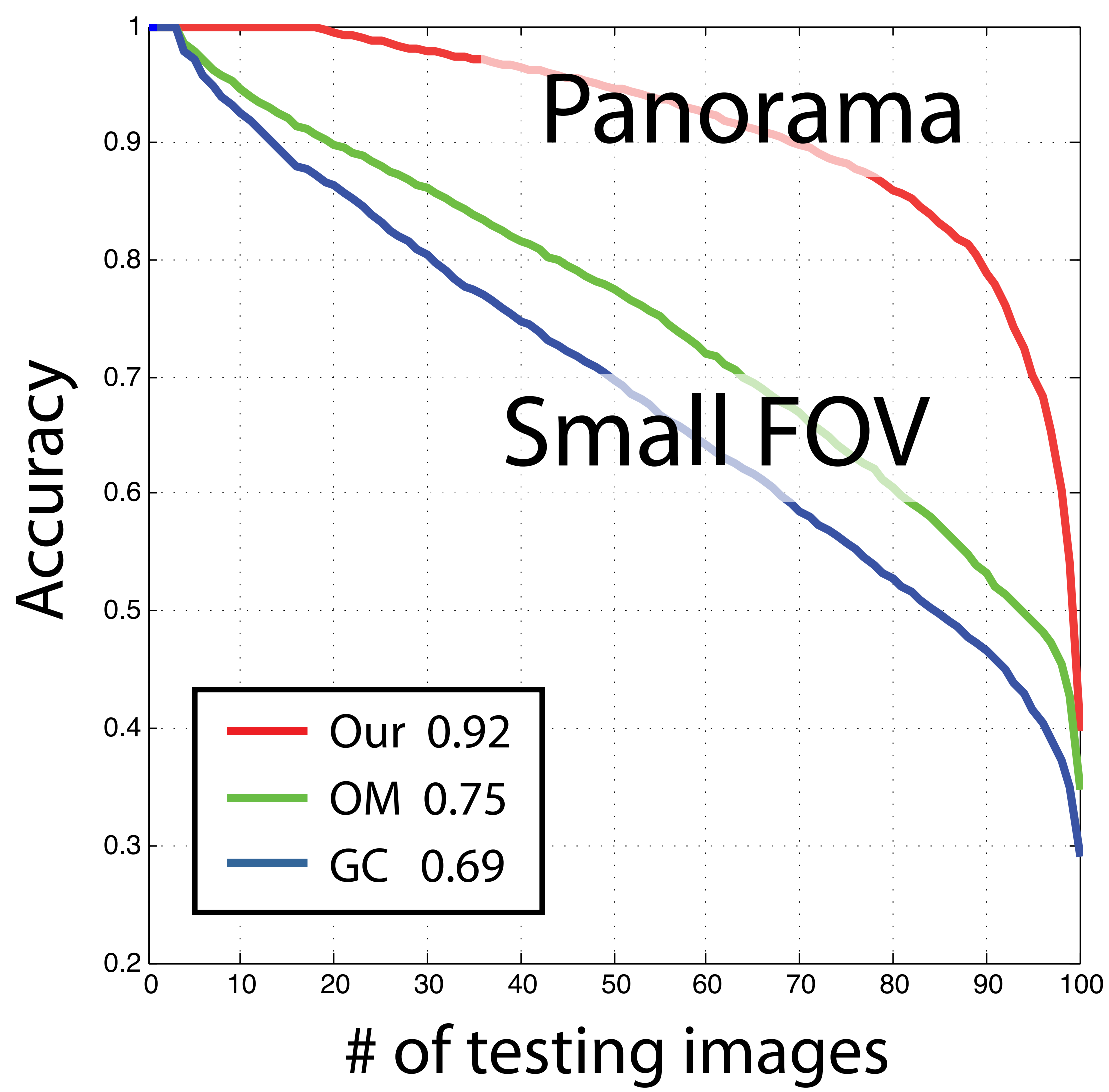


Is larger FOV helpful?



Context-based object detection

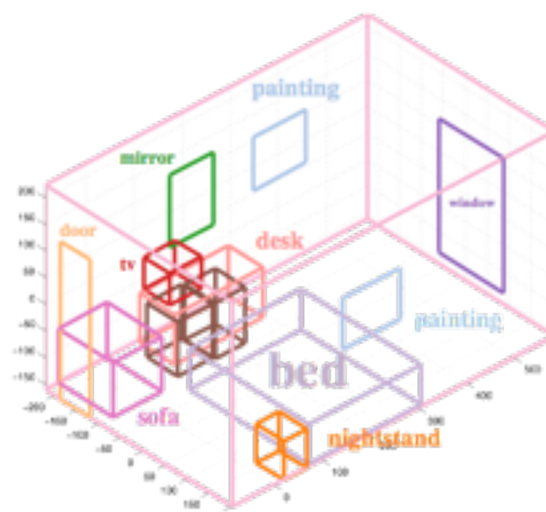
Is larger FOV helpful?

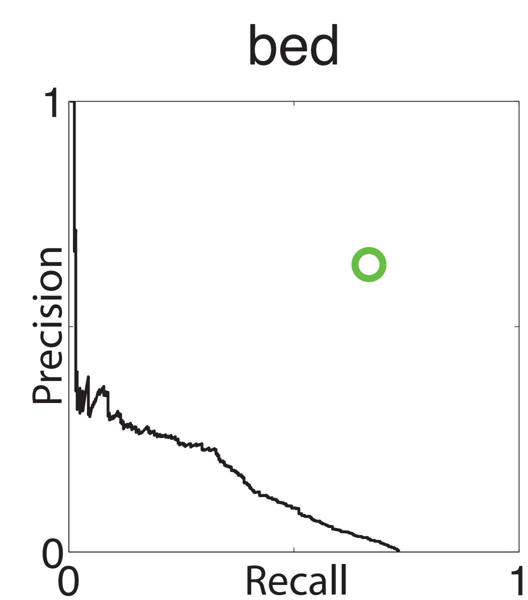


Room layout estimation

PRINCETON VISION GROUP

PanoContext





Large Field Of View

3D Whole-room

Context ≥ Detector

PRINCETON
VISION GROUP

PanoContext

Data and code available:
http://panocontext.cs.princeton.edu