PanoContext

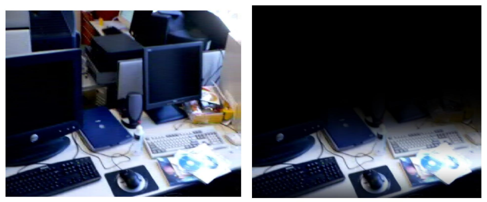
A Whole-room 3D Context Model for Panoramic Scene Understanding

by Yinda Zhang, Shuran Song, Ping Tan, Jianxiong Xiao

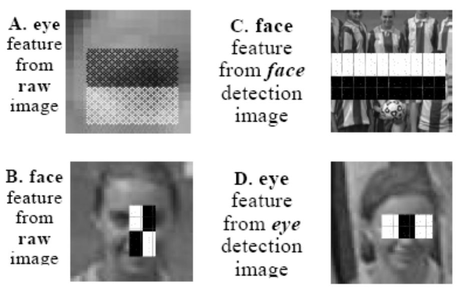
Presented by:
William Xie

Existing Context models

Torralba, Sinha (2001)

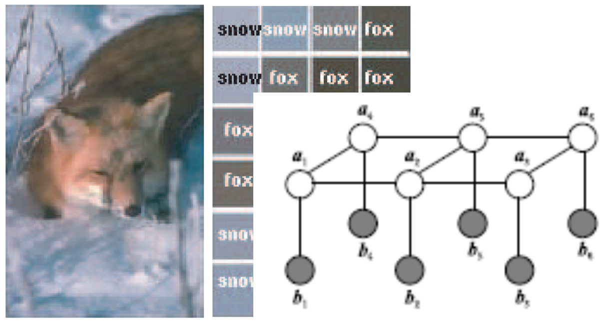


Fink & Perona (2003)

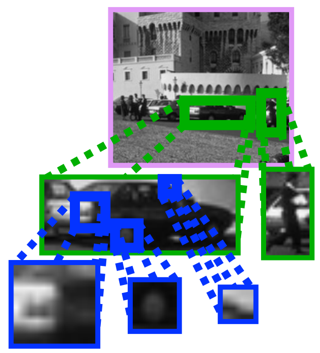


Desai, Ramanan, and Fowlkes (2009)

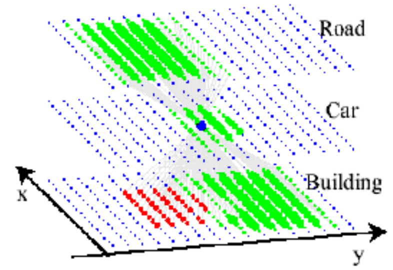
Carbonetto, de Freitas & Barnard (2004)



Sudderth, Torralba, Wilsky, Freeman (2005)

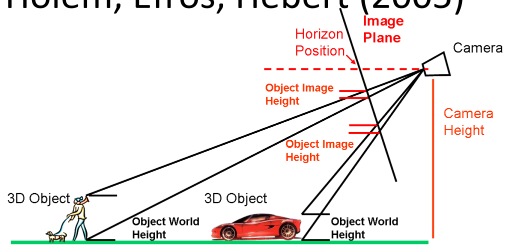


Torralba Murphy Freeman (2004)

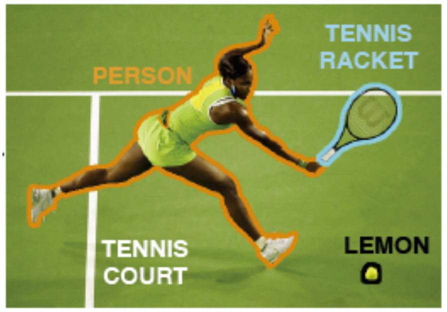


Heitz and Koller (2008)

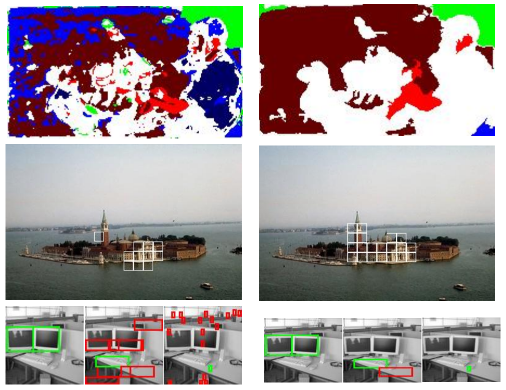
Hoiem, Efros, Hebert (2005)



Rabinovich et al (2007)



Kumar, Hebert (2005)



	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%

Slide credit: Zhang et al.

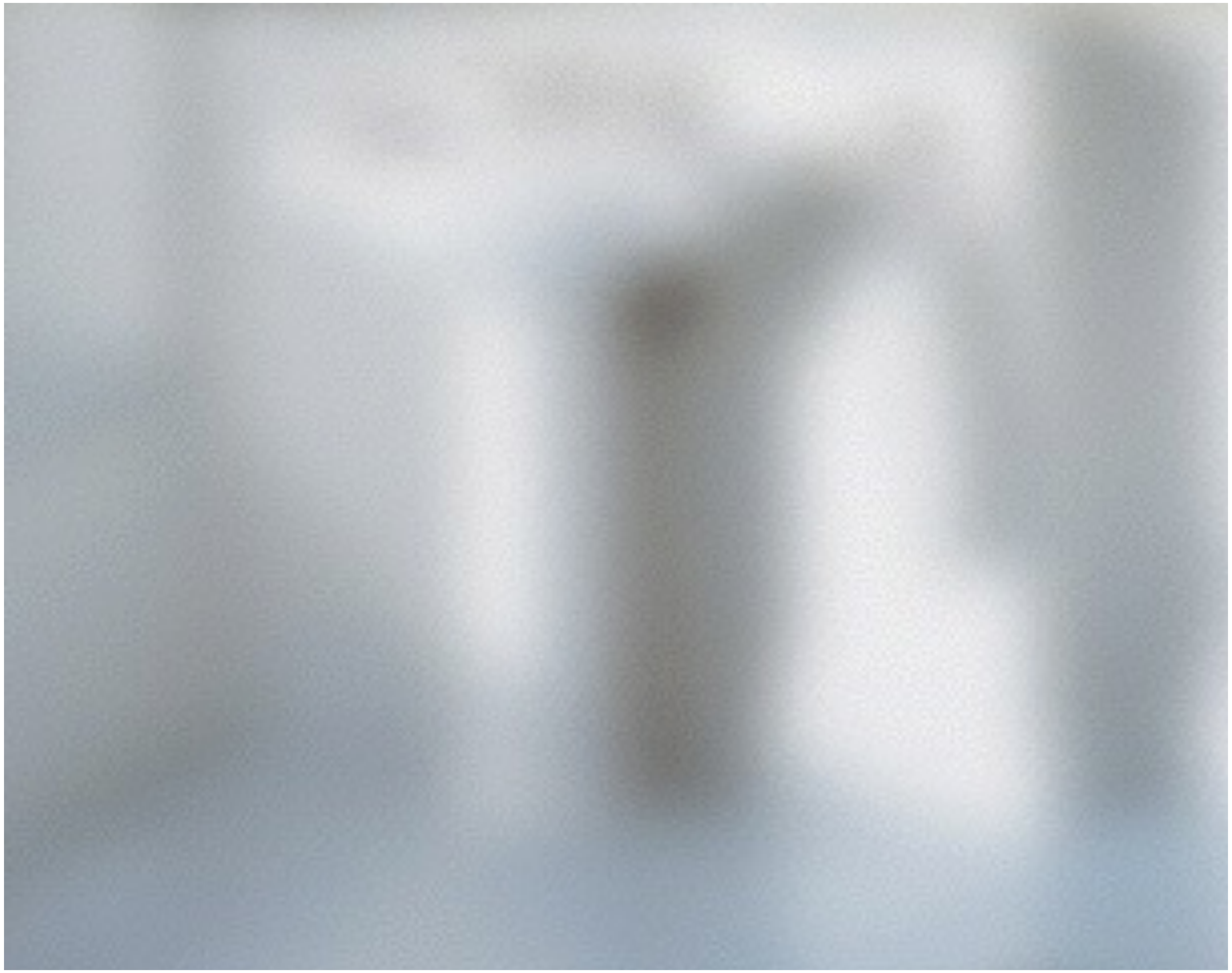
What is this object?

Slide credit: Zhang et al.

What is this object?

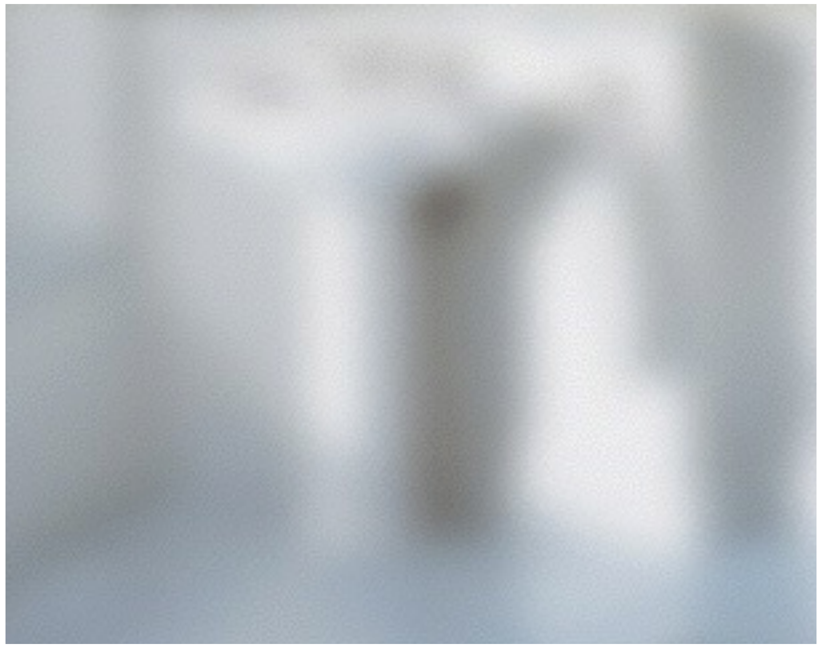
Slide credit: Zhang et al.

What is this object?



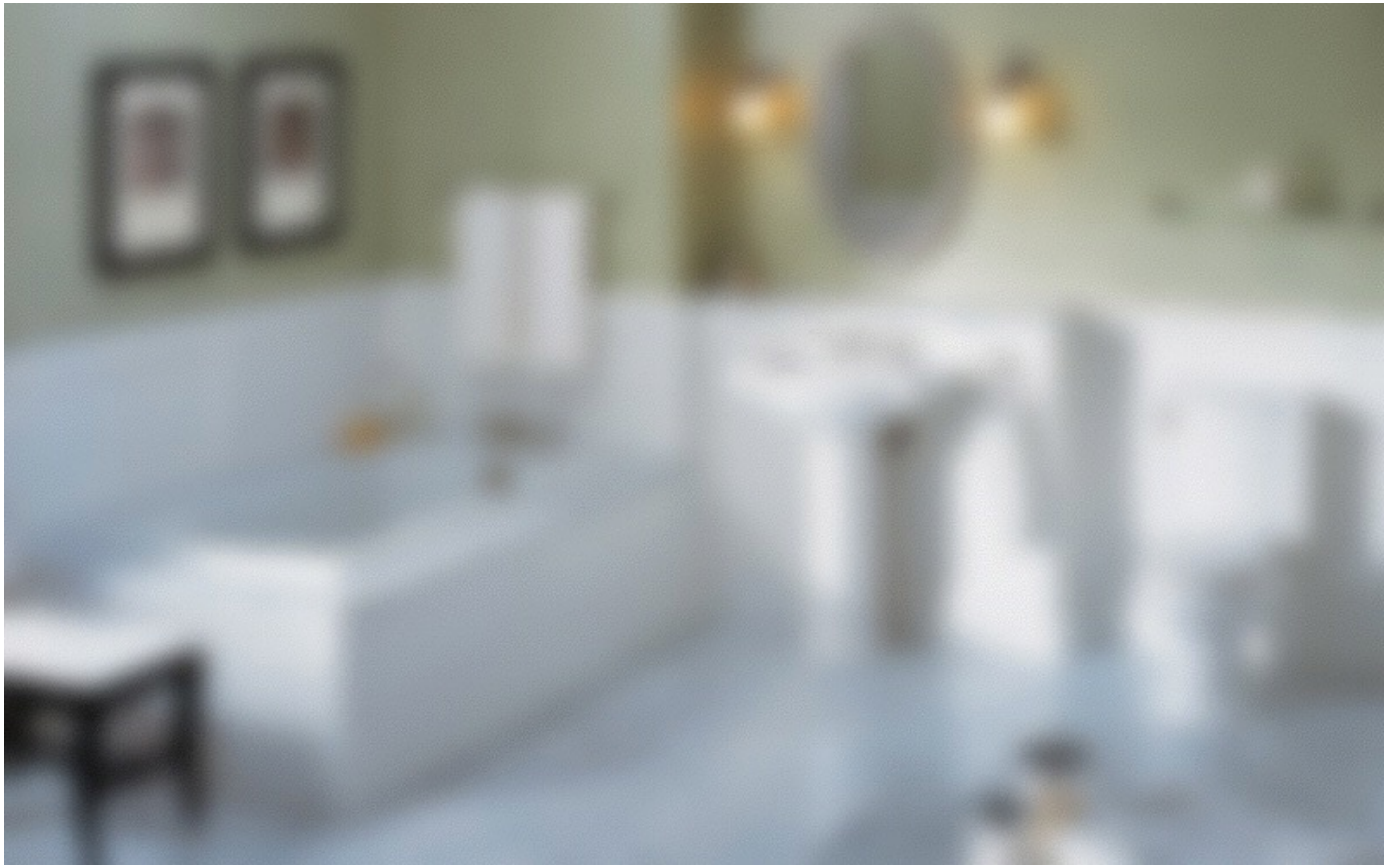
Slide credit: Zhang et al.

What is this object?



Slide credit: Zhang et al.

What is this object?



Slide credit: Zhang et al.

What is this object?

Slide credit: Zhang et al.

Why didn’ context help?

Why didn’ context help?

Perhaps we are not using the right data

PASCAL VOC

- On average: 1.5 object classes and 2.7 object instances per image
- Average camera field of view: 40° - 60° horizontal

Human Vision

- 180° horizontal field of view
- Ability to see depth
- Ability to change viewpoint

Remedy

PanoContext

Slide credit: Zhang et al.

PanoContext

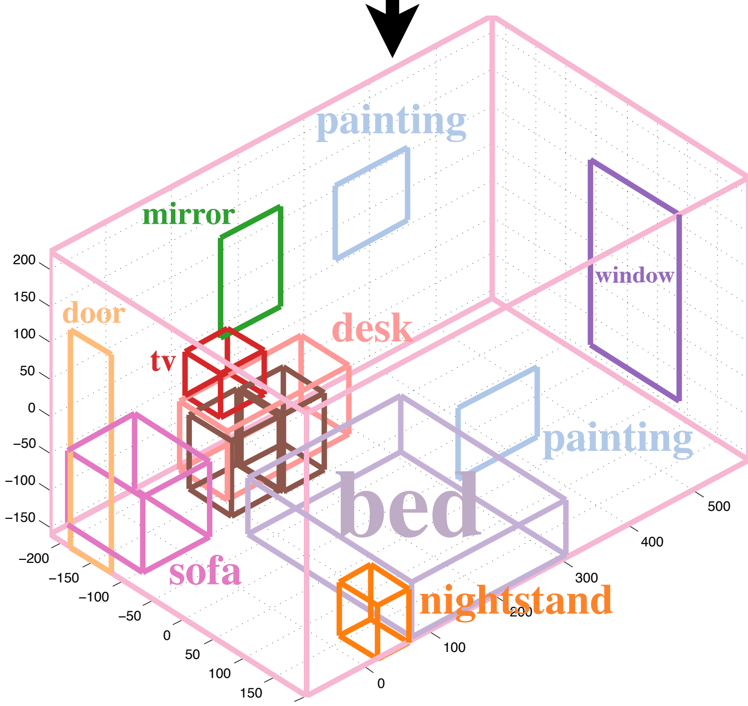
Input: Panorama

Slide credit: Zhang et al.

PanoContext

Input: Panorama

Output: 2D projected result



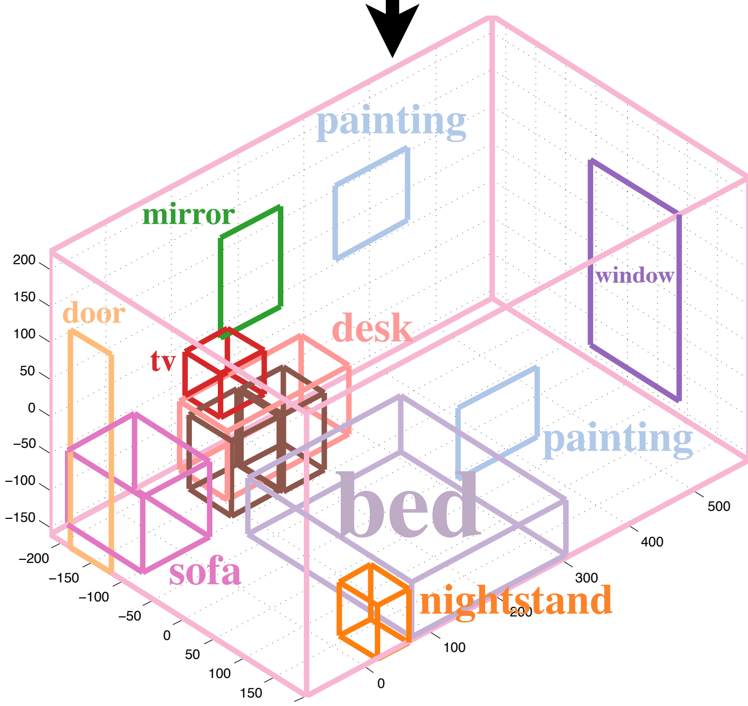
Output: 3D model

Slide credit: Zhang et al.

PanoContext

Input: Panorama

Output: 2D projected result



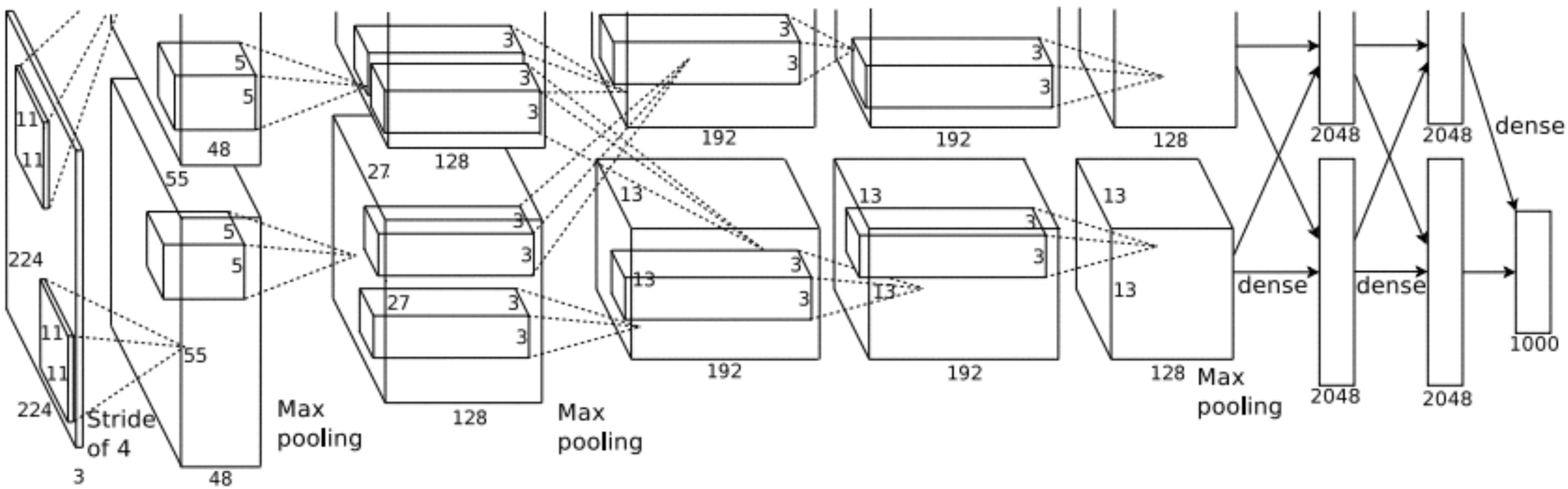
Output: 3D model

Output: 3D room exploration

Slide credit: Zhang et al.

Pipeline

Pipeline



Krizhevsky, Alex, et al. "Imagenet classification with deep convolutional neural networks." NIPS. 2012.

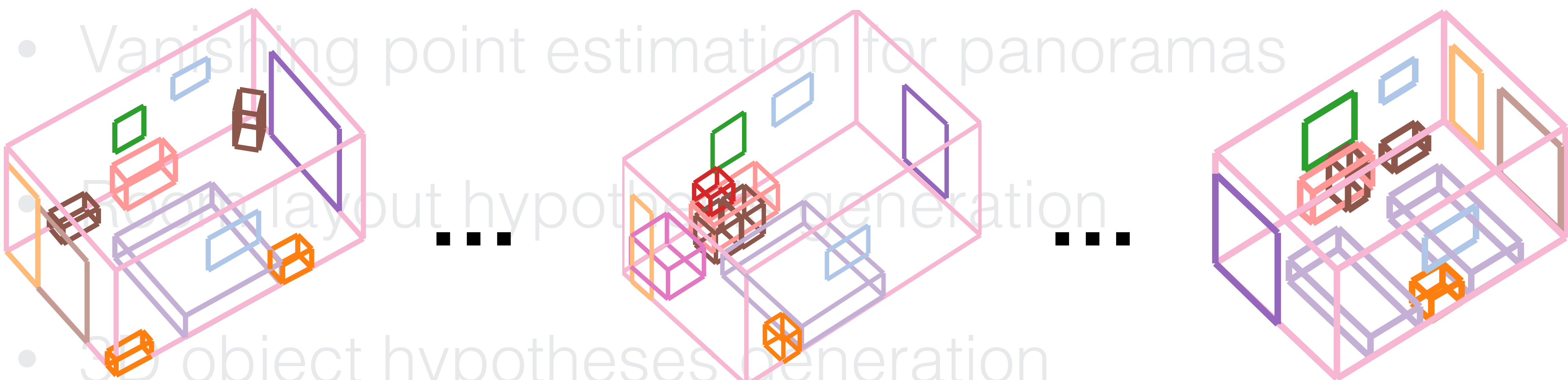
Pipeline

- Vanishing point estimation for panoramas
- Room layout hypothesis generation
- 3D object hypotheses generation
- Whole-room scene hypotheses generation
- Data-driven holistic ranking

Pipeline

- Vanishing point estimation for panoramas
- Room layout hypothesis generation
- 3D object hypotheses generation
- Whole-room scene hypotheses generation
- Data-driven holistic ranking

Pipeline



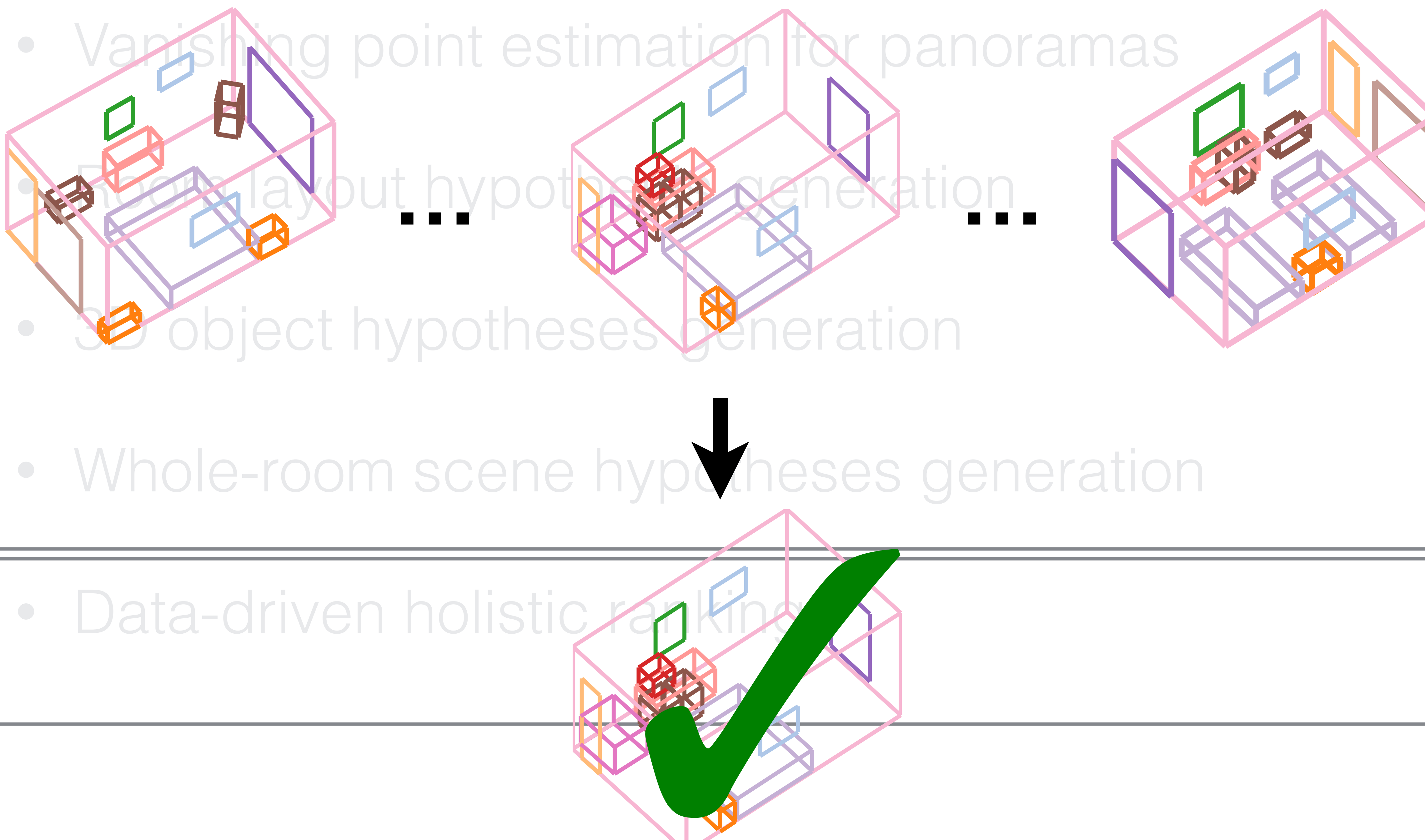
- Vanishing point estimation for panoramas
- Room layout hypotheses generation
- 3D object hypotheses generation
- Whole-room scene hypotheses generation

- Data-driven holistic ranking

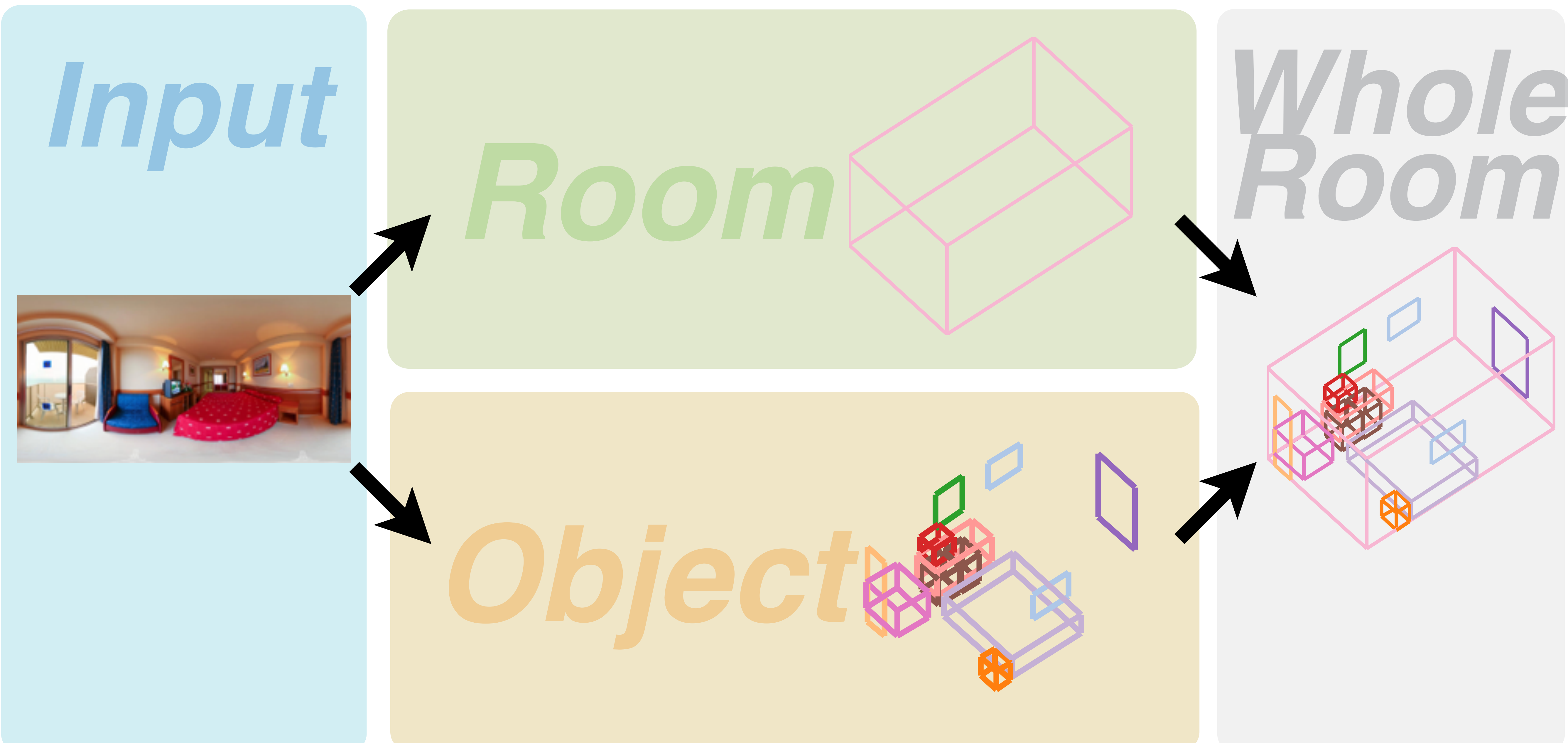
Pipeline

- Vanishing point estimation for panoramas
- Room layout hypotheses generation
- 3D object hypotheses generation
- Whole-room scene hypotheses generation

- Data-driven holistic ranking

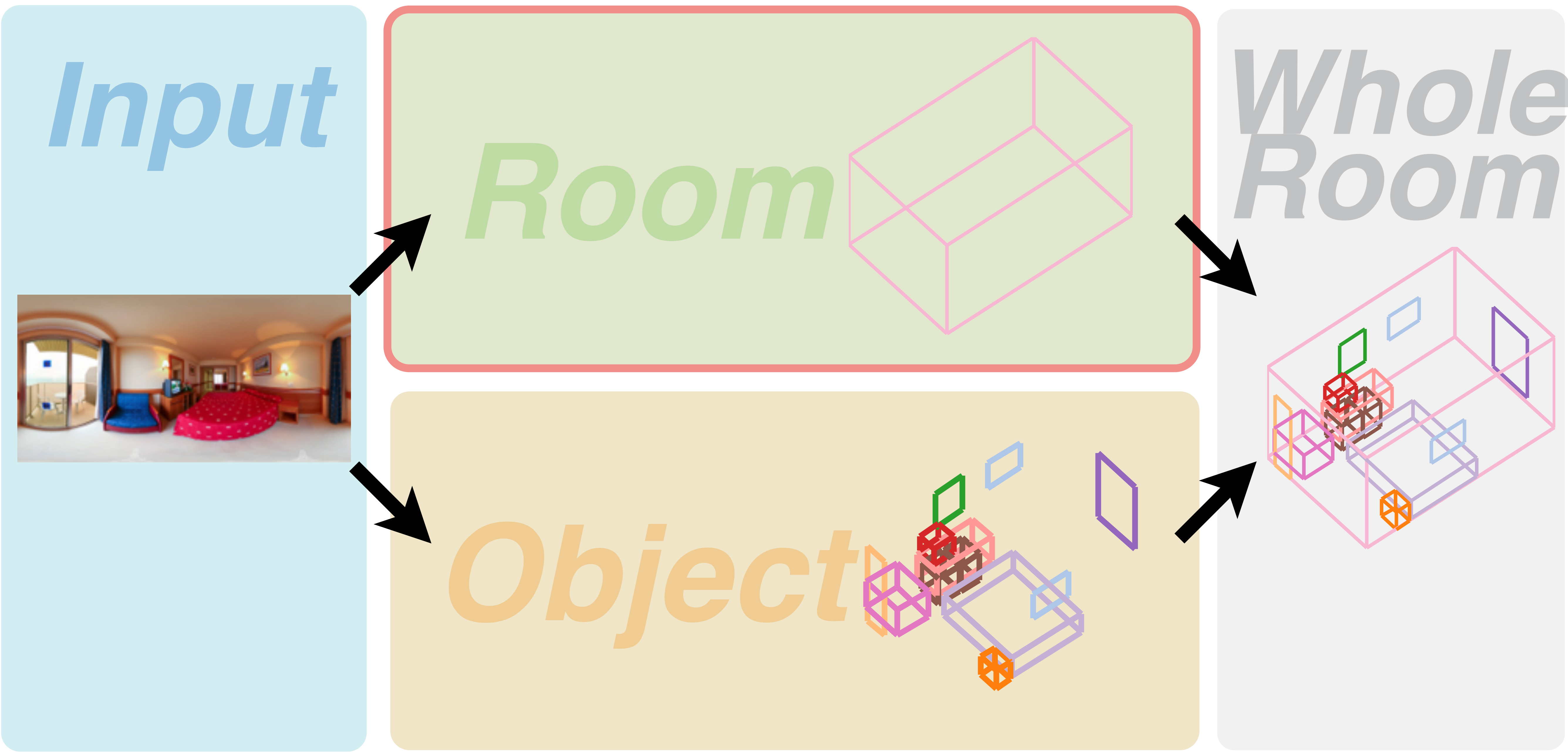


Generate a pool of hypotheses



Slide credit: Zhang et al.

Generate a pool of hypotheses

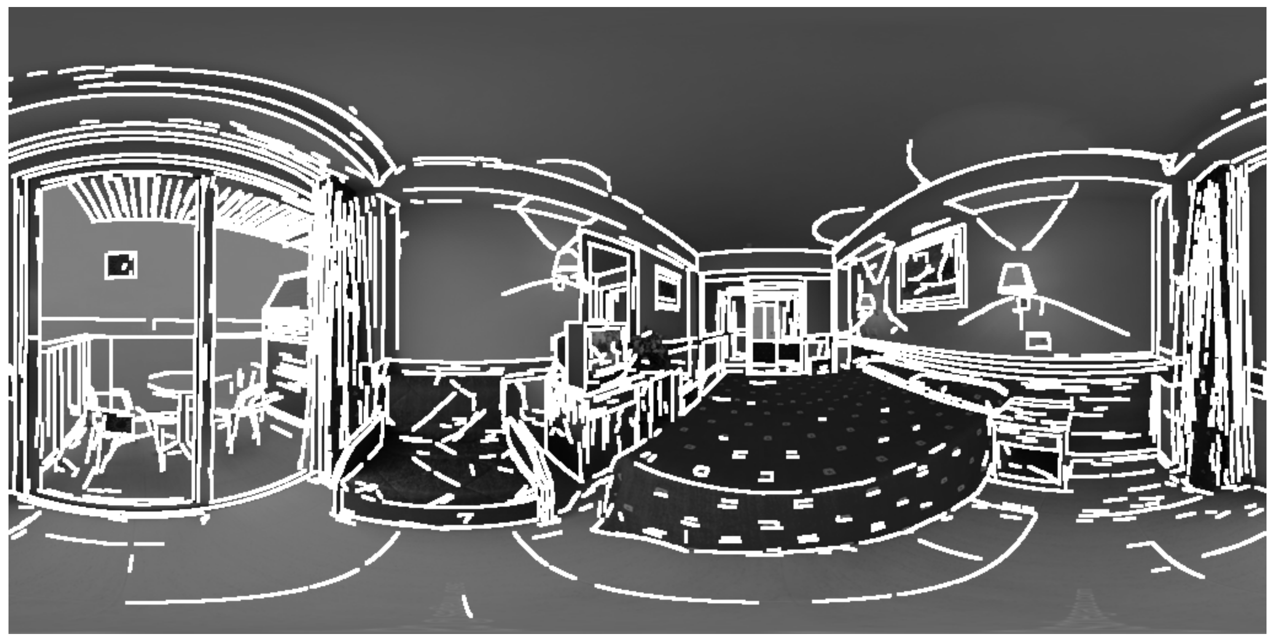


Slide credit: Zhang et al.

Room layout hypothesis

Slide credit: Zhang et al.

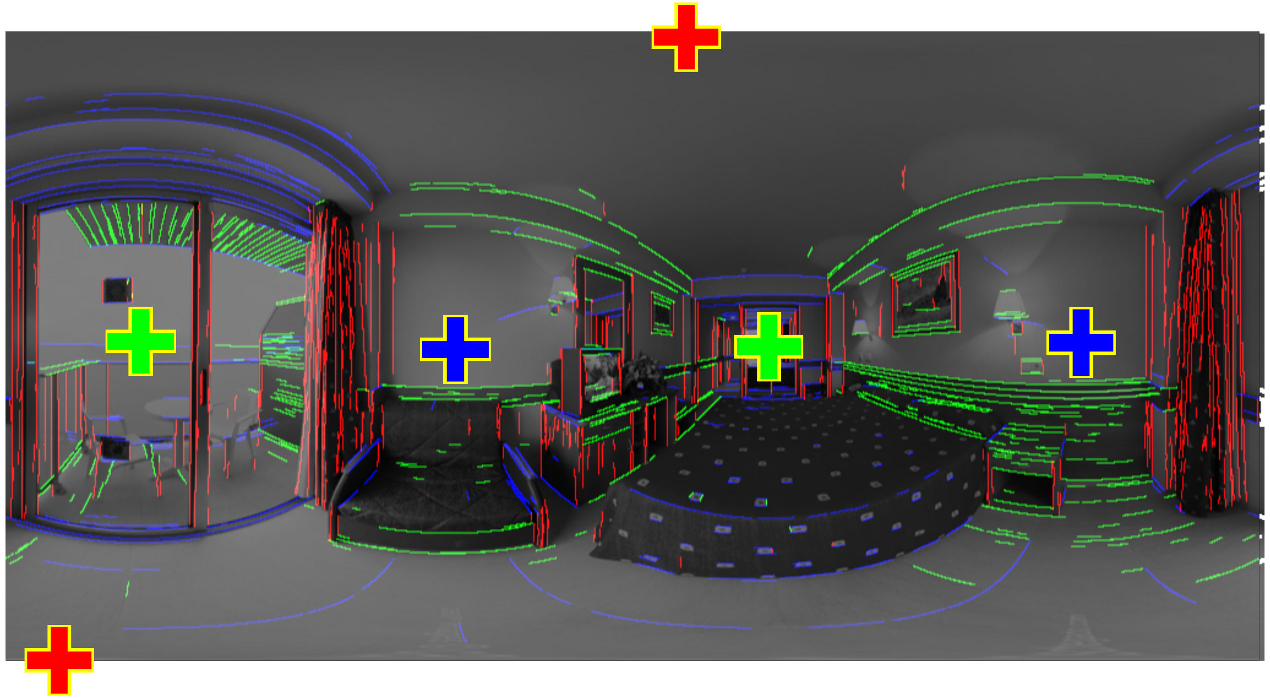
Room layout hypothesis



Line segments detection Algorithm

Slide credit: Zhang et al.

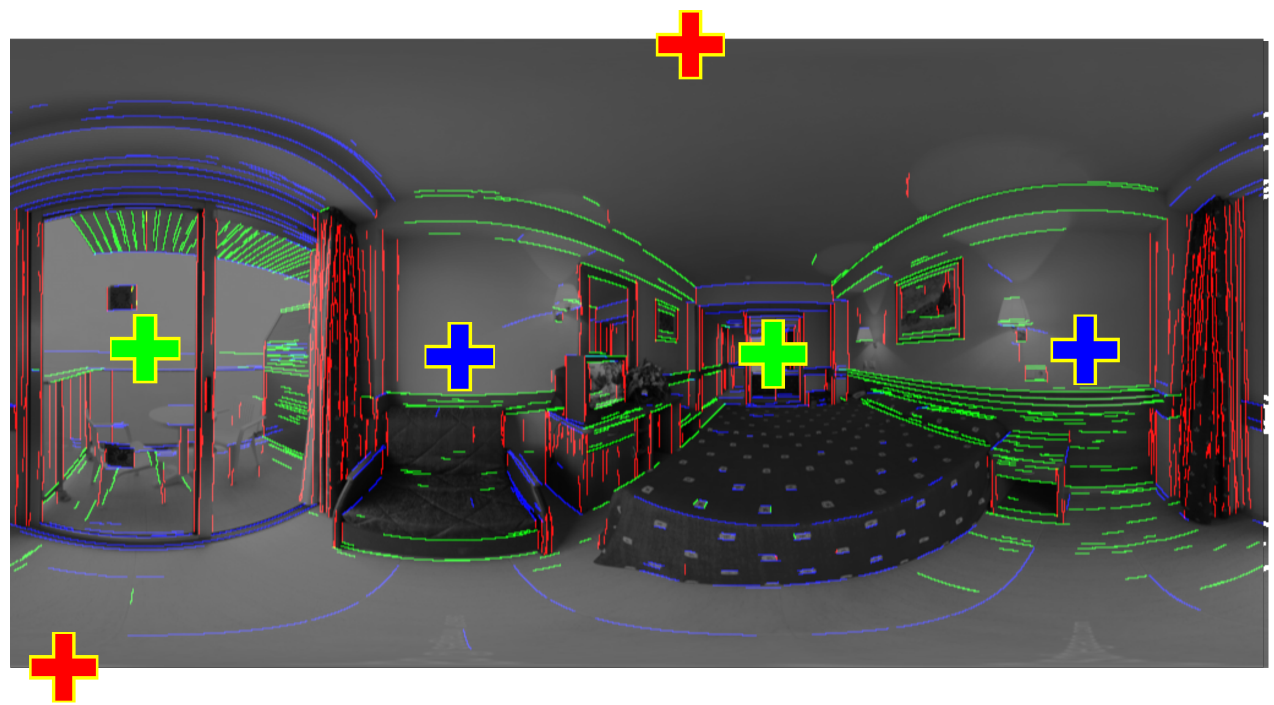
Room layout hypothesis



Hough transform for vanishing point

Slide credit: Zhang et al.

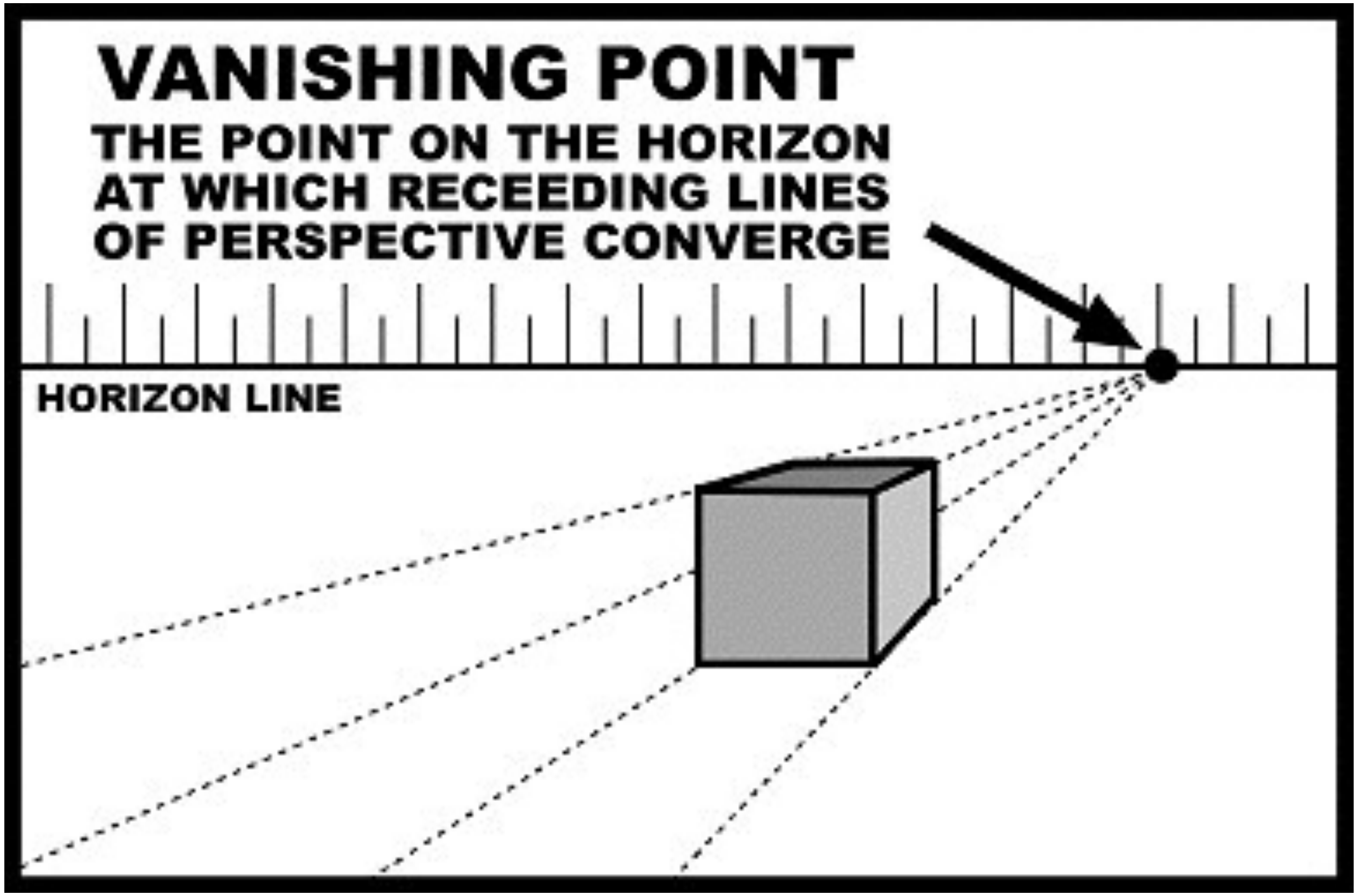
Room layout hypothesis

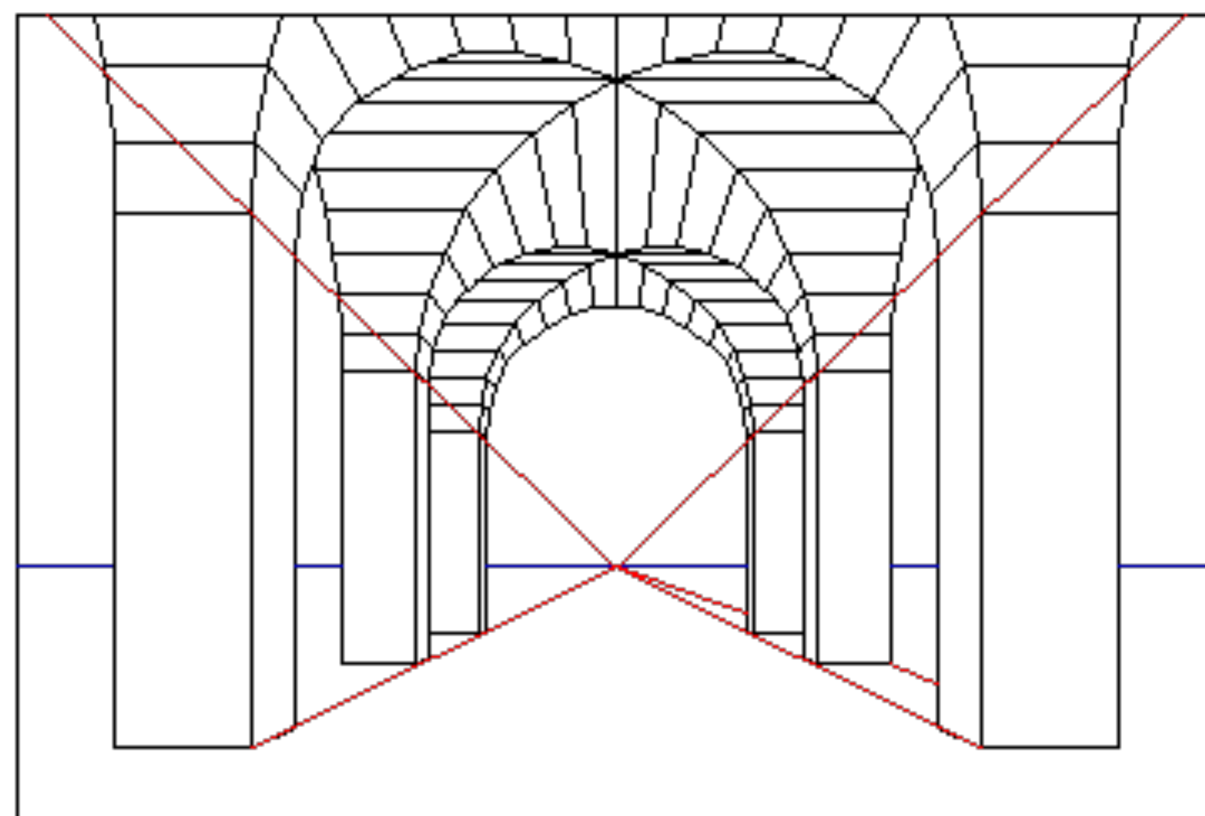


Hough transform for vanishing point

Classify a vanishing direction for each line

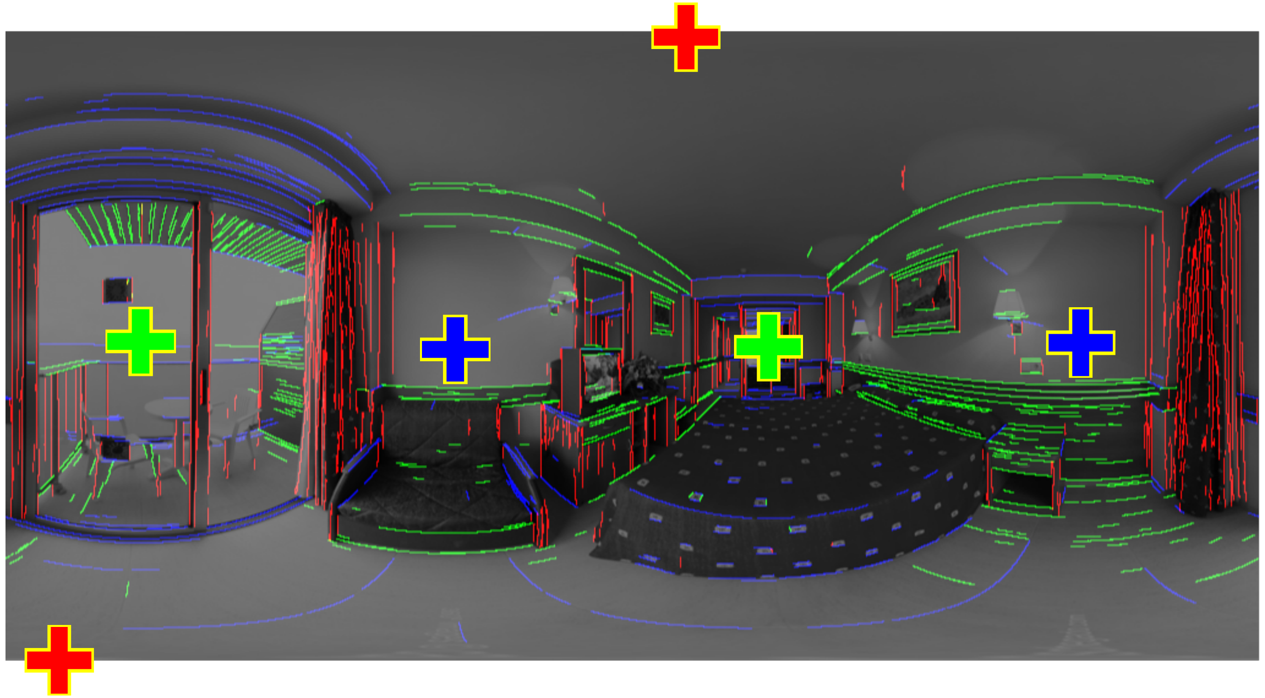
Slide credit: Zhang et al.





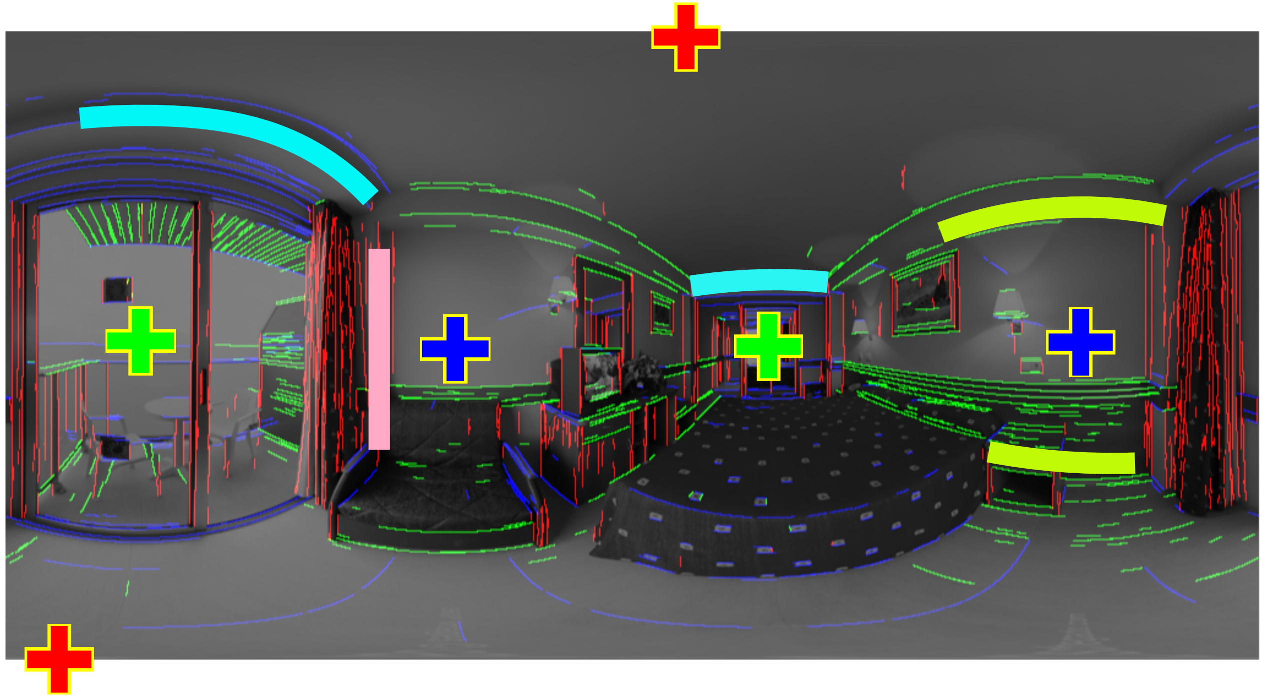
Source: Wikipedia, Emaze

Room layout hypothesis



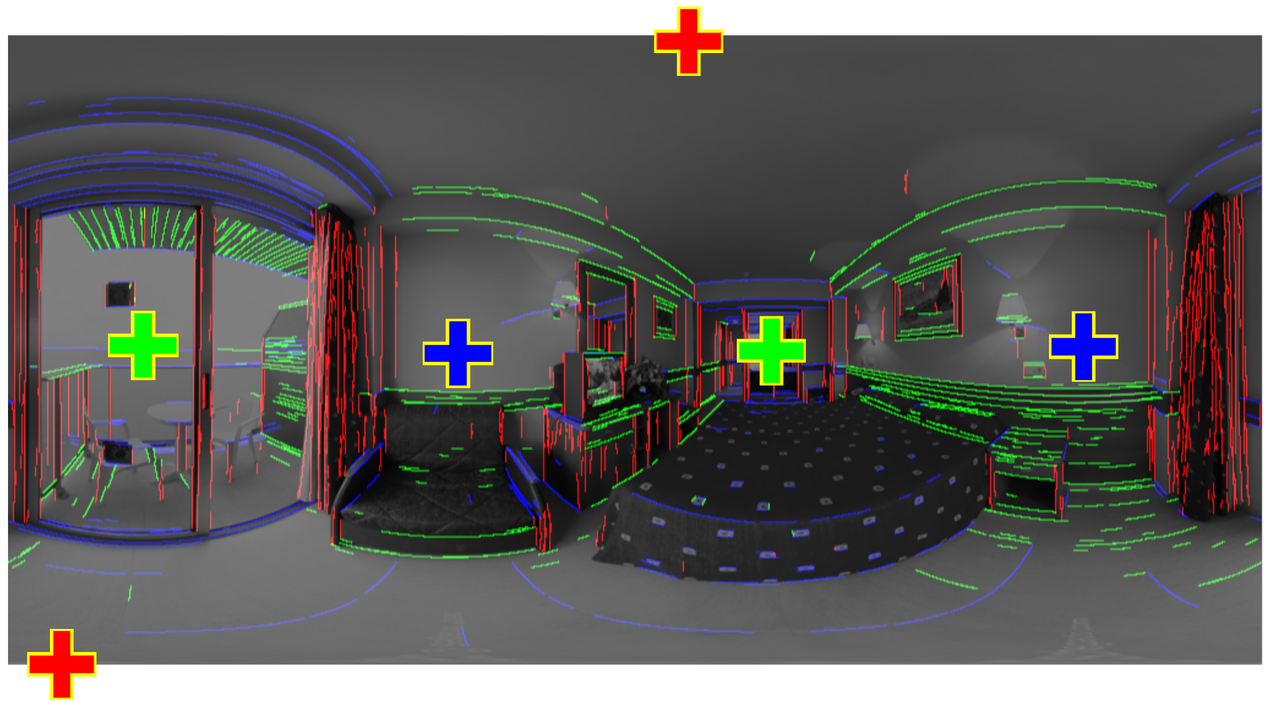
Sample 5 line segments to generate a room layout

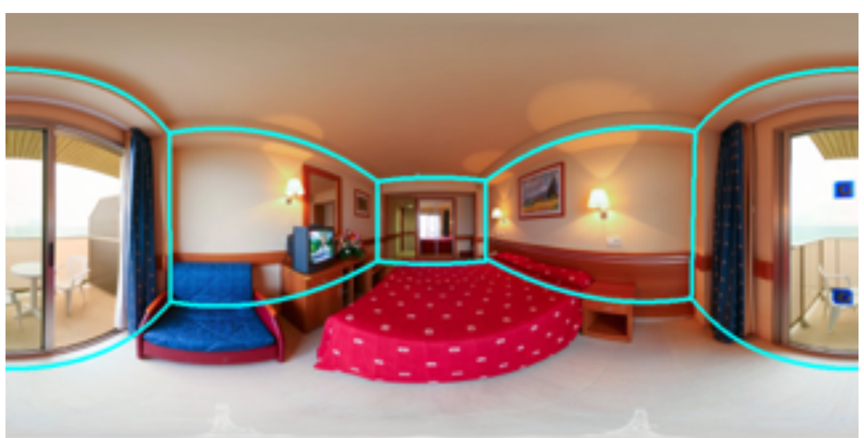
Room layout hypothesis



Sample 5 line segments to generate a room layout

Room layout hypothesis

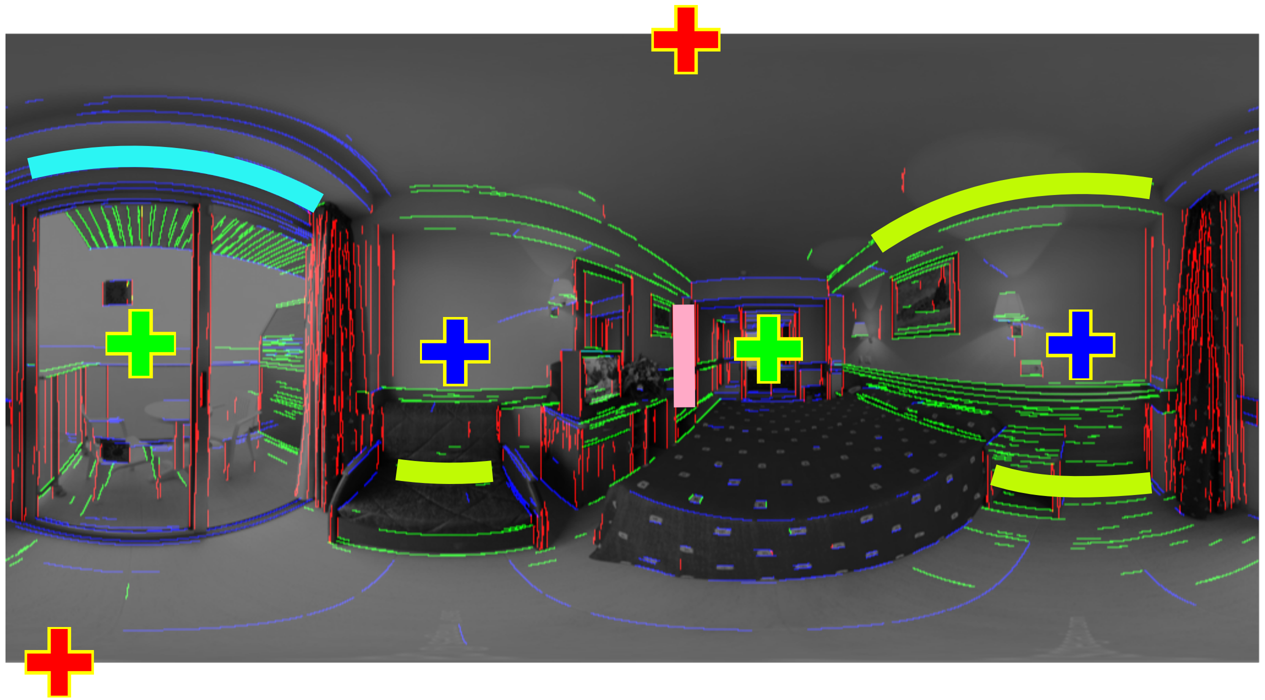




Sample 5 line segments to generate a room layout

Slide credit: Zhang et al.

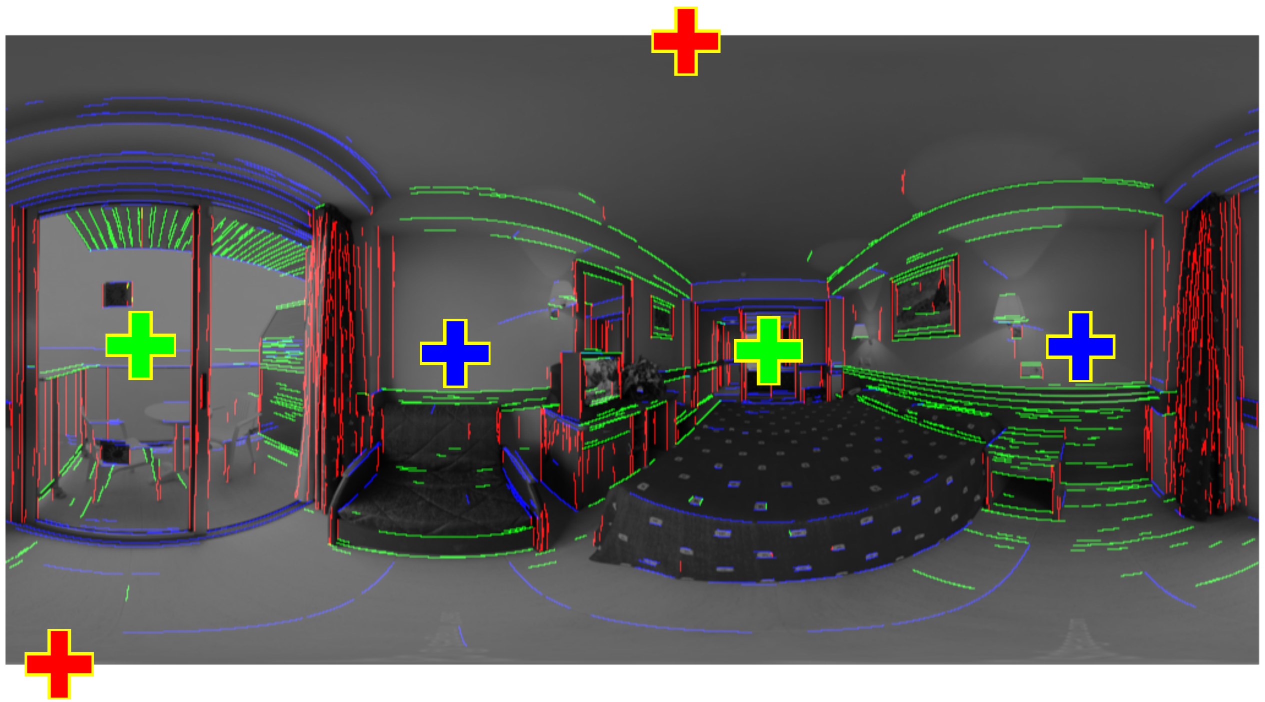
Room layout hypothesis



Sample 5 line segments to generate a room layout

Slide credit: Zhang et al.

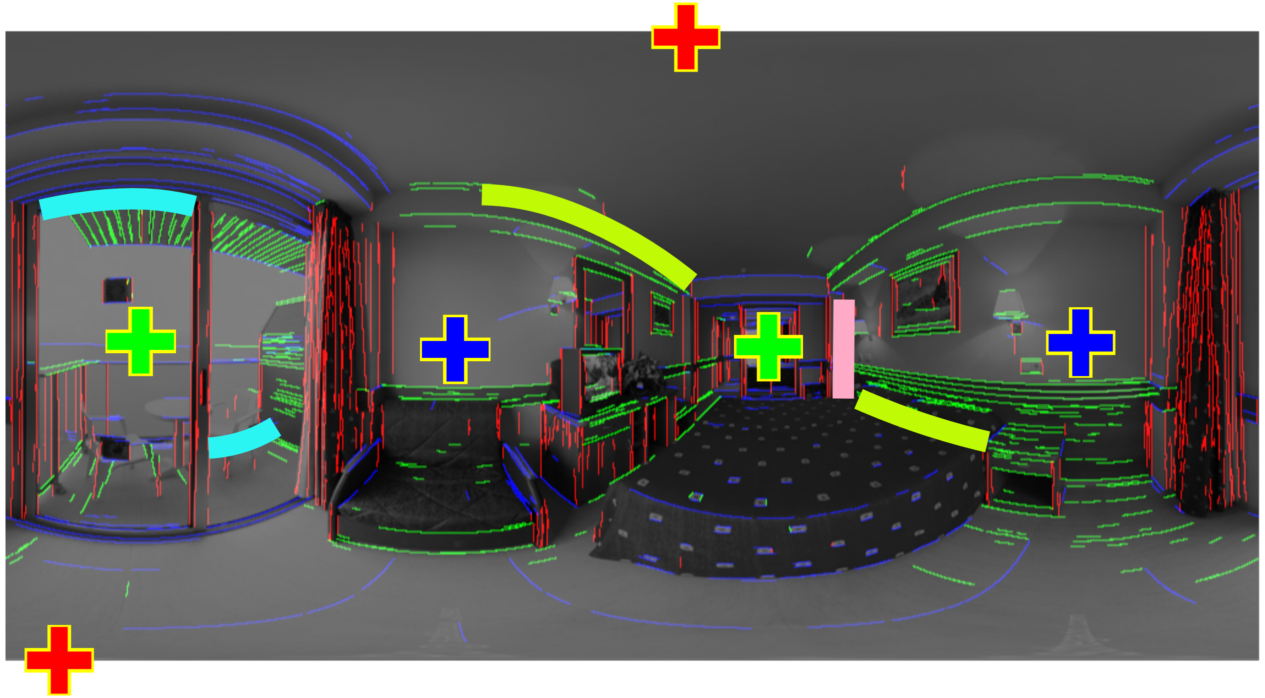
Room layout hypothesis



Sample 5 line segments to generate a room layout

Slide credit: Zhang et al.

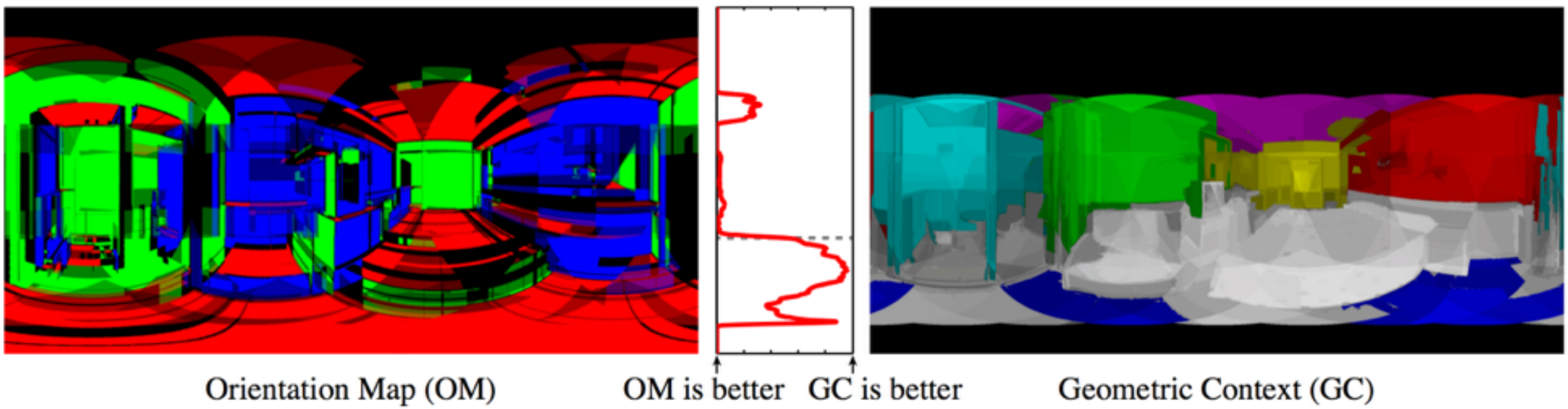
Room layout hypothesis



Sample 5 line segments to generate a room layout

Slide credit: Zhang et al.

Room layout hypothesis



Pixel-wise surface direction estimation

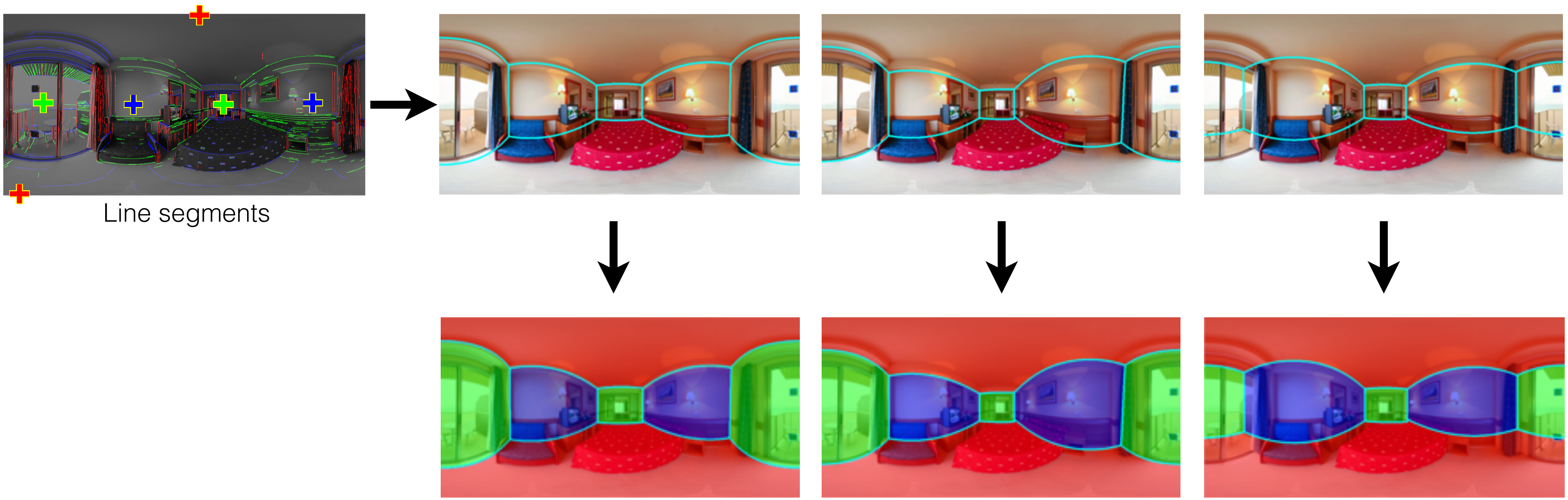
Slide credit: Zhang et al.

Room layout hypothesis

Line segments

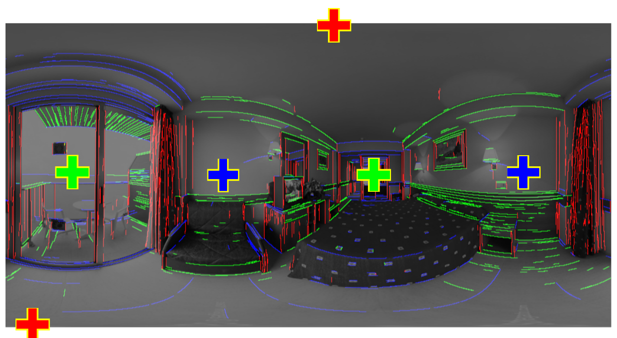
Slide credit: Zhang et al.

Room layout hypothesis

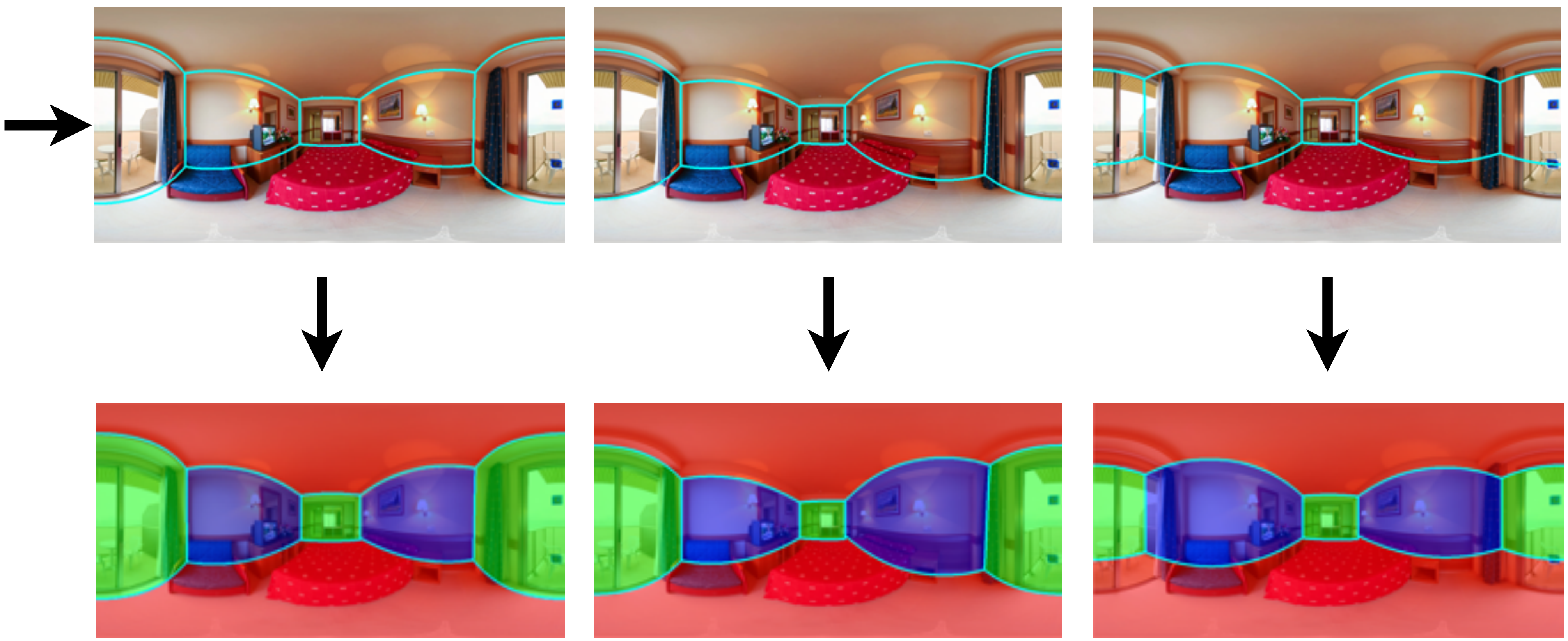


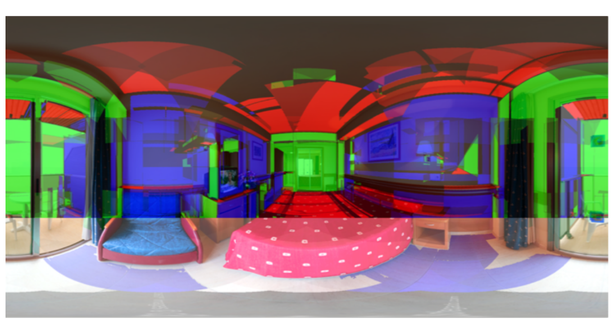
Slide credit: Zhang et al.

Room layout hypothesis



Line segments

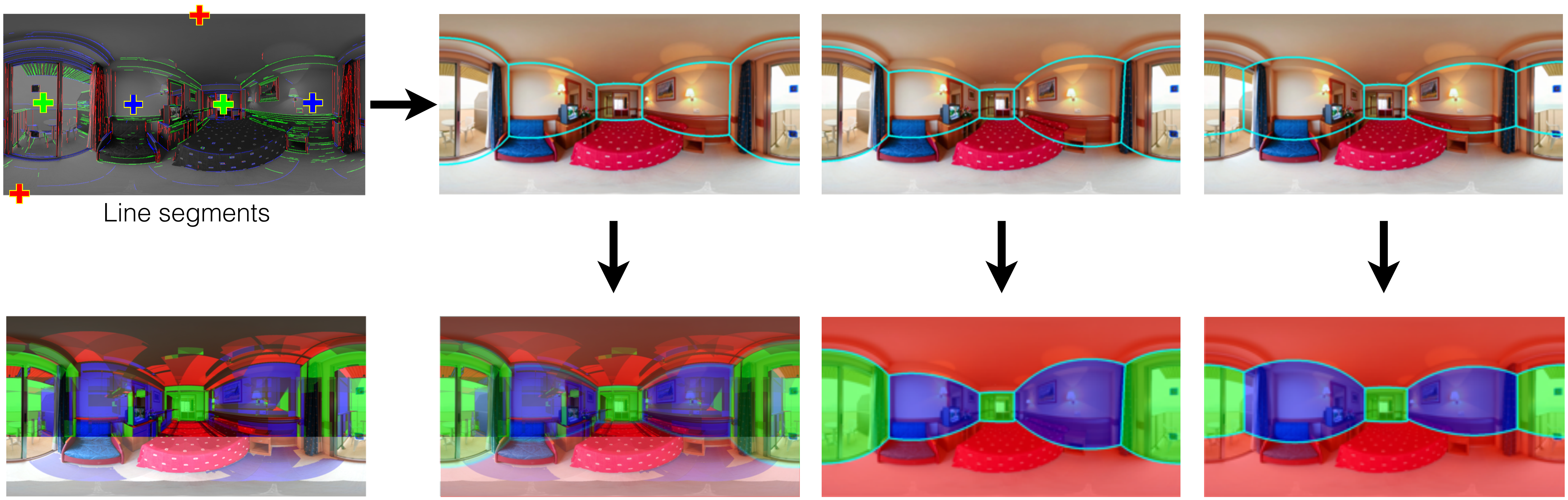




Surface normal estimation

Slide credit: Zhang et al.

Room layout hypothesis

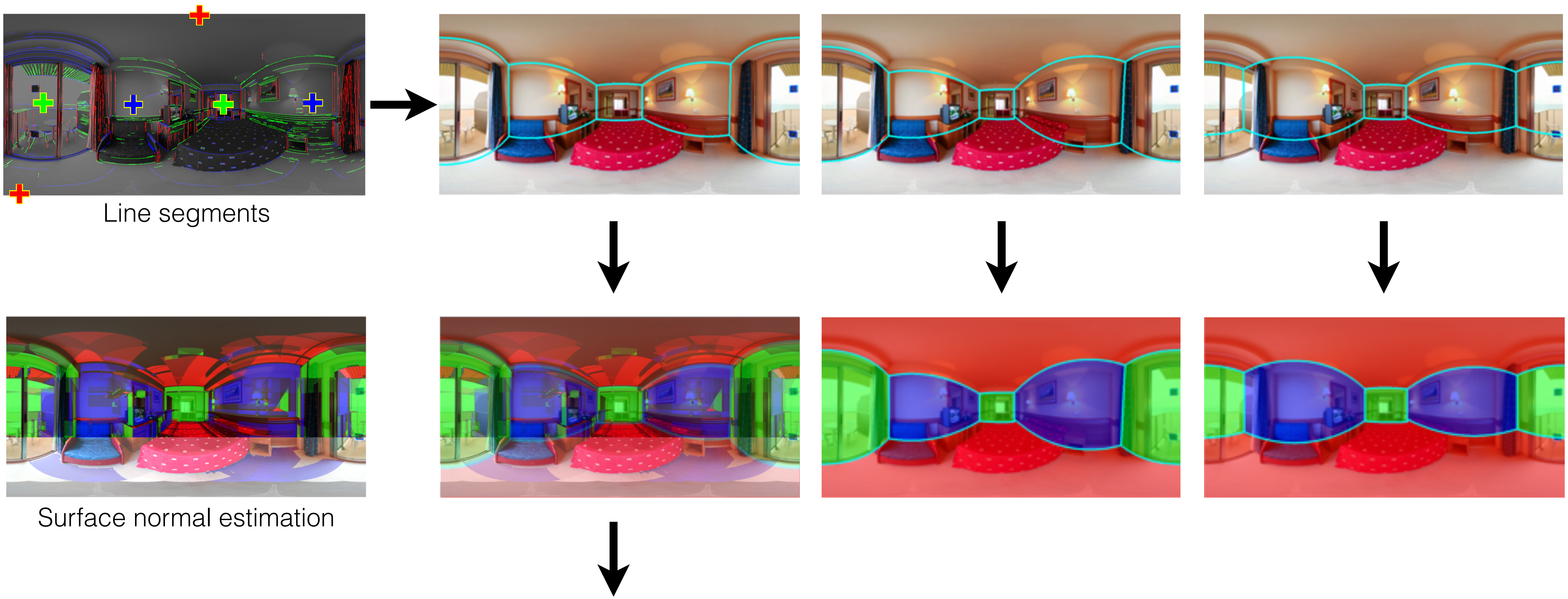


Line segments

Surface normal estimation

Slide credit: Zhang et al.

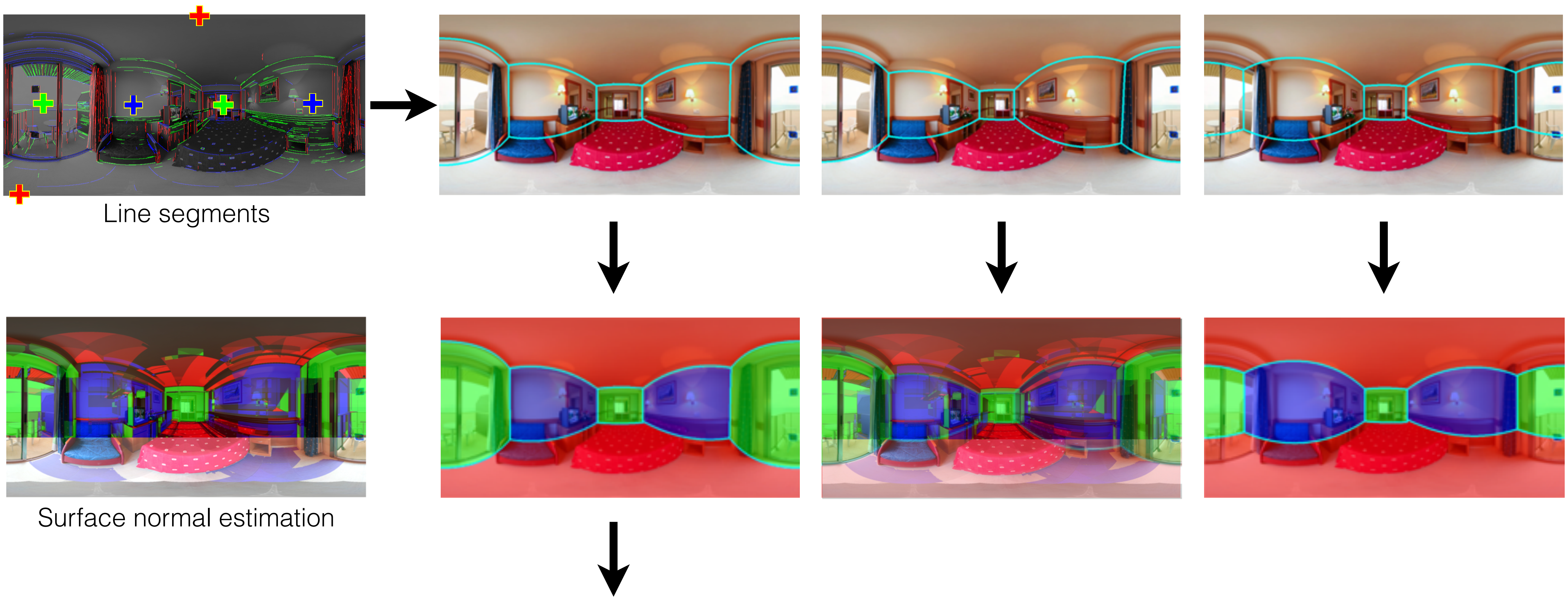
Room layout hypothesis



Consistency Score: **0.770**

Slide credit: Zhang et al.

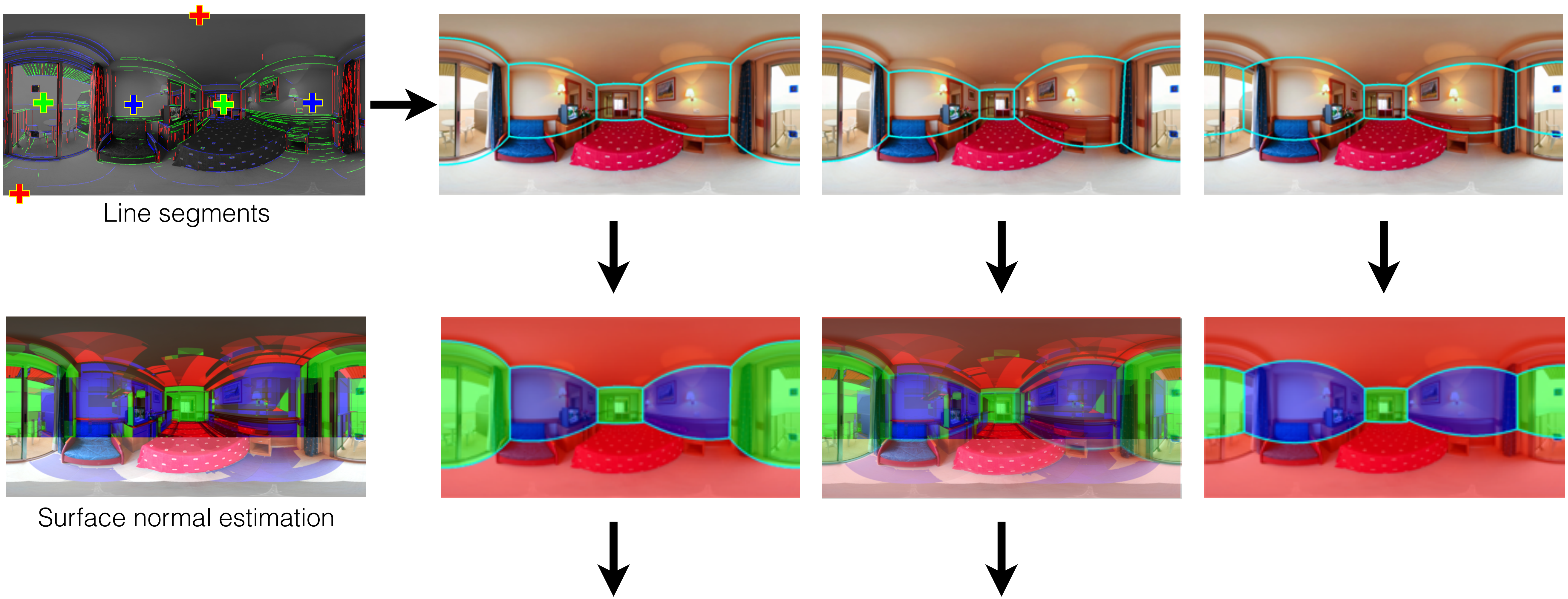
Room layout hypothesis



Consistency Score: **0.770**

Slide credit: Zhang et al.

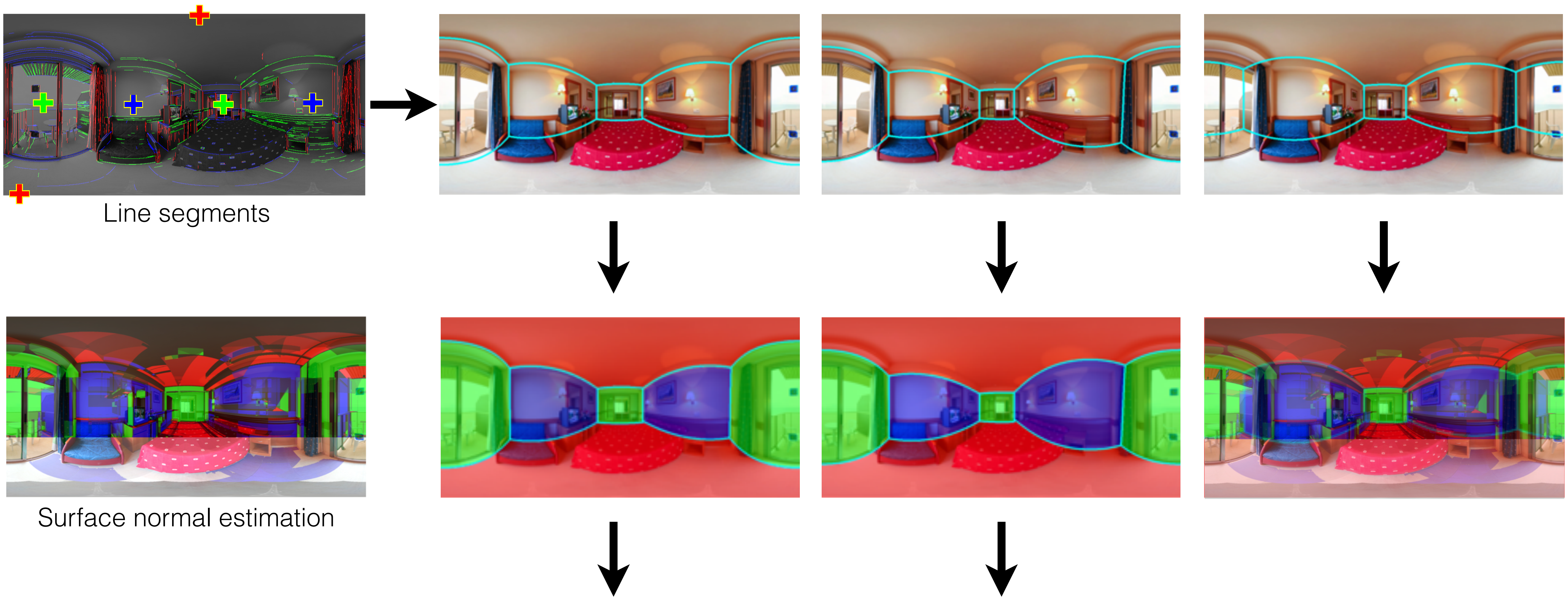
Room layout hypothesis



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

Slide credit: Zhang et al.

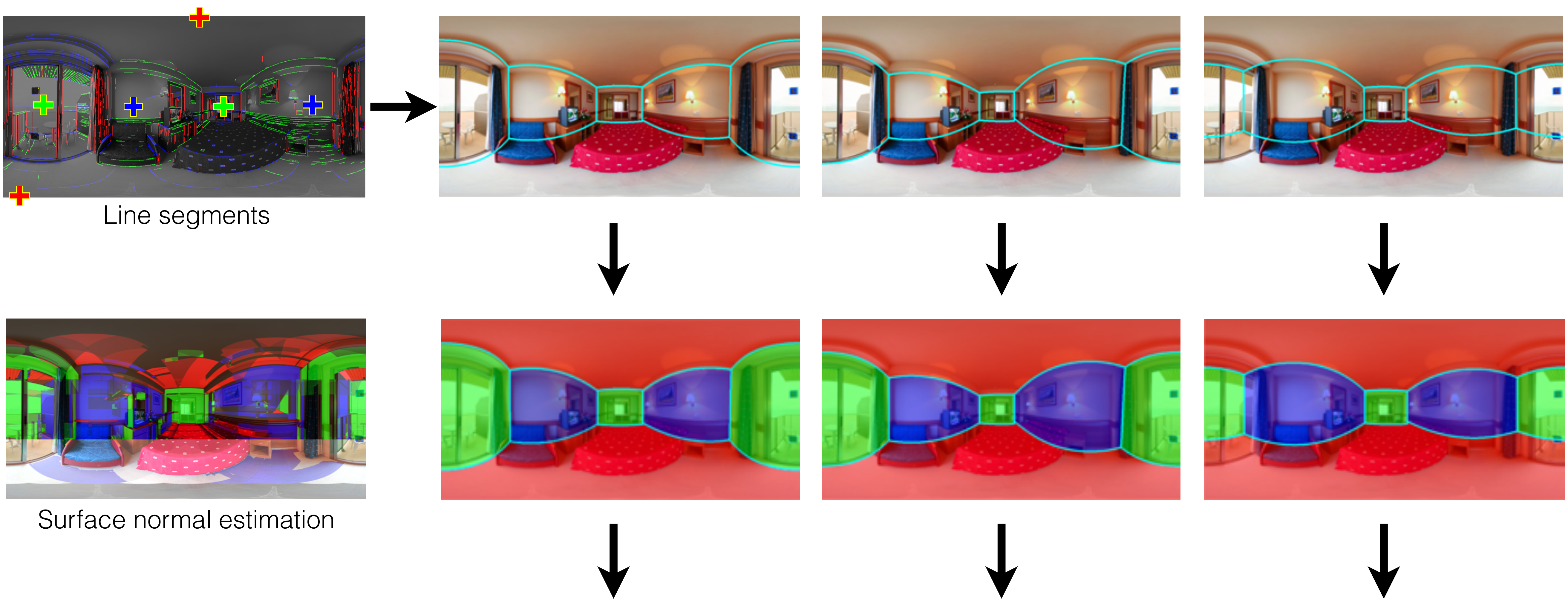
Room layout hypothesis



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

Slide credit: Zhang et al.

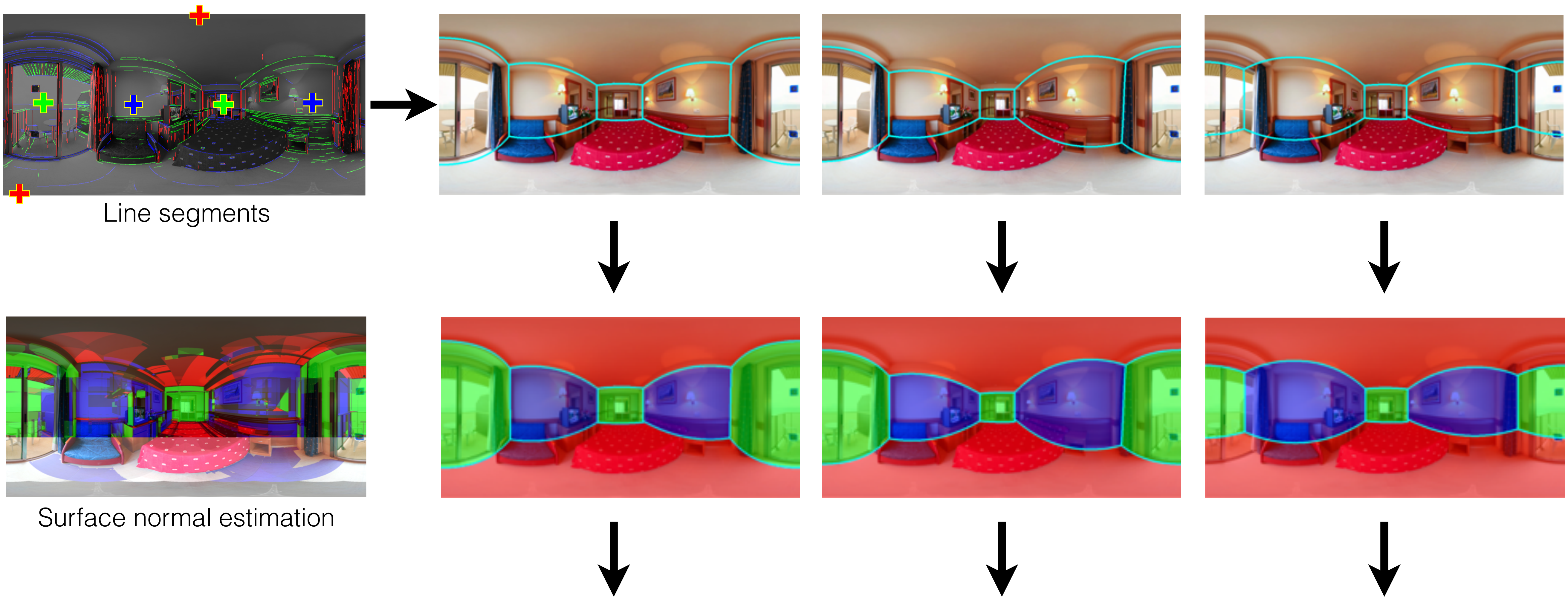
Room layout hypothesis



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

Slide credit: Zhang et al.

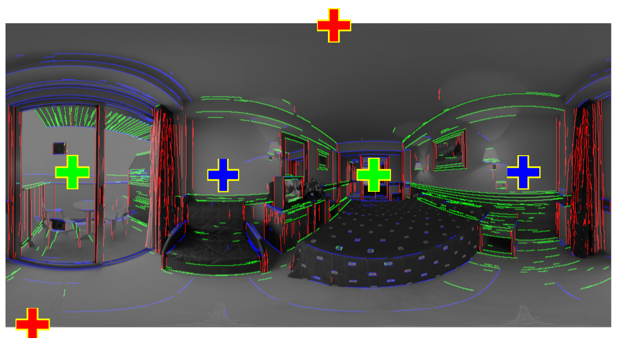
Room layout hypothesis



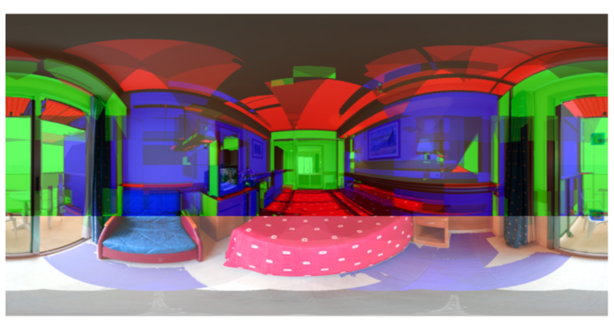
Consistency Score: **0.770** **0.711** **0.504**

Slide credit: Zhang et al.

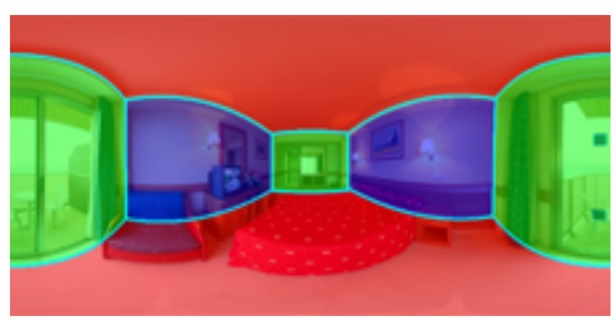
Room layout hypothesis



Line segments



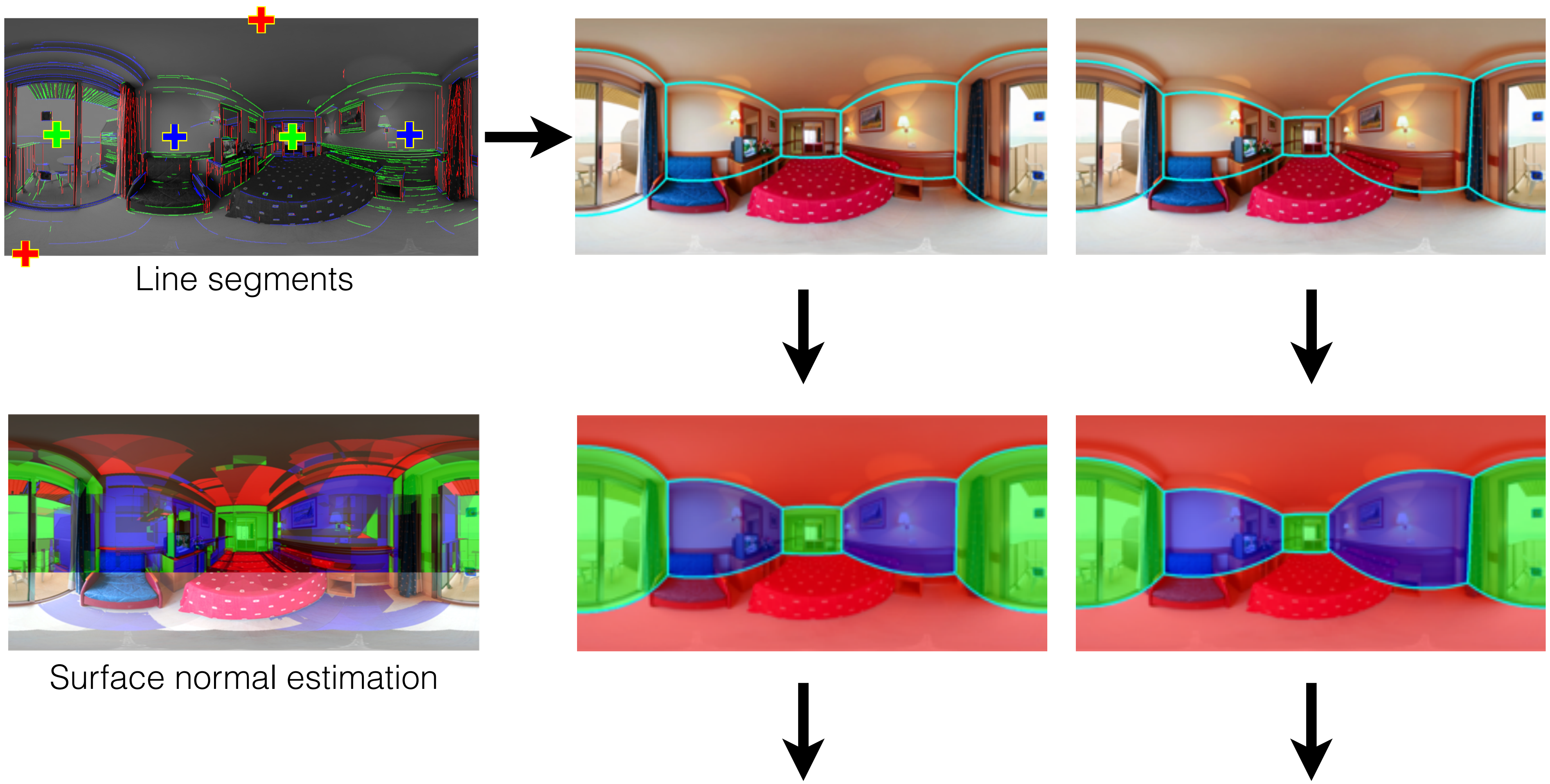
Surface normal estimation



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

Slide credit: Zhang et al.

Room layout hypothesis

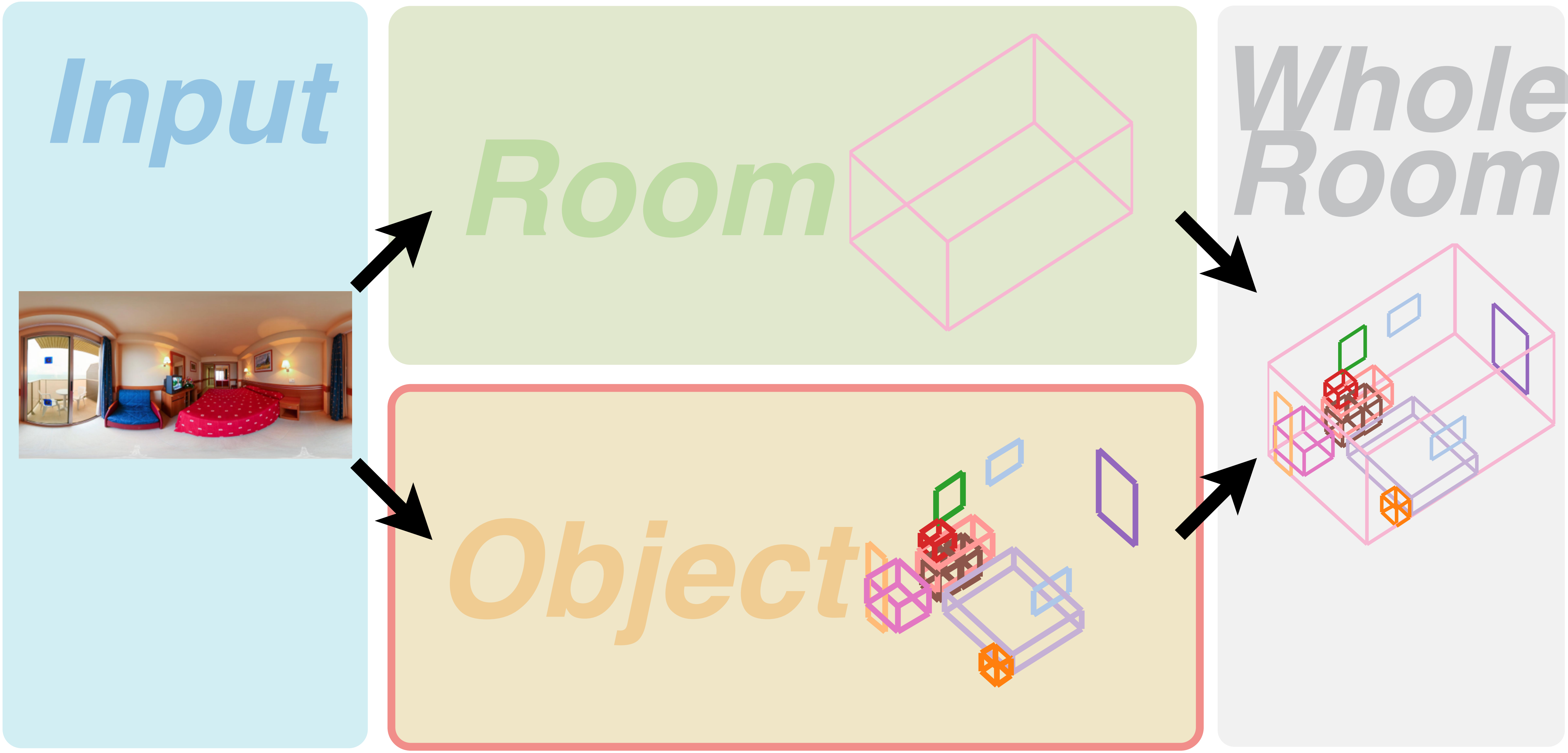


Top 50 only

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

Slide credit: Zhang et al.

Generate a pool of hypotheses

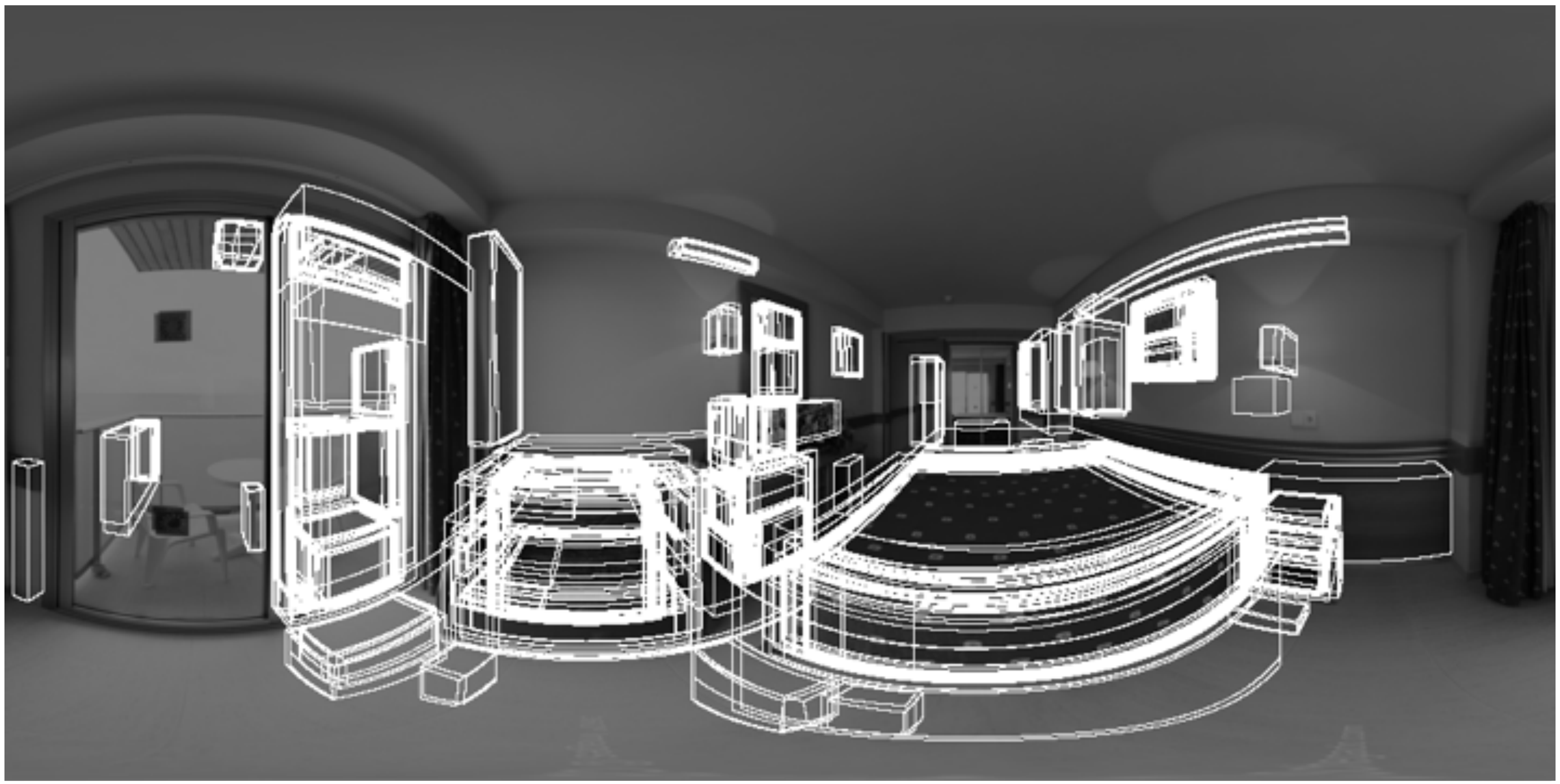


Slide credit: Zhang et al.

Cuboid detection

Slide credit: Zhang et al.

Cuboid detection

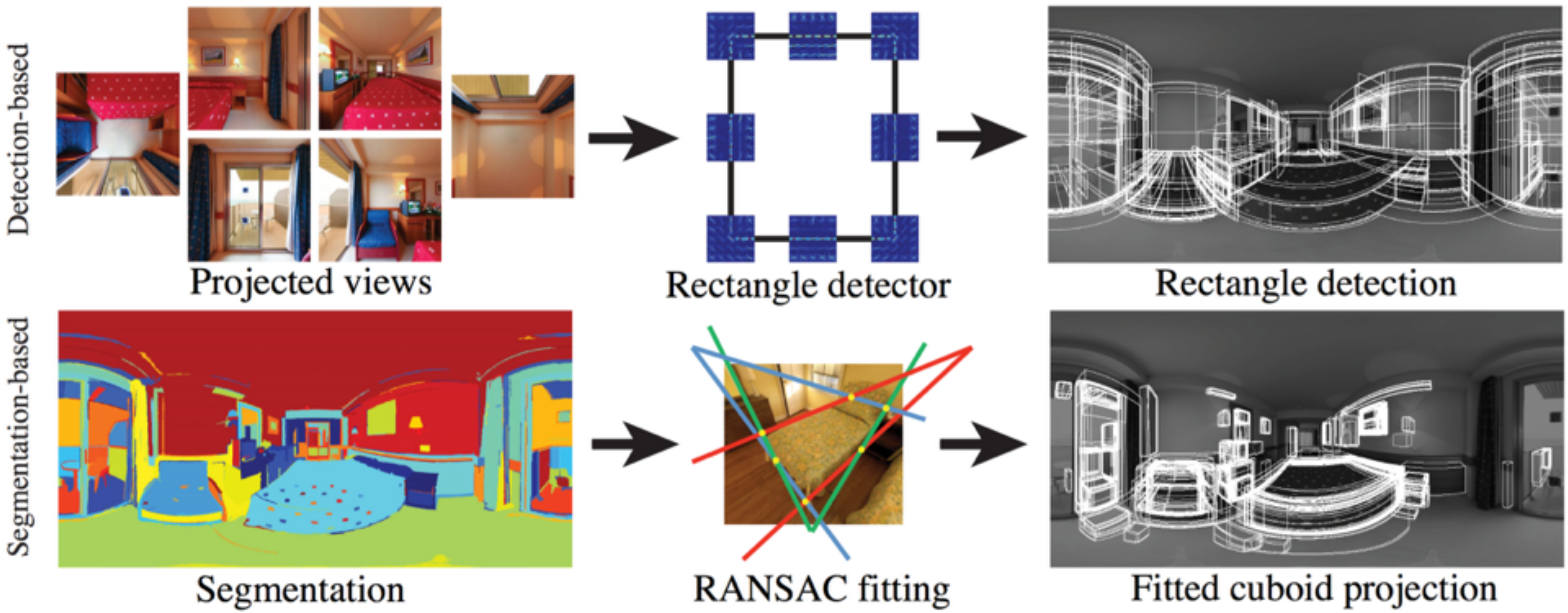


Fitted cuboids

Slide credit: Zhang et al.

Cuboid detection

DPM-esque



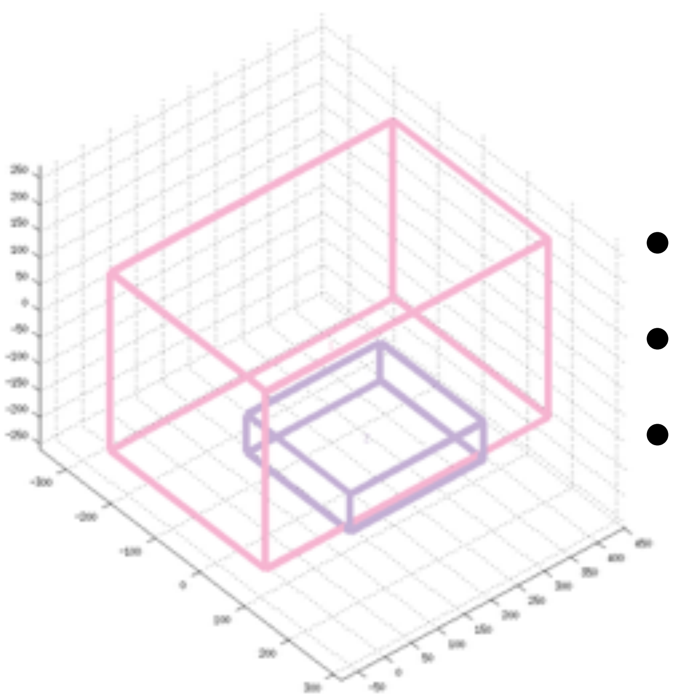
Selective search

6 rays and
3 vanishing points

Largest IoU
with the segment

Slide credit: Zhang et al.

Semantic classification



Features

- Size
- Aspect ratio & Area
- Distance to walls

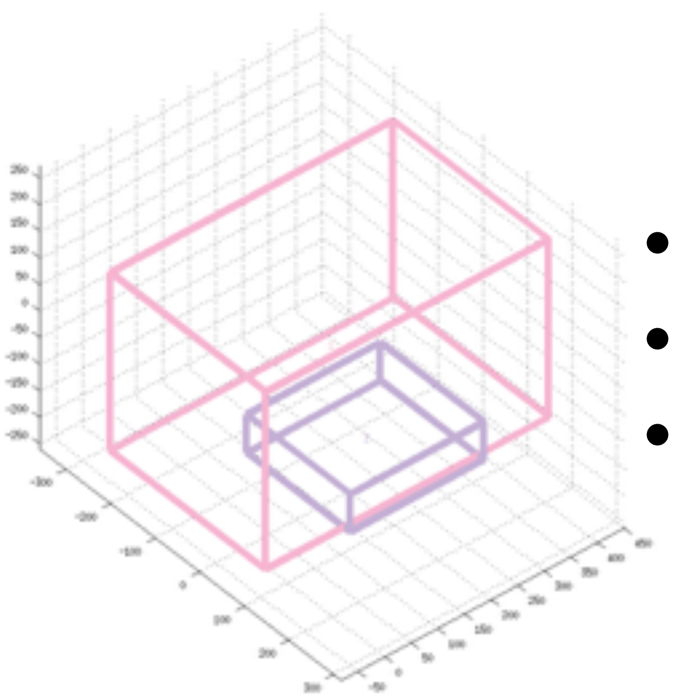
Random forest

Object categories

bed
desk
sofa
…
chair

Slide credit: Zhang et al.

Semantic classification



Features

- Size
- Aspect ratio & Area
- Distance to walls

Random forest

Object categories

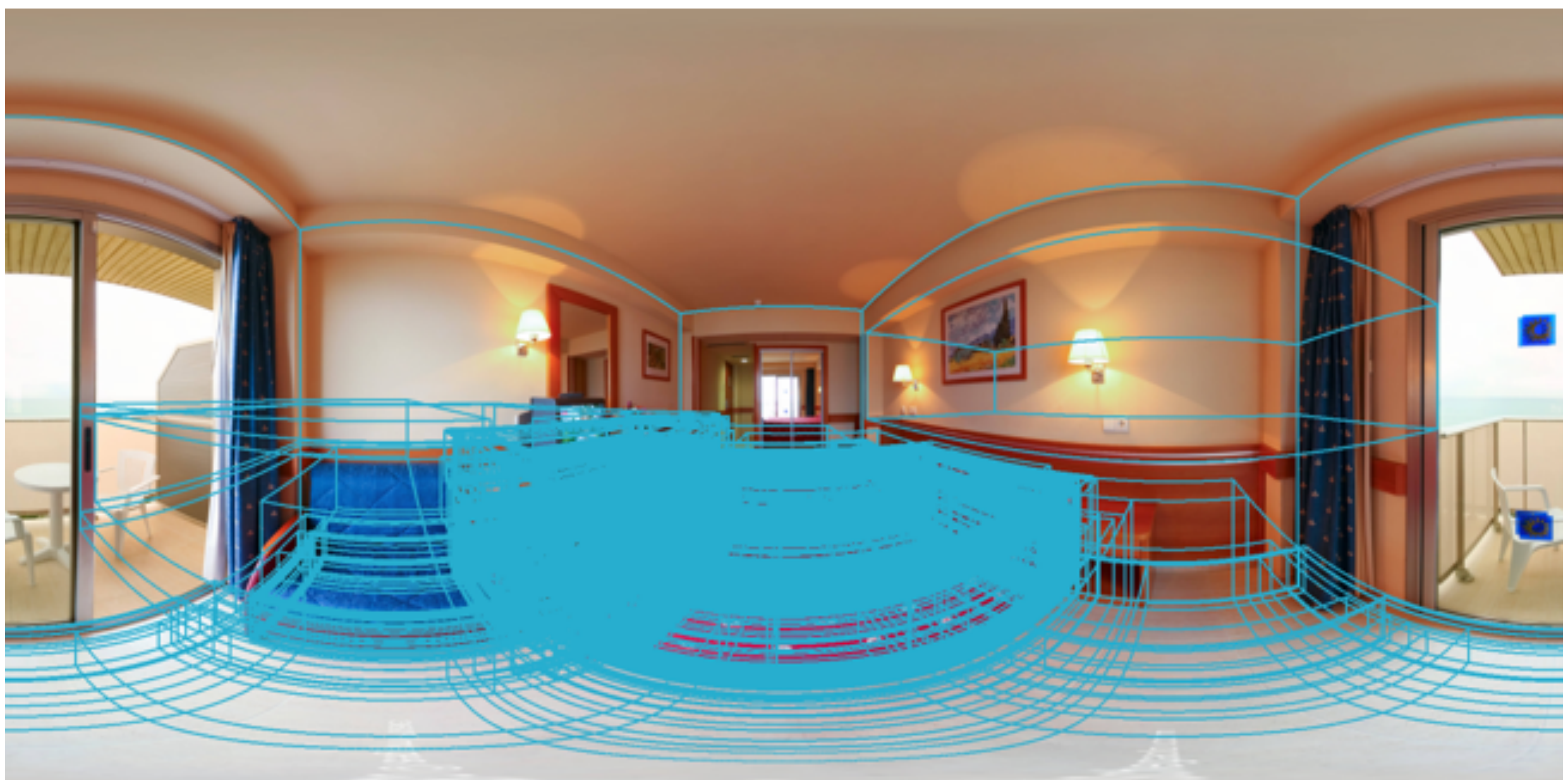
bed
desk
sofa
…
chair

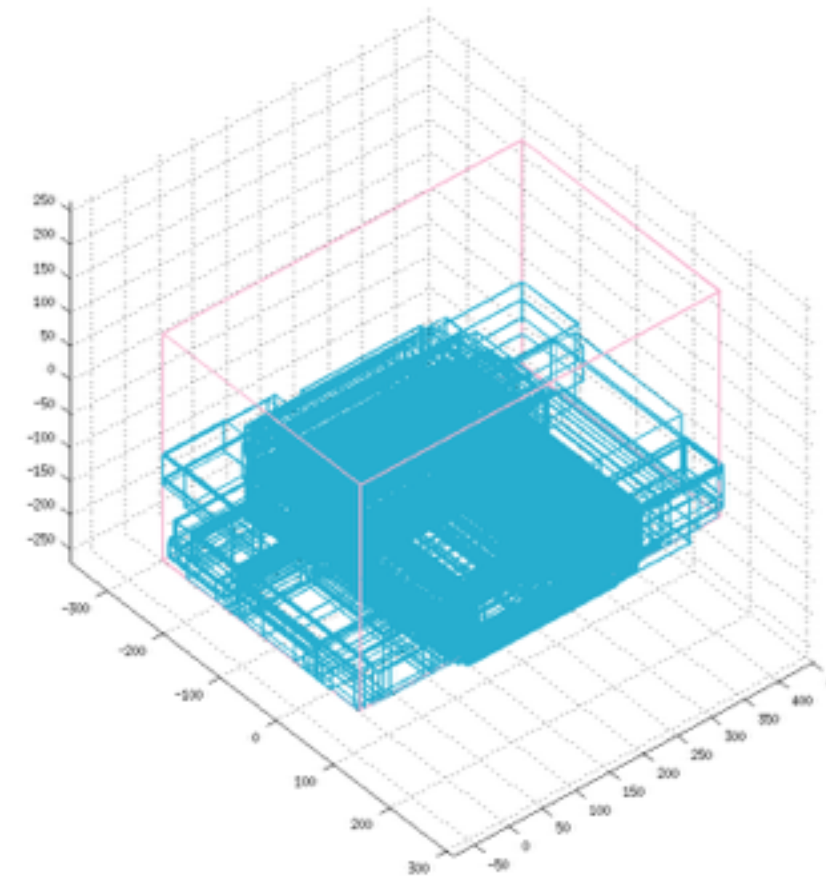
70% Accuracy

Slide credit: Zhang et al.

Semantic classification

bed

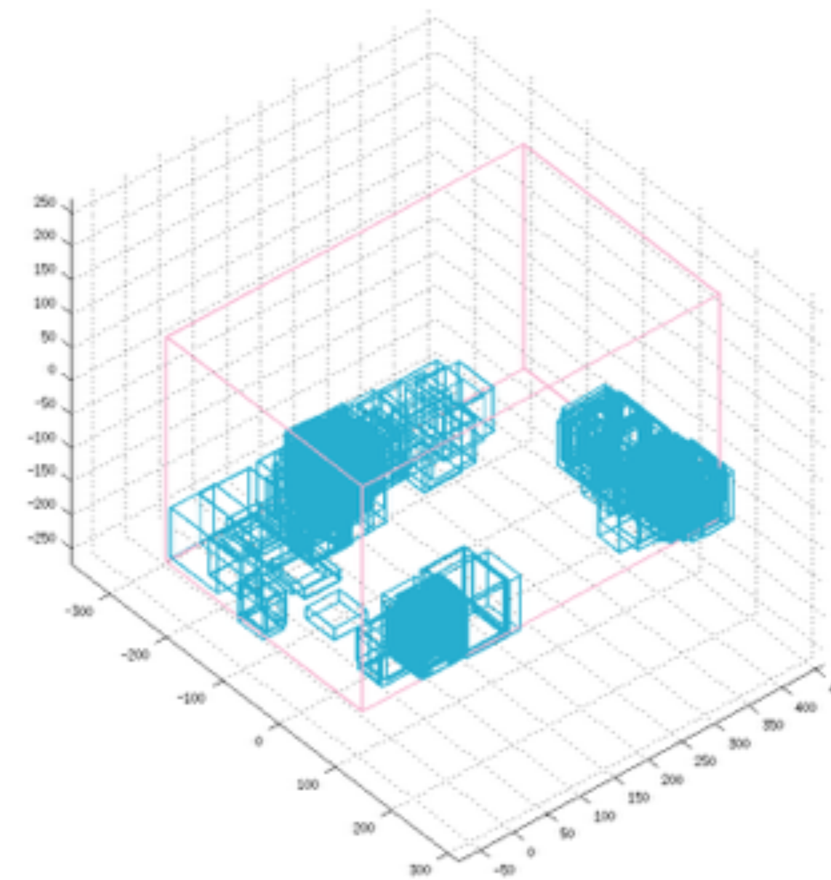




Slide credit: Zhang et al.

Semantic classification

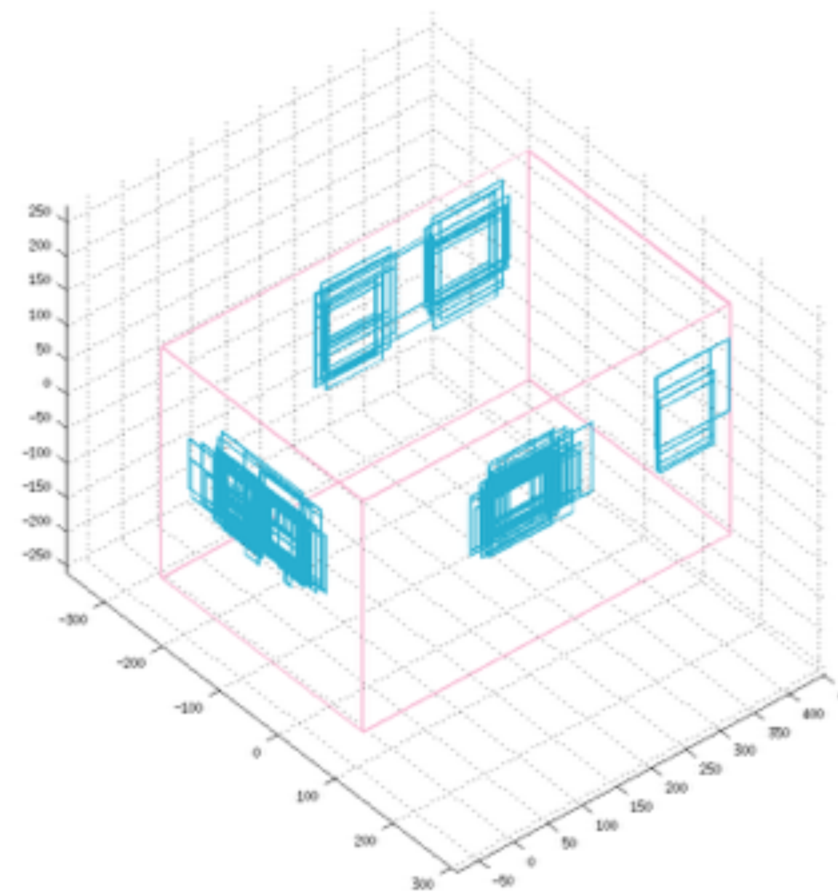
nightstand



Slide credit: Zhang et al.

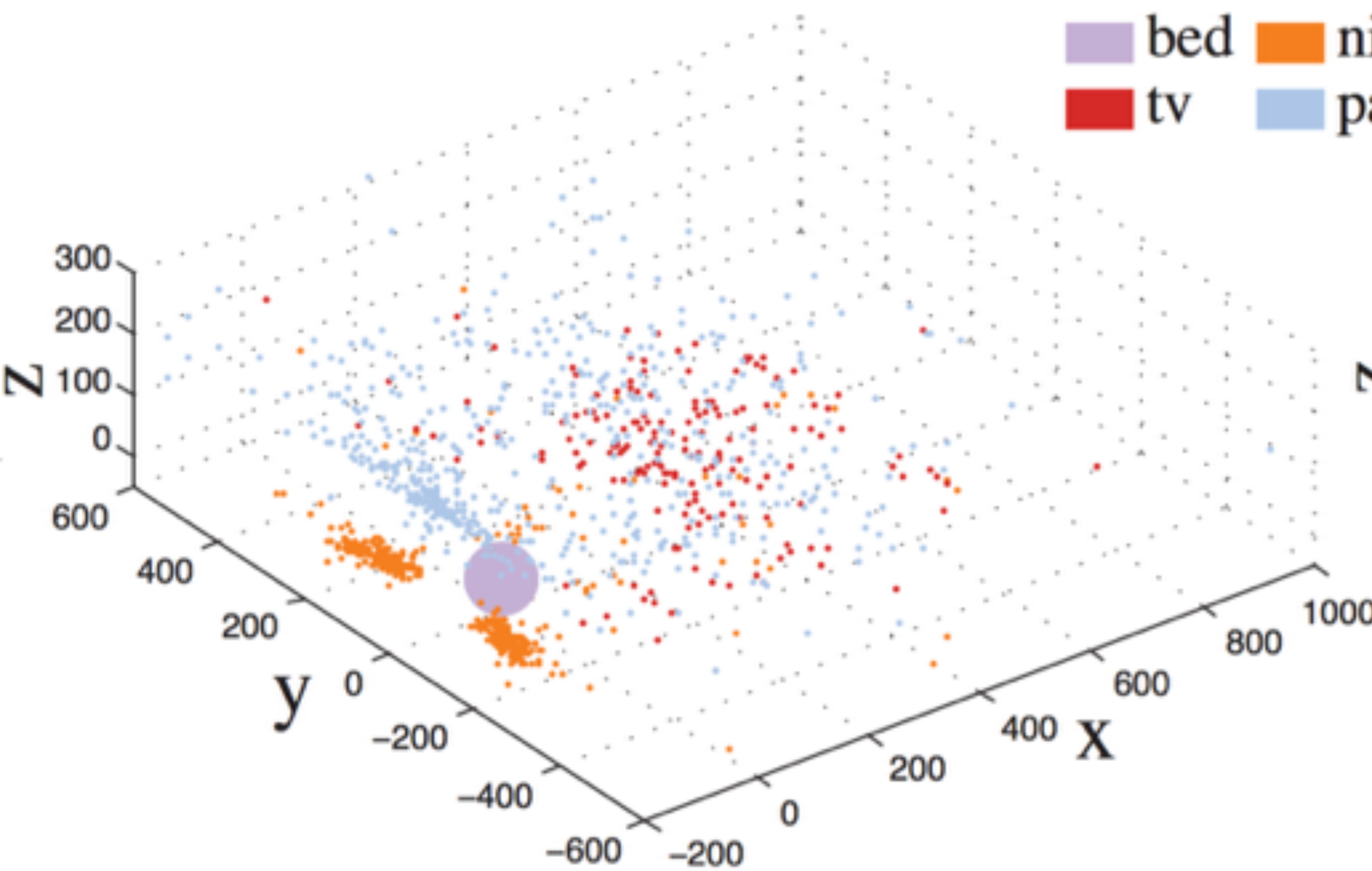
Semantic classification

painting

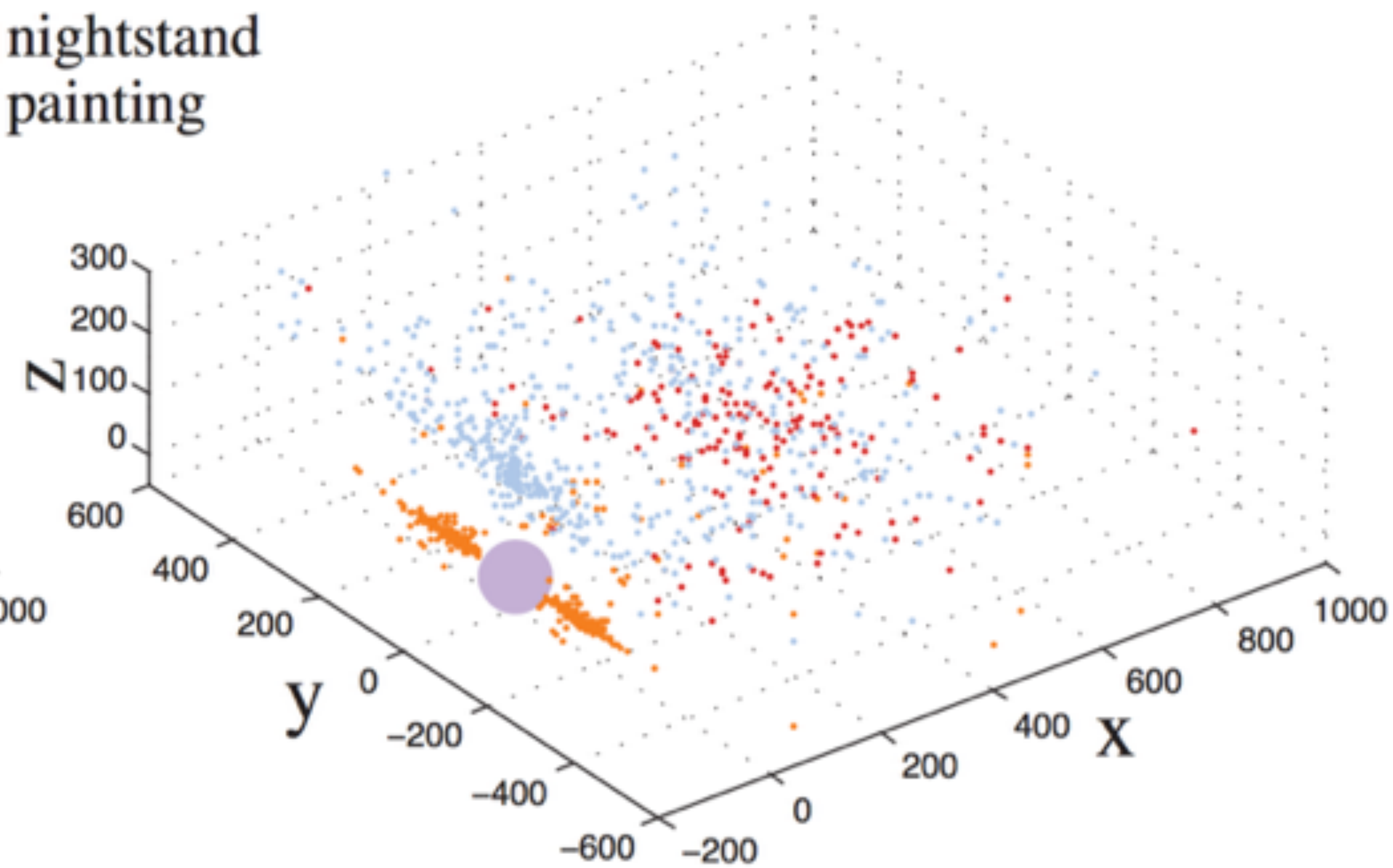


Slide credit: Zhang et al.

Pairwise constraint



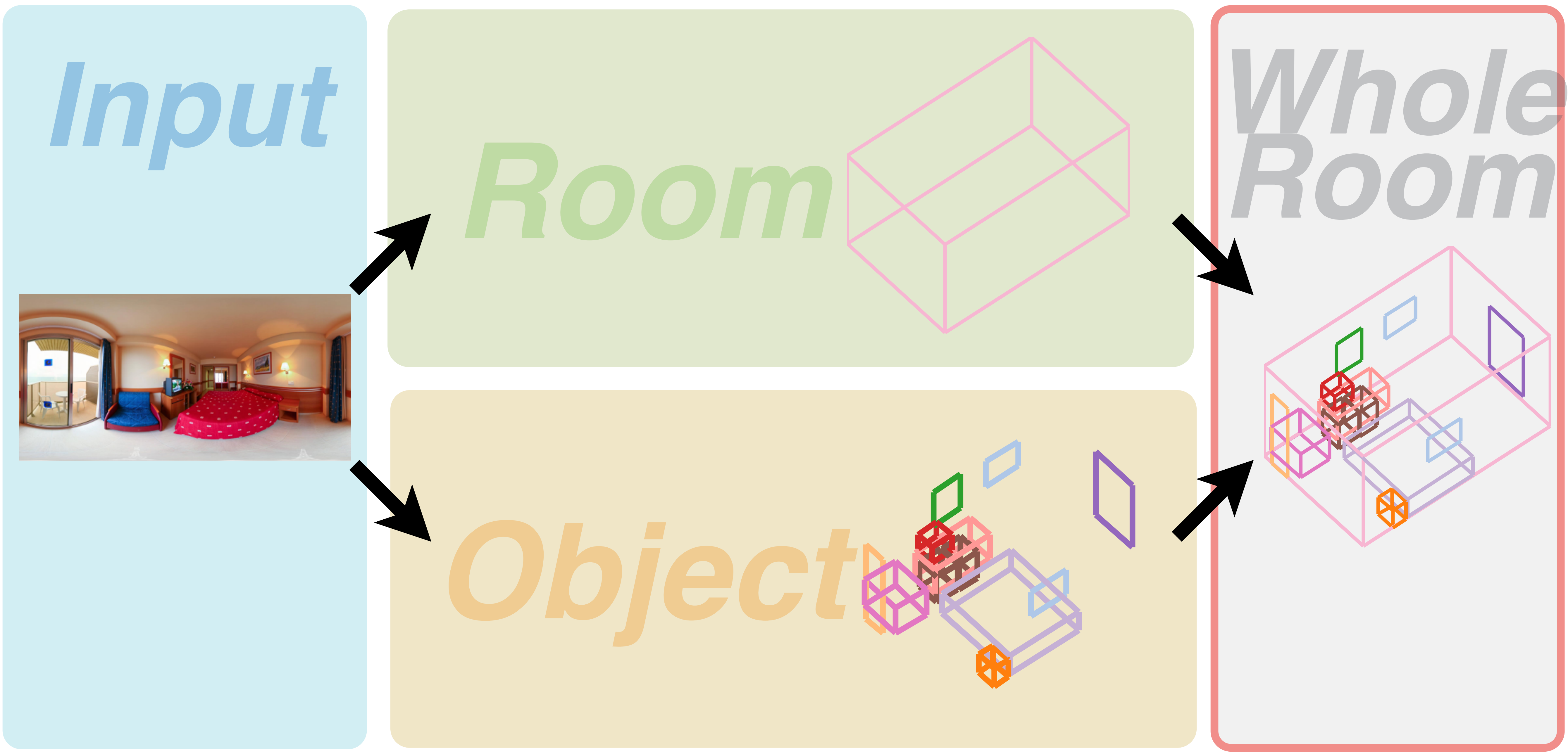
Object centroids



Face centroid with the closest wall

Slide credit: Zhang et al.

Generate a pool of hypotheses



Slide credit: Zhang et al.

Data-driven sampling

Randomly sample
a room layout

With P(layout) $\propto$ normal consistency score

Slide credit: Zhang et al.

Data-driven sampling

Randomly sample
a room layout

With P(layout) $\propto$ normal consistency score

Slide credit: Zhang et al.

Data-driven sampling

Slide credit: Zhang et al.

Data-driven sampling

Decide number of object
based on prior distribution:

paintin	2
bed	1
desk	1
nightst	1
mirror	1
sofa	1
tv	1
window	1

Slide credit: Zhang et al.

Data-driven sampling

Decide number of object based on prior distribution:

paintin	2
bed	1
desk	1
nightst	1
mirror	1
sofa	1
tv	1
window	1

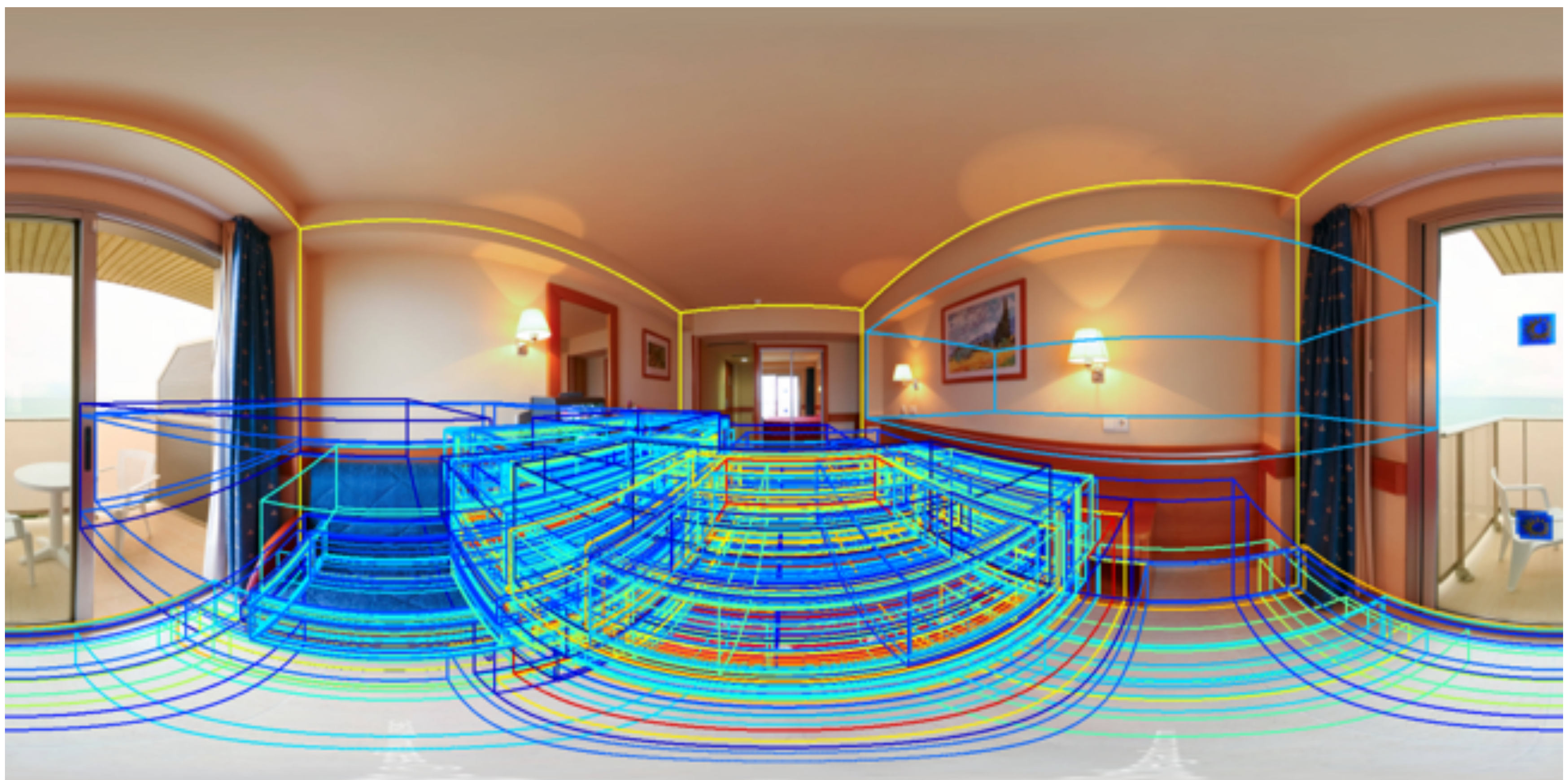
Decide object sampling sequence based on bottom up scores:

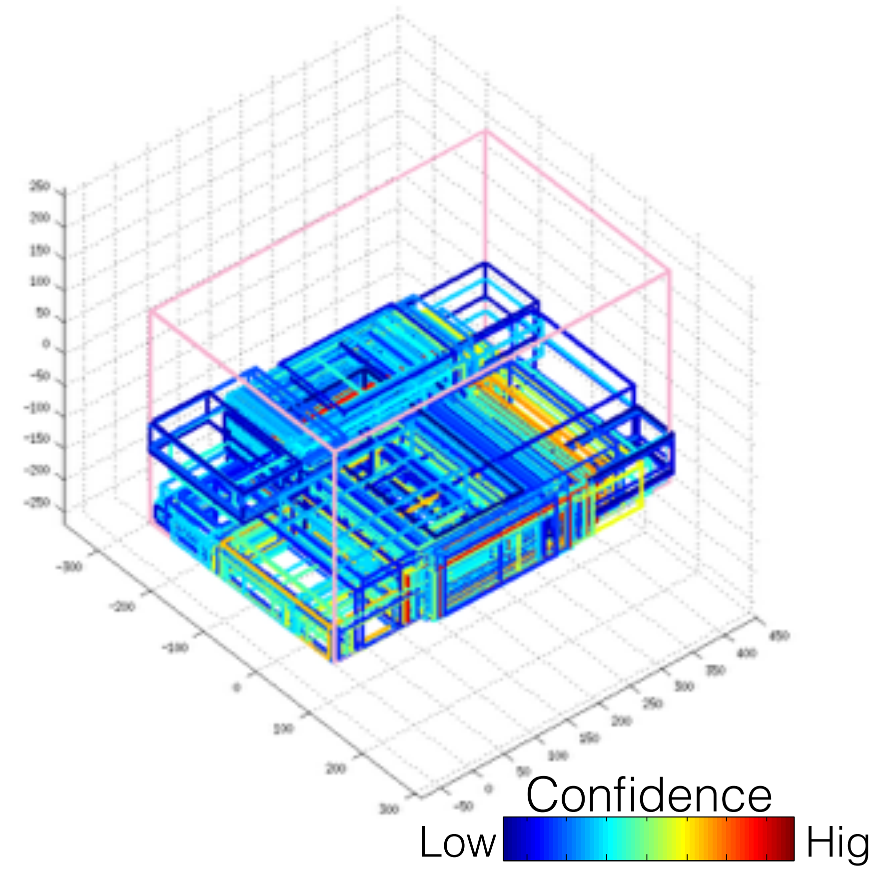
bed
nightstand
painting
desk
window
painting
tv
sofa
mirror

Slide credit: Zhang et al.

Data-driven sampling

Sample a bed in empty room first…



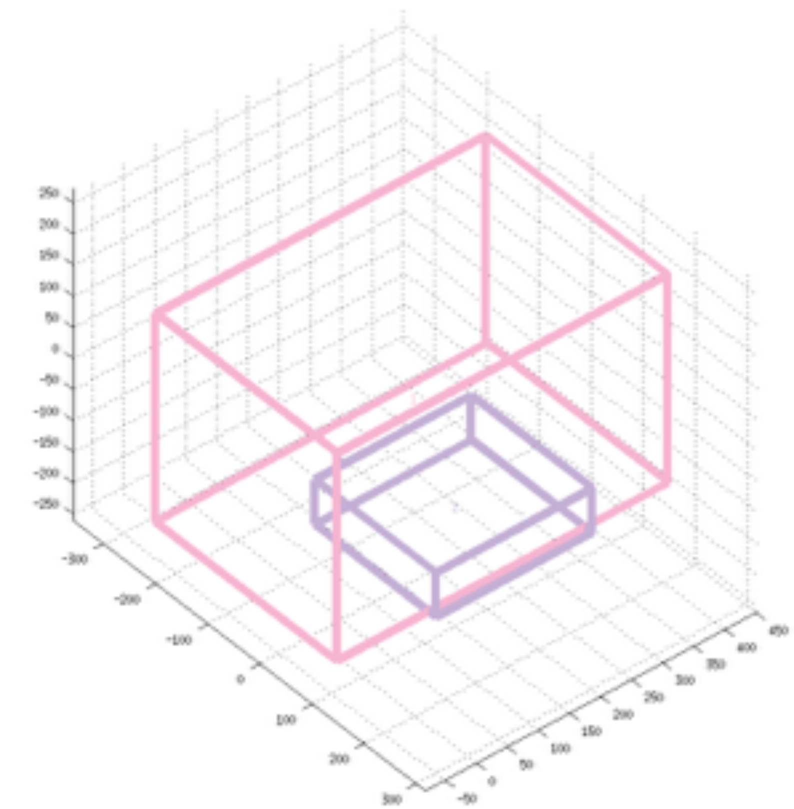


Bottom-up score as bed

Slide credit: Zhang et al.

Data-driven sampling

Sample a bed in empty room first…



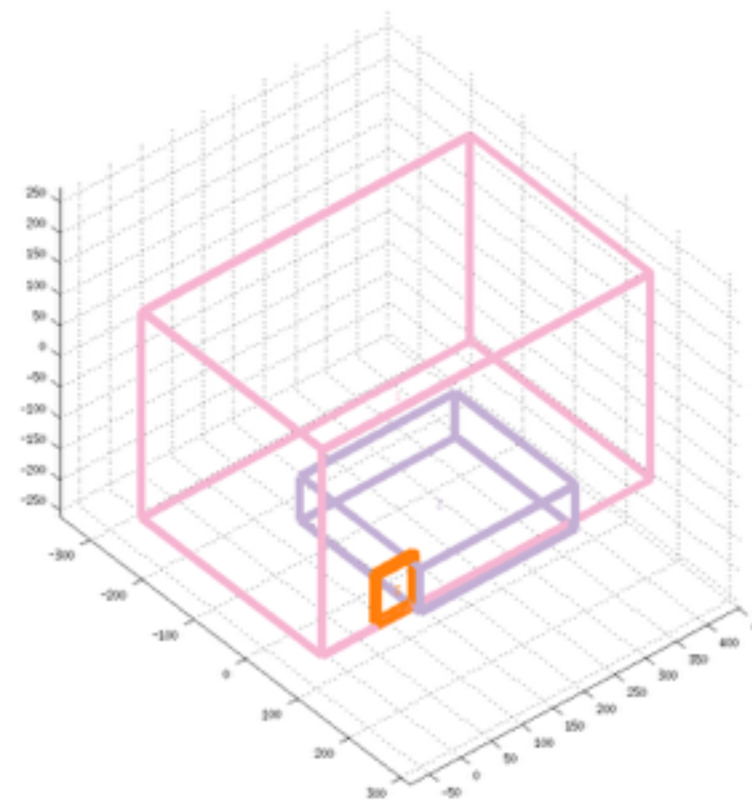
Randomly select one according to **bottom up** priority

rectangle detection score, semantic classifier score

Slide credit: Zhang et al.

Data-driven sampling

Then, sample a nightstand given a bed

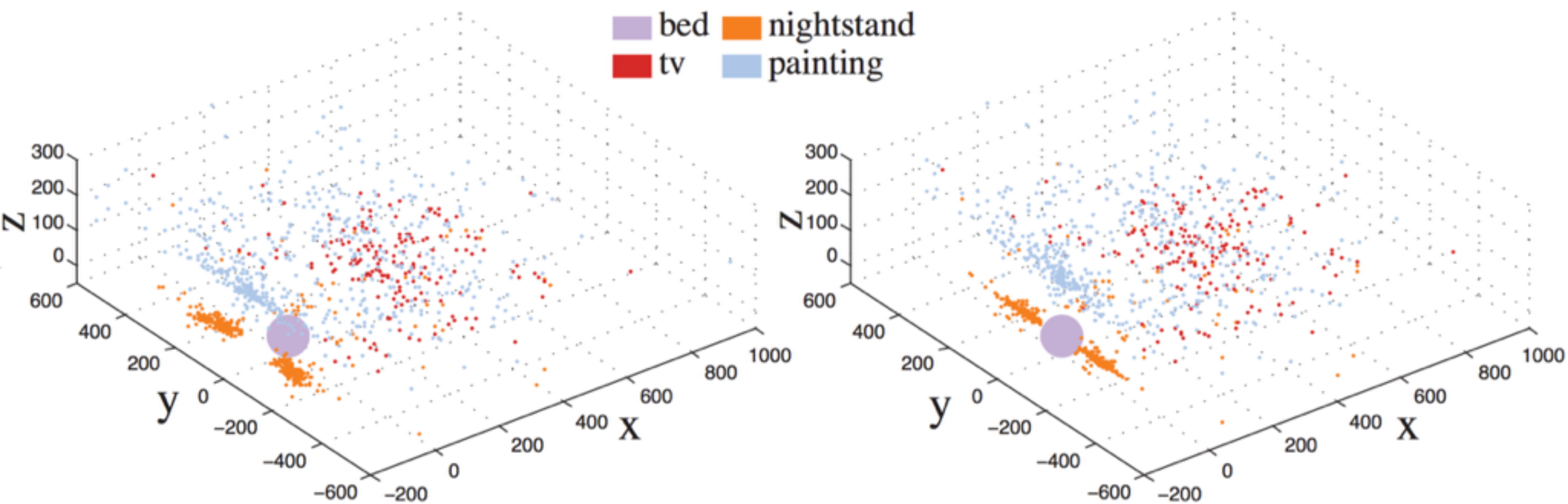


Randomly select one according to the bottom up **+ pair-wise** priority

mean distance to the K nearest neighbors

Slide credit: Zhang et al.

Pairwise constraint



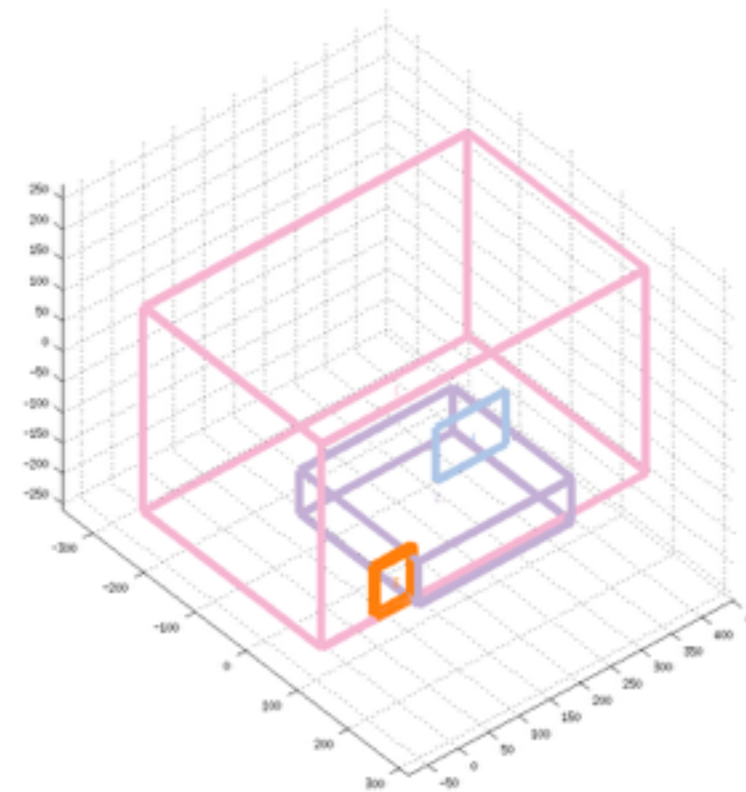
Object centroids

Face centroid with the closest wall

Slide credit: Zhang et al.

Data-driven sampling

Keep on sampling until finishing the list…

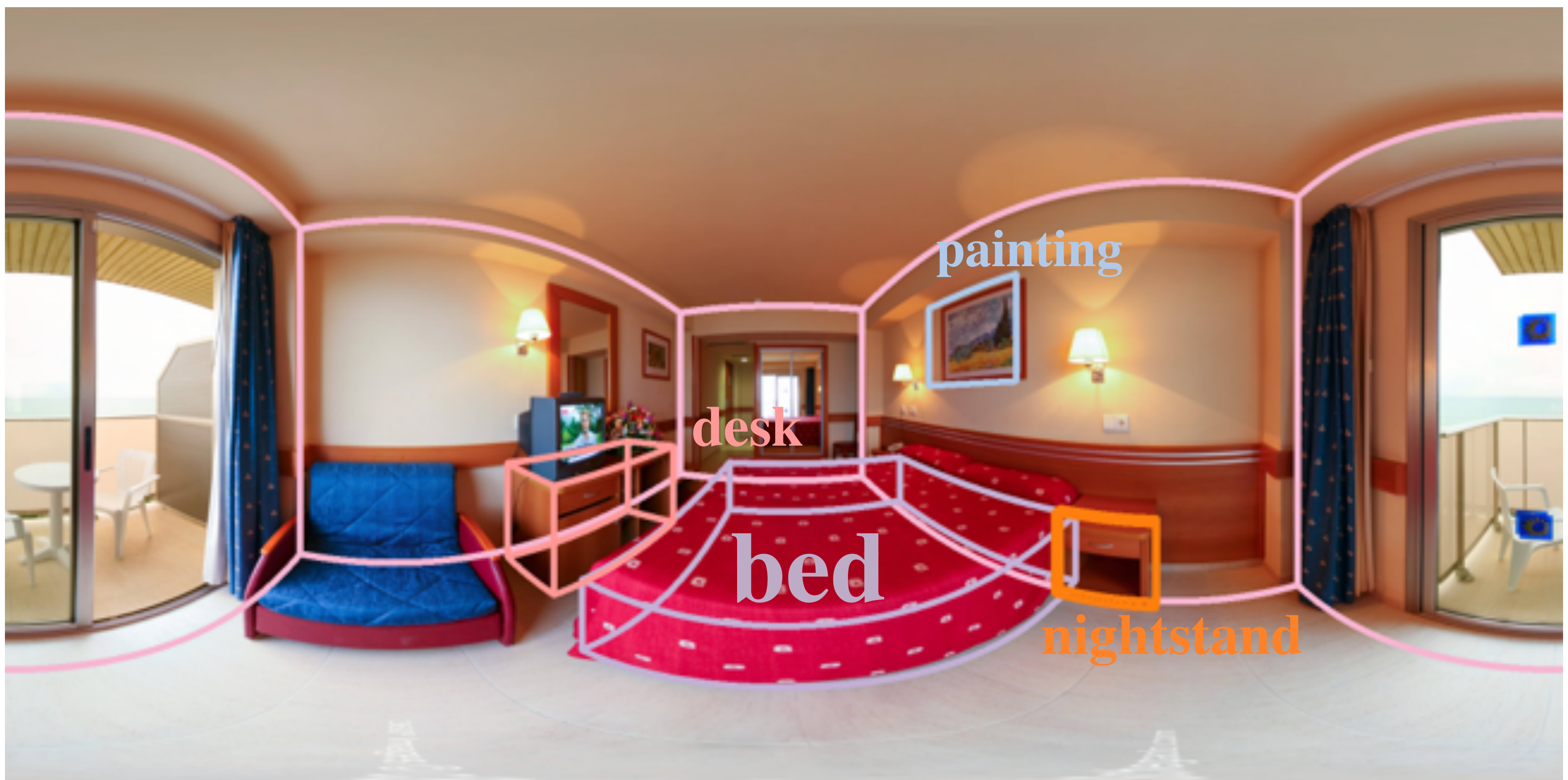


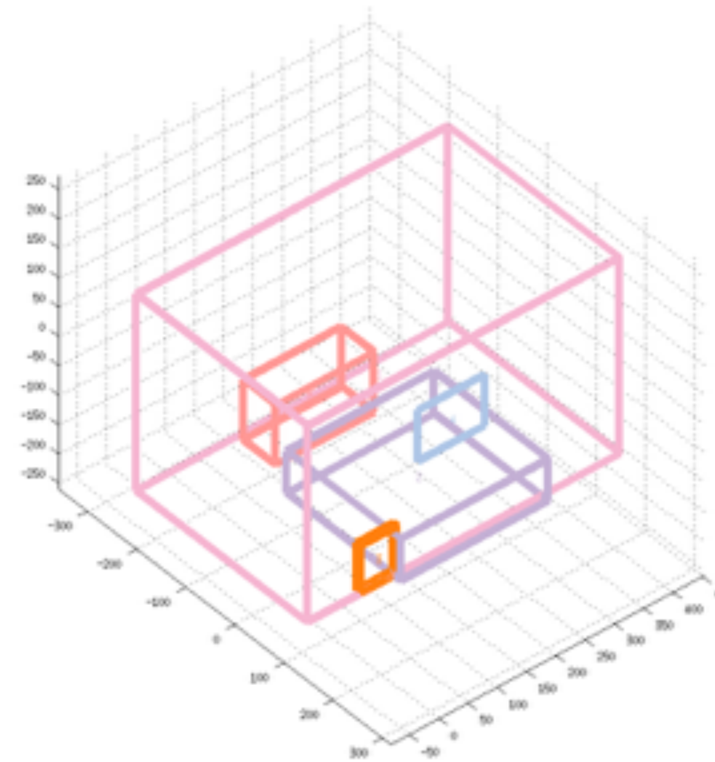
List: bed, nightstand, painting, desk, window, painting, TV, sofa, mirror

Slide credit: Zhang et al.

Data-driven sampling

Keep on sampling until finishing the list…



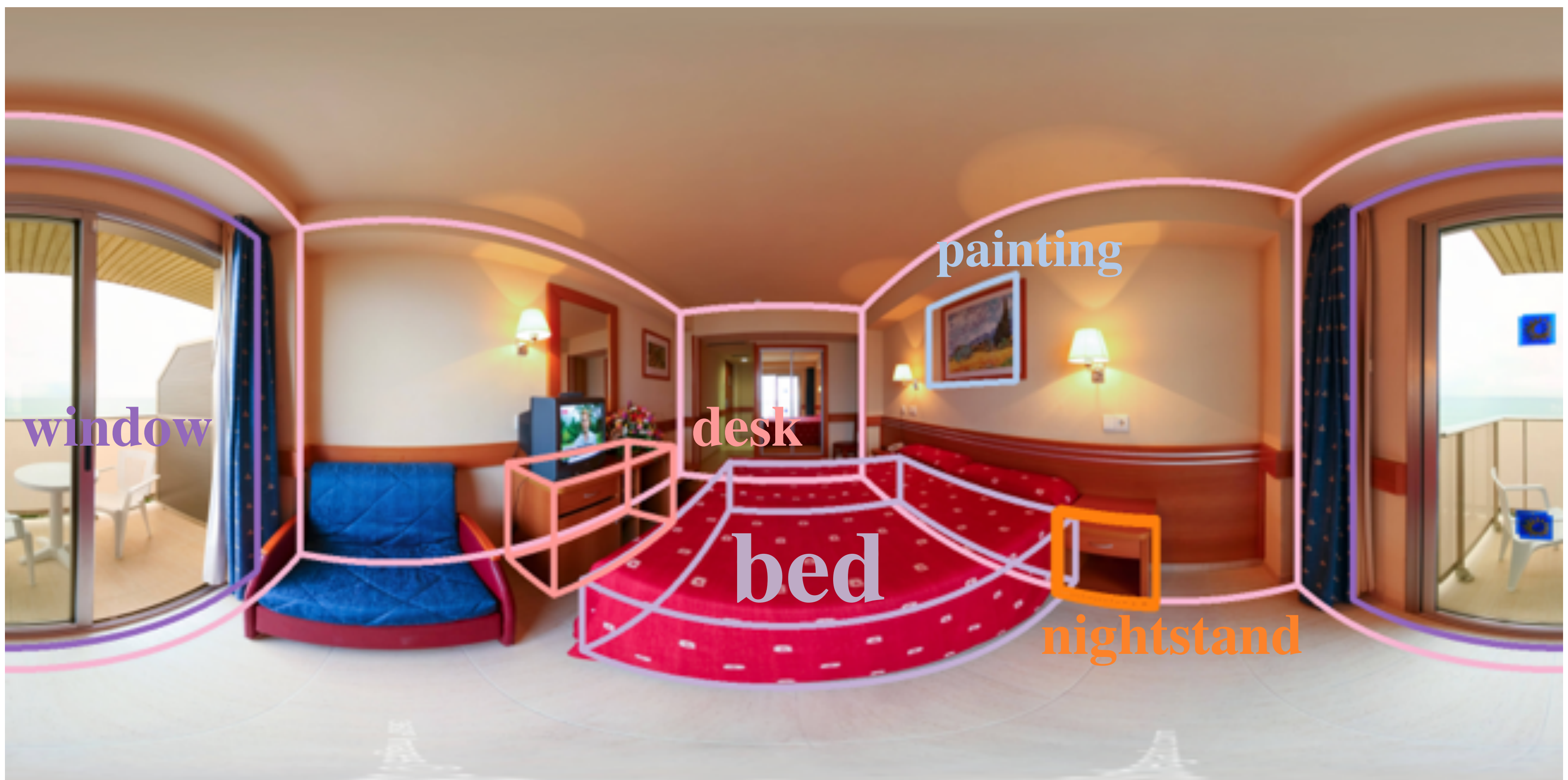


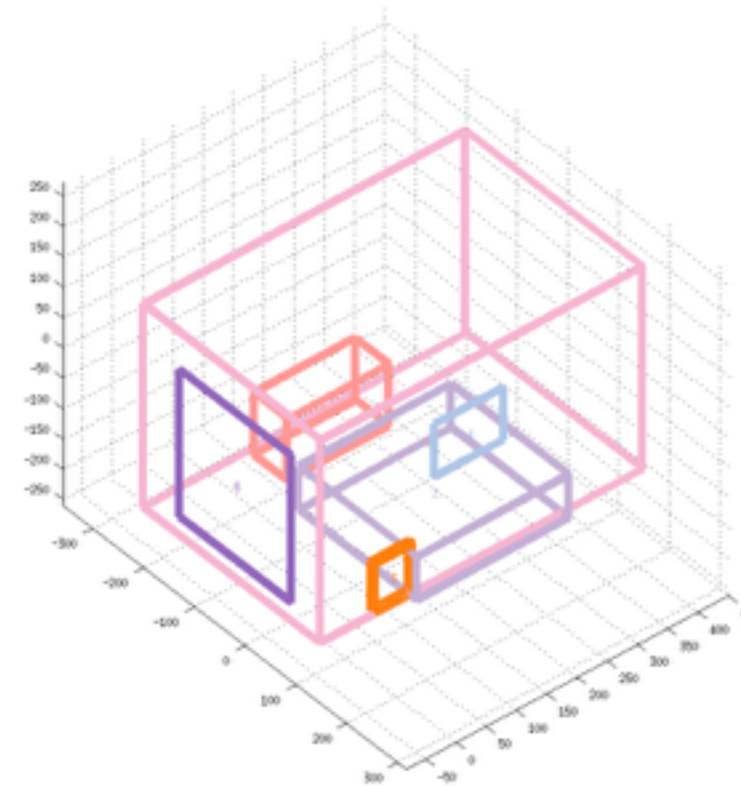
List: bed, nightstand, painting, desk, window, painting, TV, sofa, mirror

Slide credit: Zhang et al.

Data-driven sampling

Keep on sampling until finishing the list…



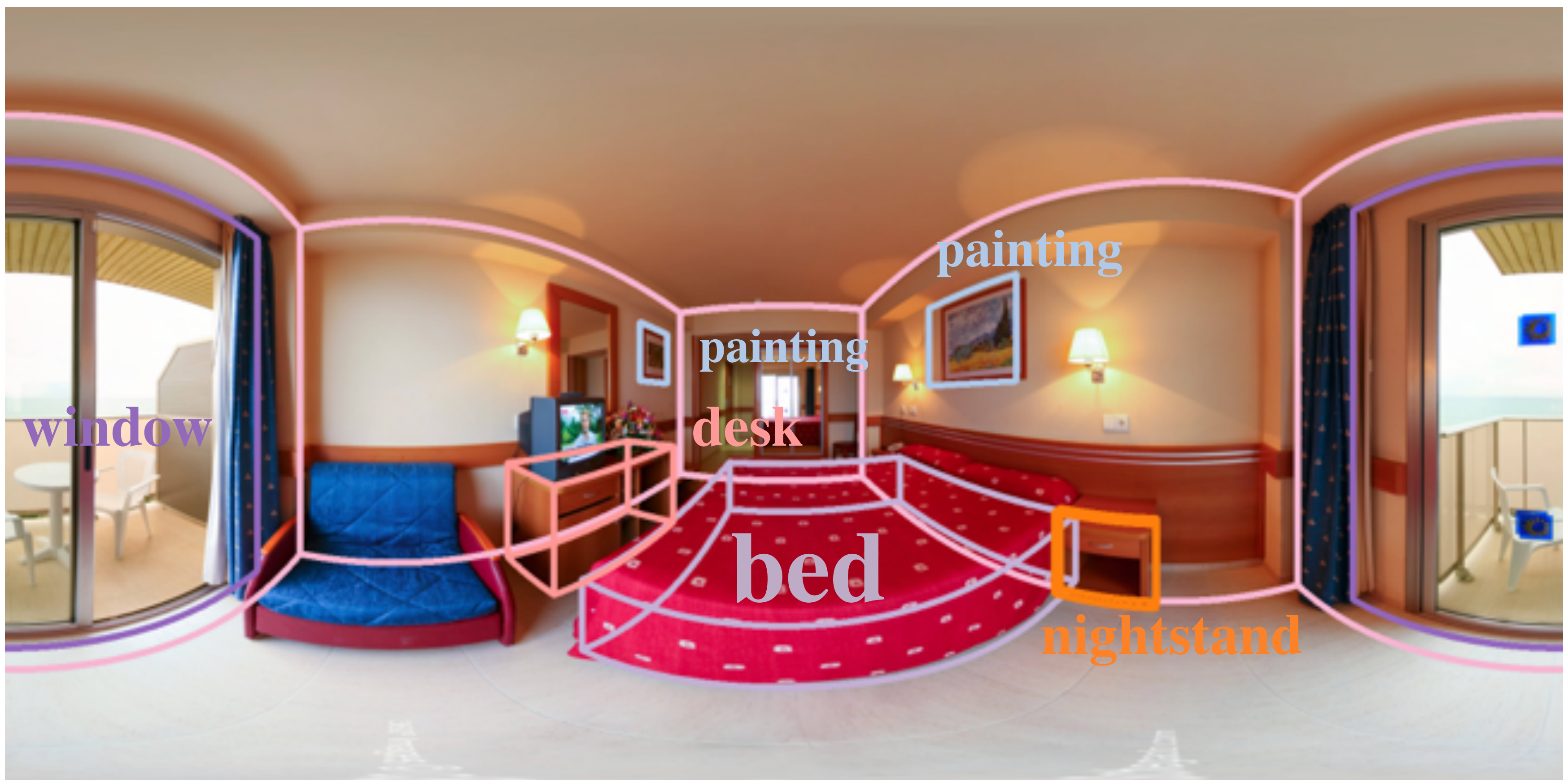


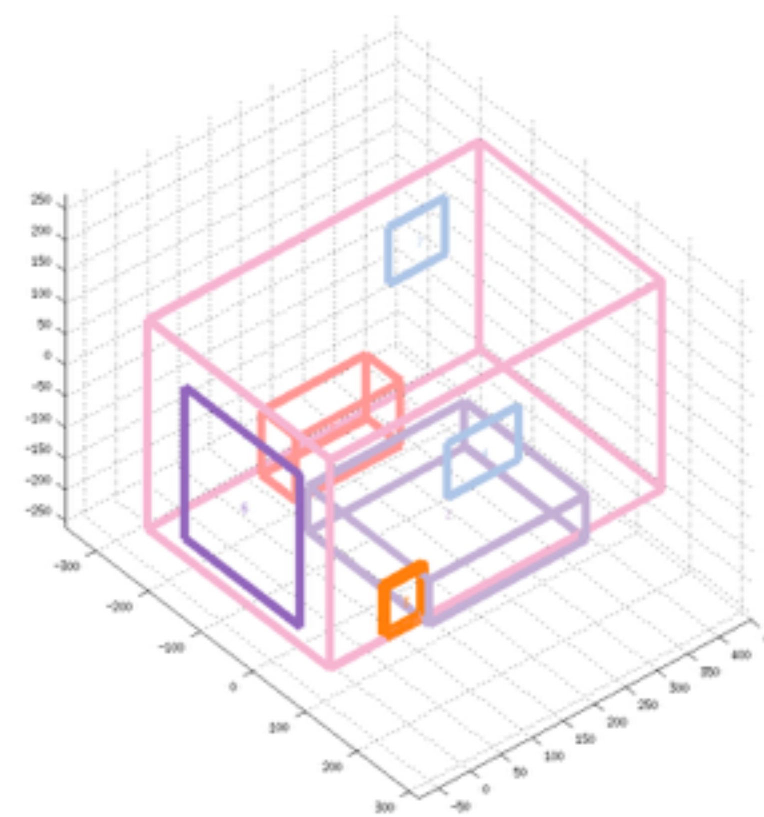
List: bed, nightstand, painting, desk, window, painting, TV, sofa, mirror

Slide credit: Zhang et al.

Data-driven sampling

Keep on sampling until finishing the list…



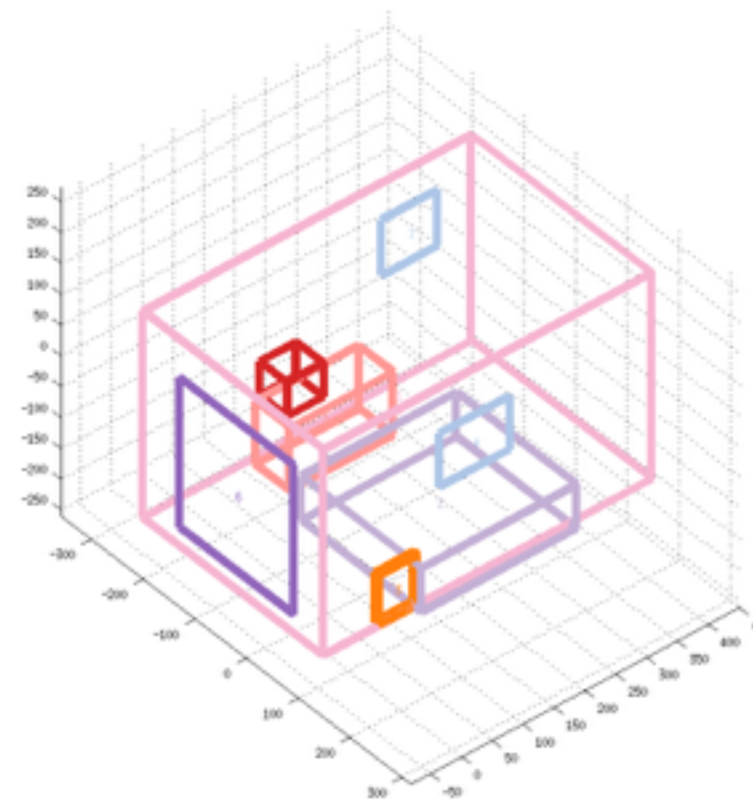


List: bed, nightstand, painting, desk, window, painting, TV, sofa, mirror

Slide credit: Zhang et al.

Data-driven sampling

Keep on sampling until finishing the list…

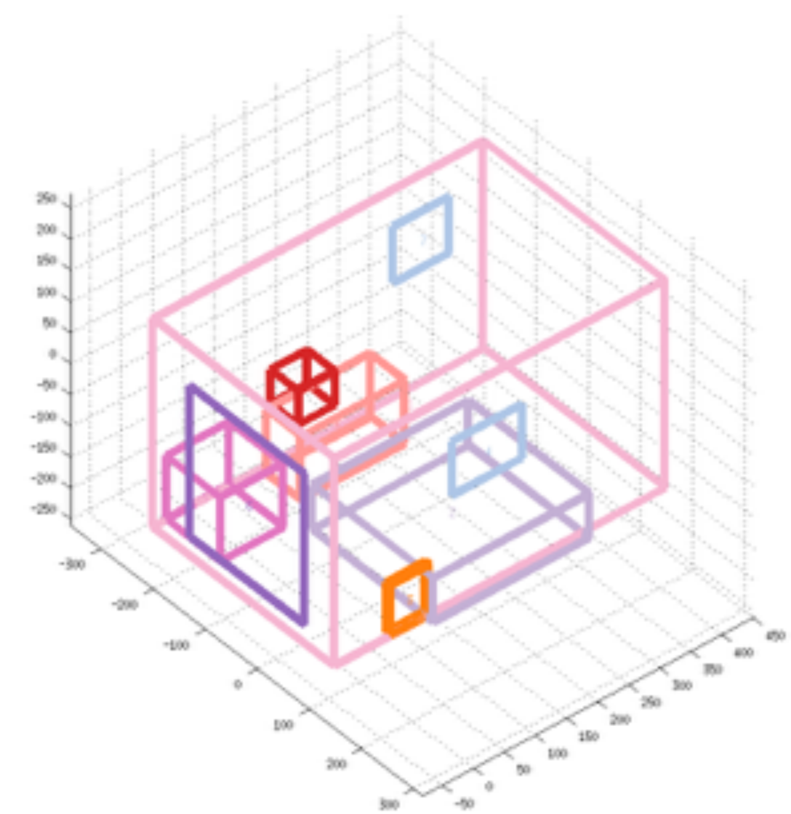


List: bed, nightstand, painting, desk, window, painting, TV, sofa, mirror

Slide credit: Zhang et al.

Data-driven sampling

Keep on sampling until finishing the list…

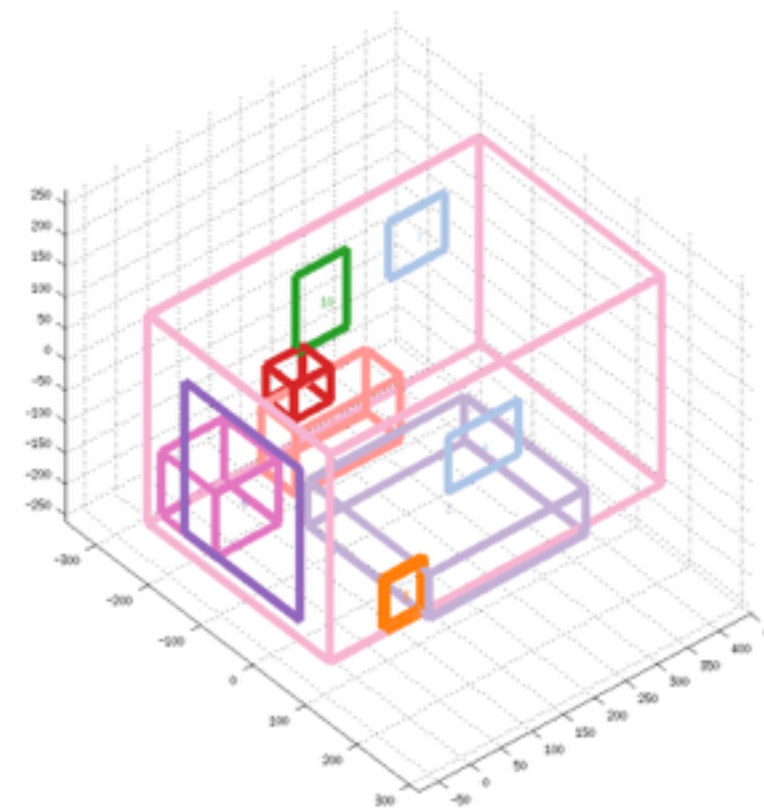


List: bed, nightstand, painting, desk, window, painting, TV, sofa, mirror

Slide credit: Zhang et al.

Data-driven sampling

Keep on sampling until finishing the list…

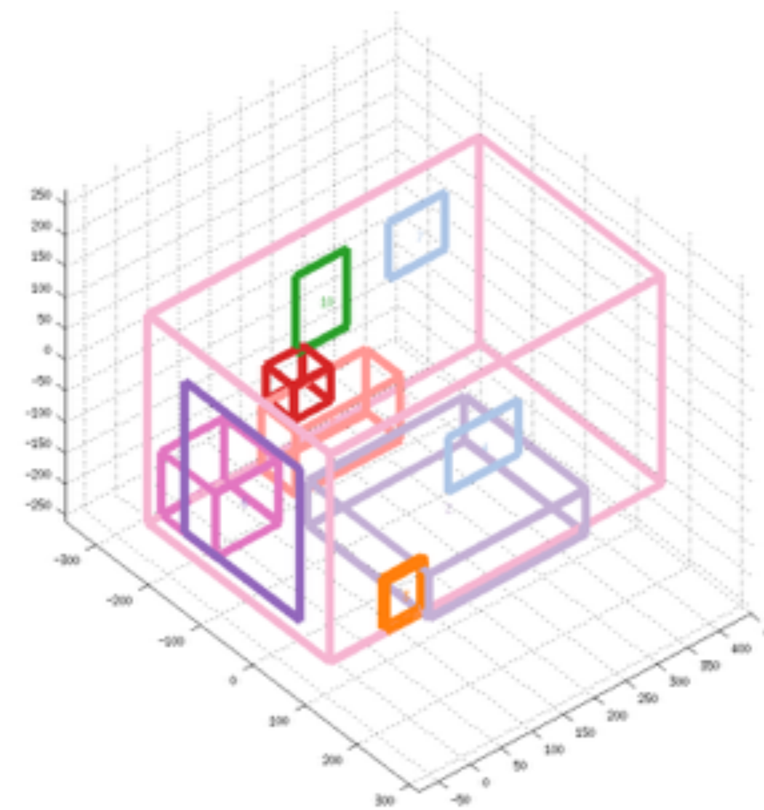


List: bed, nightstand, painting, desk, window, painting, TV, sofa, mirror

Slide credit: Zhang et al.

Data-driven sampling

Keep on sampling until finishing the list…



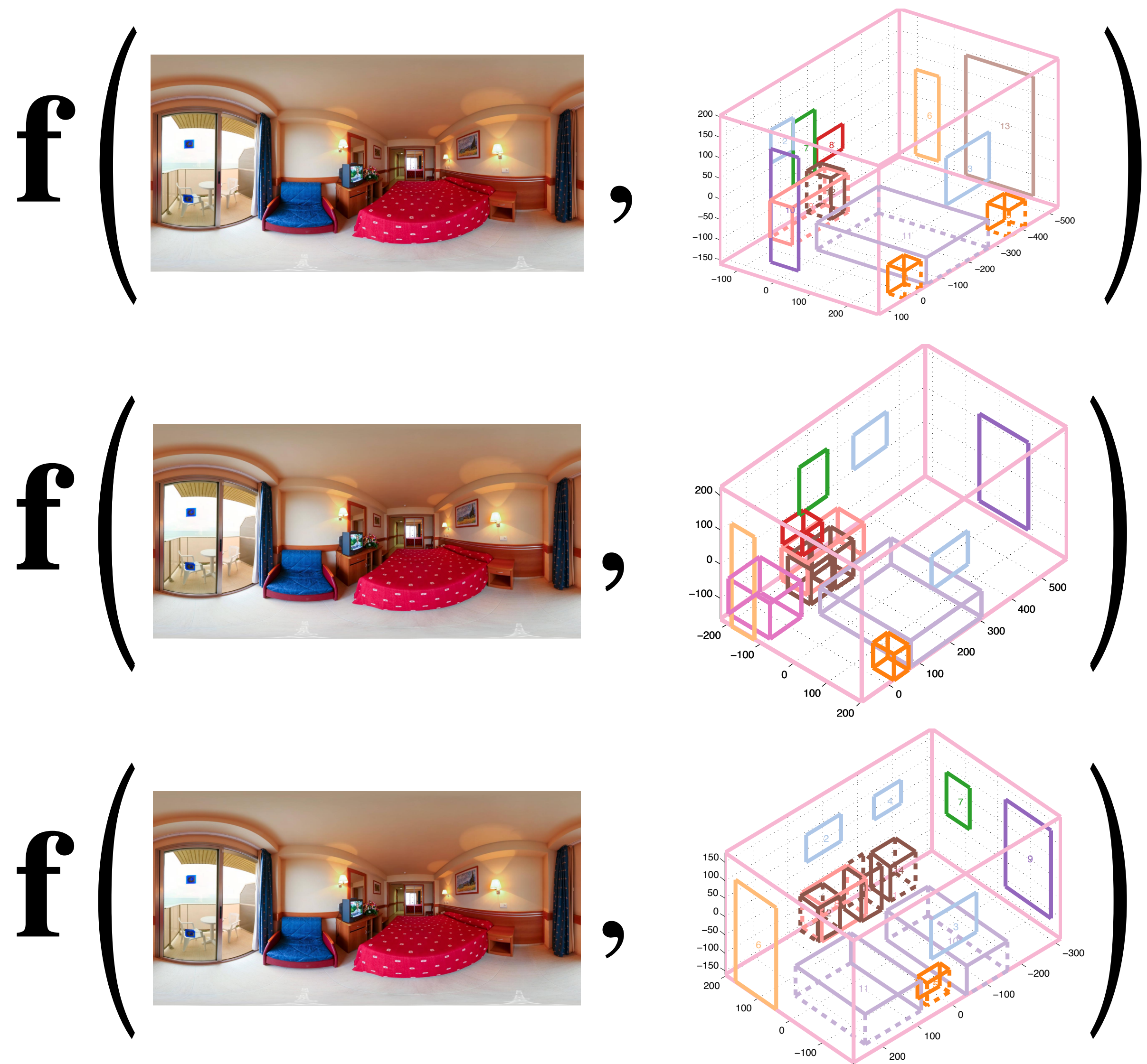
Whole-room sampling is finished.

Slide credit: Zhang et al.

Holistic ranking

Learn a linear SVM for scoring and take the best

N MN

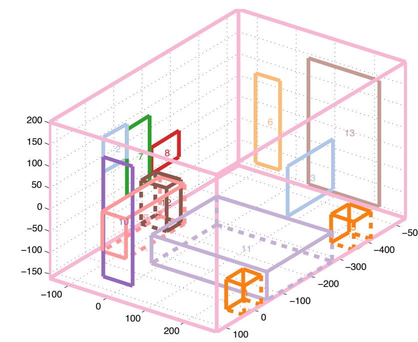


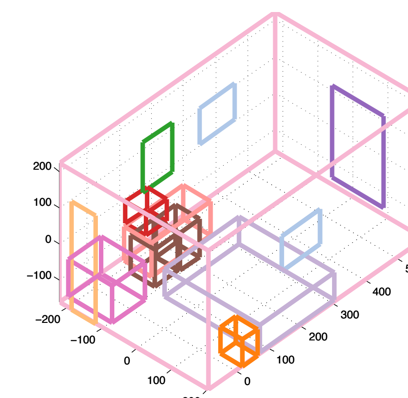
Slide credit: Zhang et al.

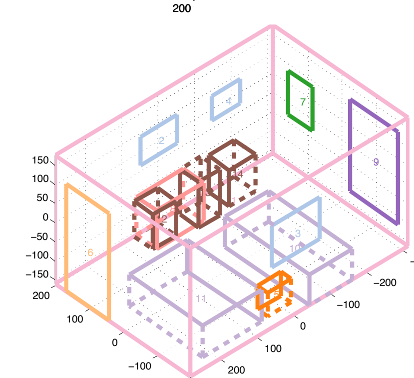
Holistic ranking

Learn a linear SVM for scoring and take the best

N MN

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

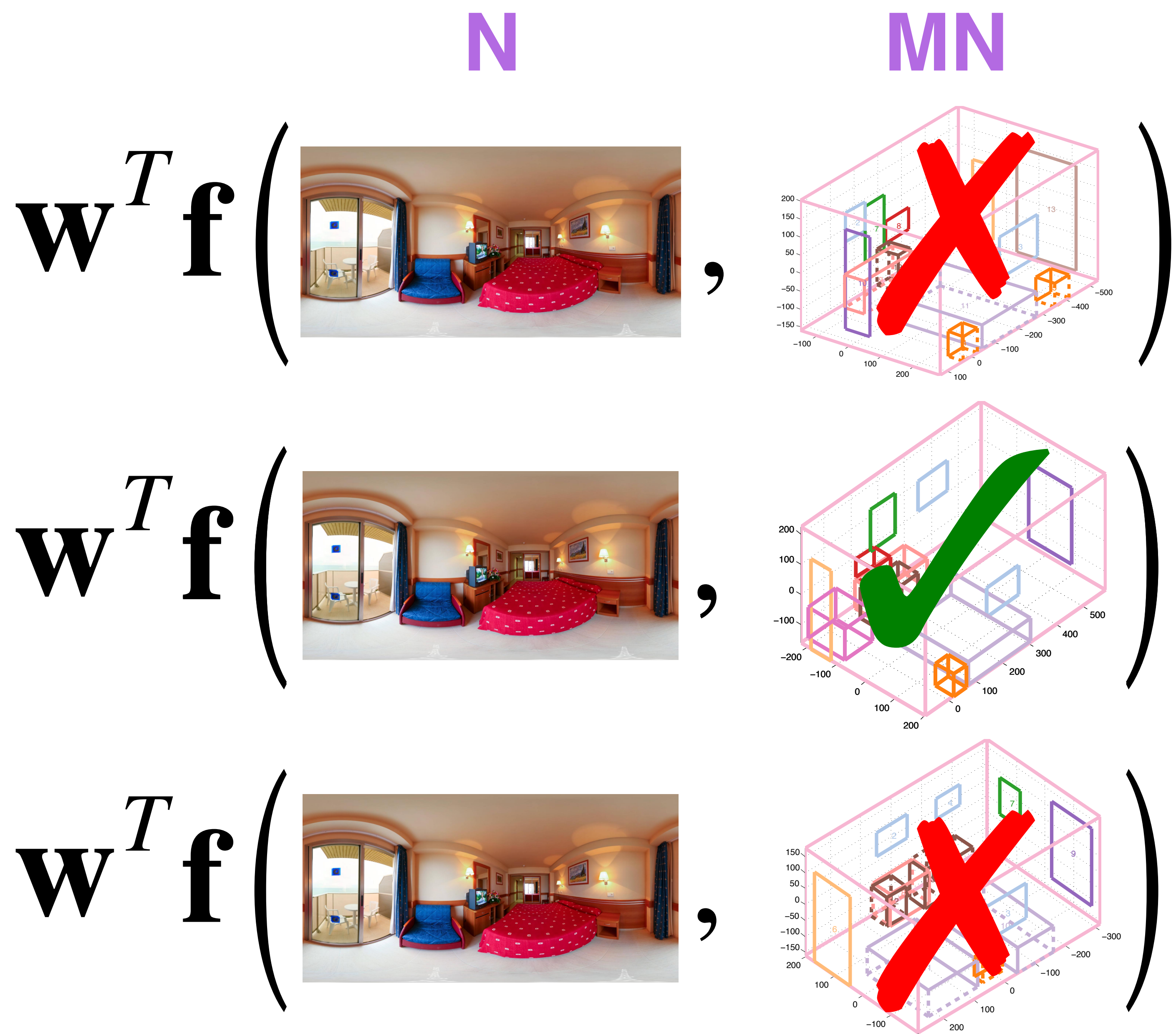
$\mathbf{w}^T\mathbf{f}\Big($  ,  $\Big)$

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

Slide credit: Zhang et al.

Holistic ranking

Learn a linear SVM for scoring and take the best

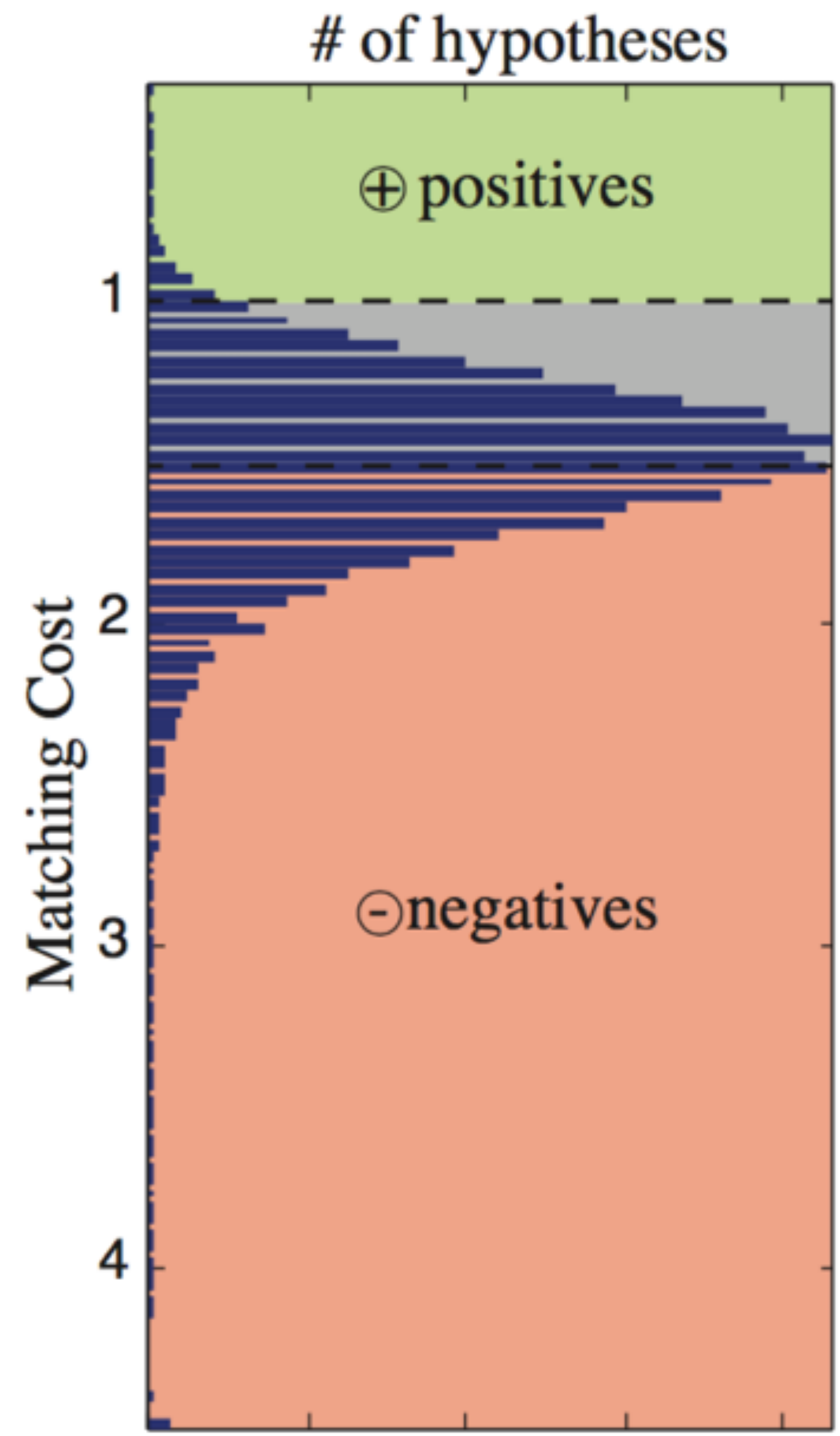


Slide credit: Zhang et al.

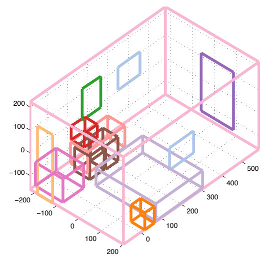
Holistic ranking

$$l = [\Delta(\mathbf{y}, \mathbf{y}^*) < \epsilon].$$

Binary label **Matching cost**

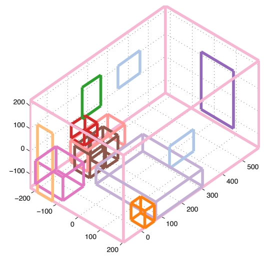


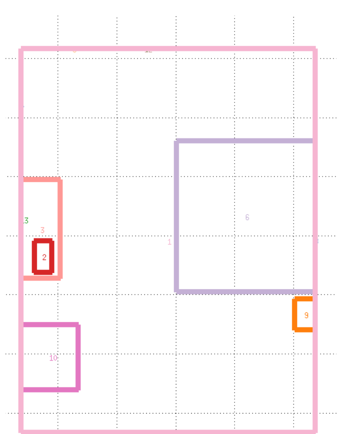
Holistic feature

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

Slide credit: Zhang et al.

Holistic feature

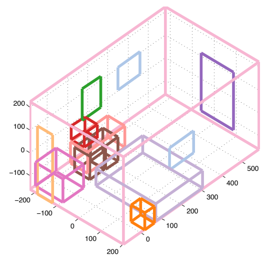
f(, ) = bottom-up feature + **top-down feature**

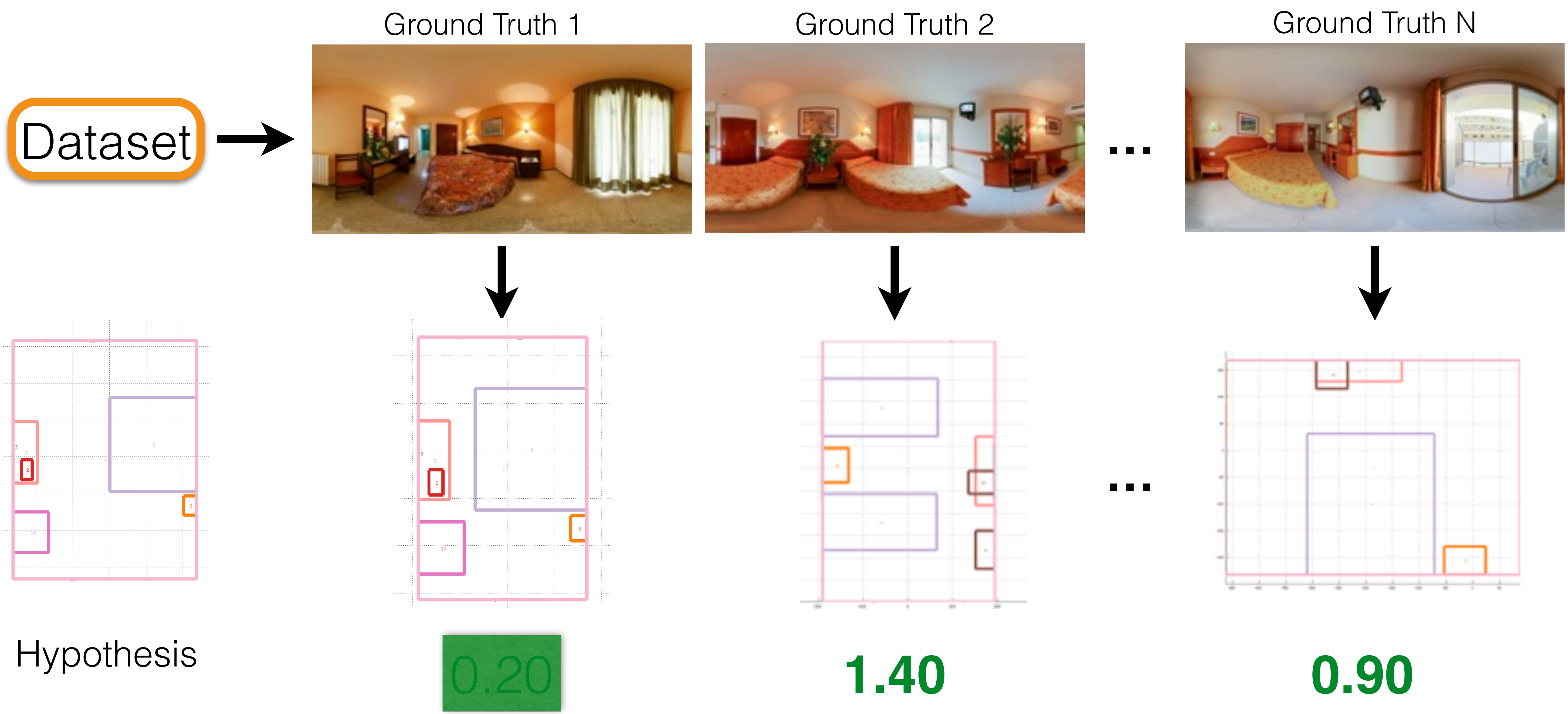


Hypothesis

Slide credit: Zhang et al.

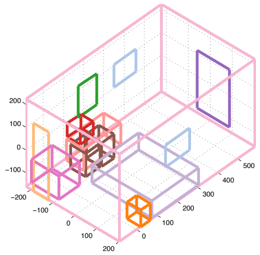
Holistic feature

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



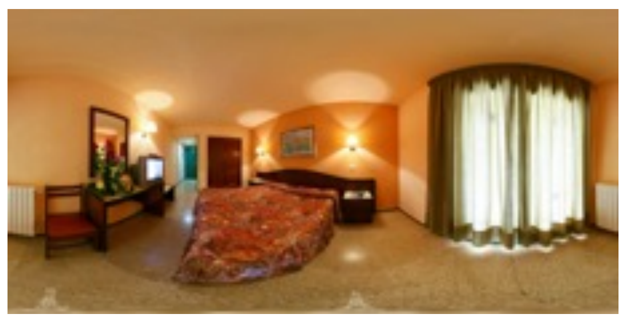
Slide credit: Zhang et al.

Holistic feature

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

Dataset →

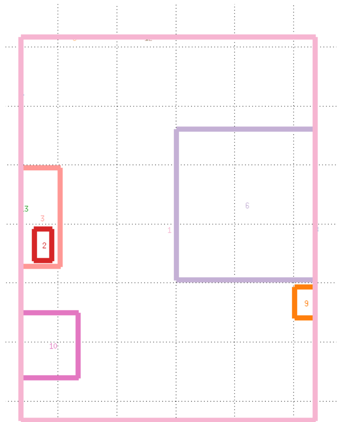
Ground Truth 1



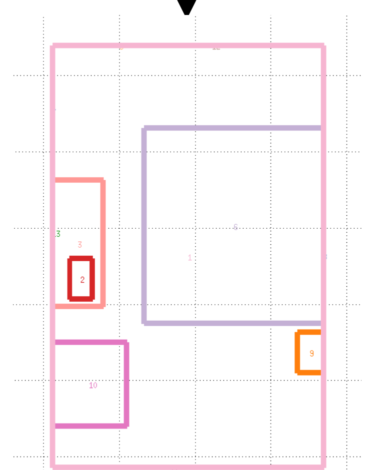
Ground Truth 2

...

Ground Truth N



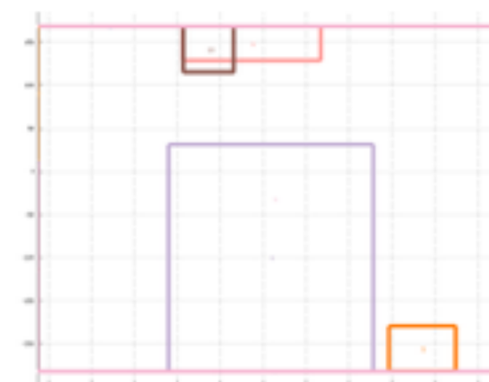
Hypothesis



0.20

1.40

...

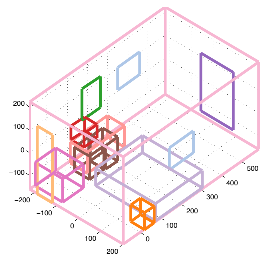


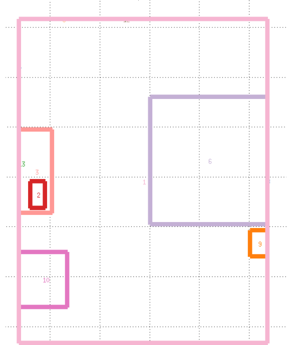
0.90

Centroid distance, IoU, semantic type consistency

Slide credit: Zhang et al.

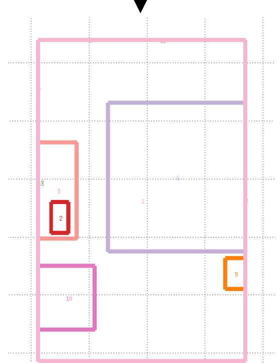
Holistic feature

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



Hypothesis

Dataset



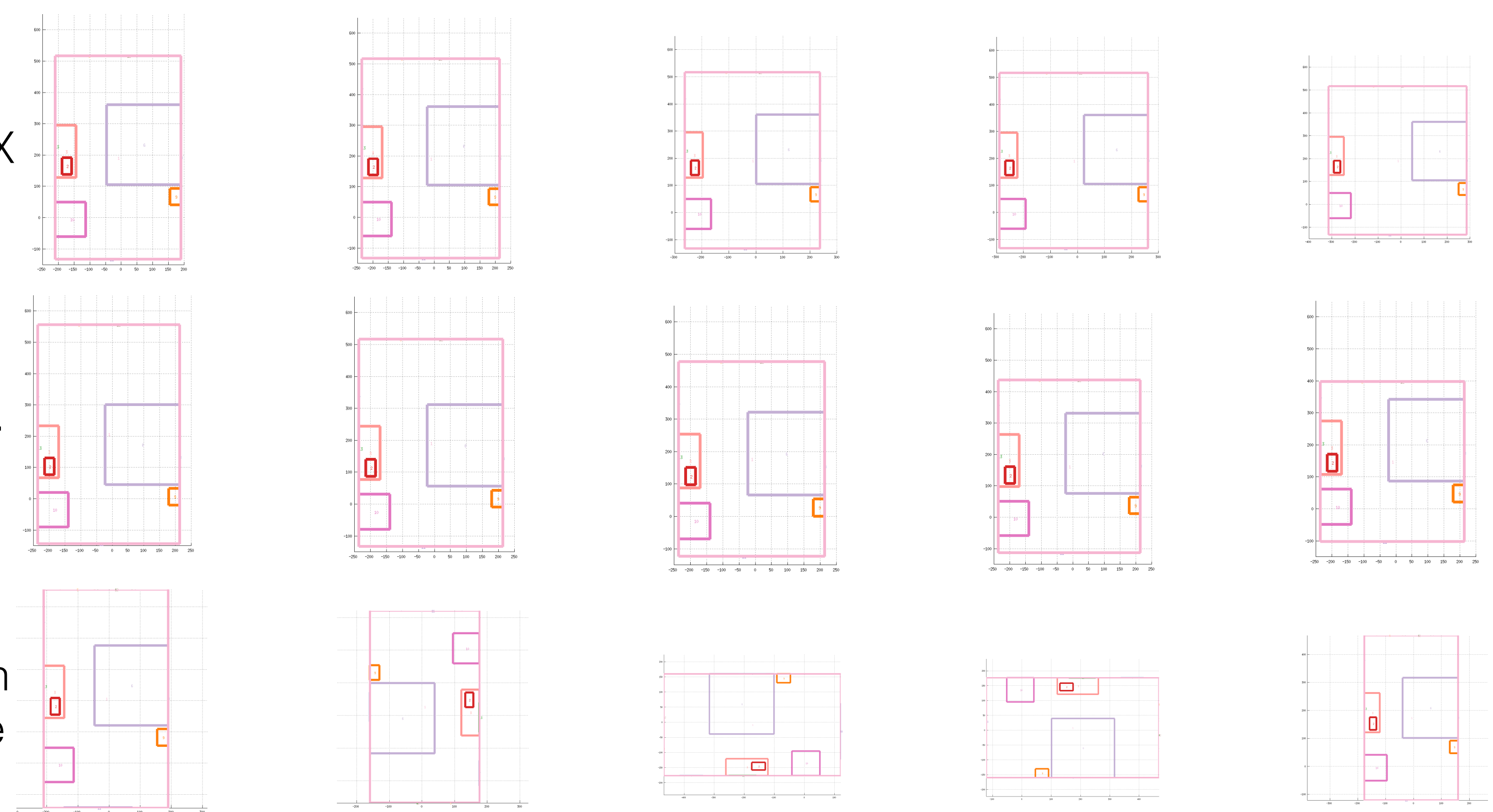
A ground truth room

Transformed ground truth

Resize X

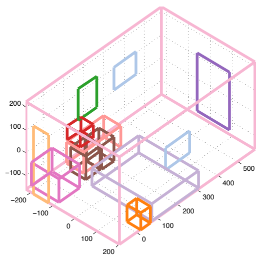
Fix dist. to wall

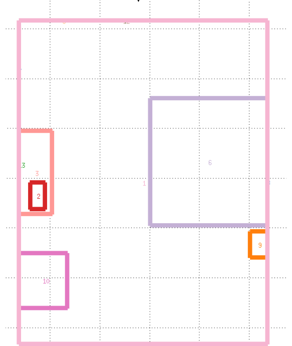
Rotation & Scale



Slide credit: Zhang et al.

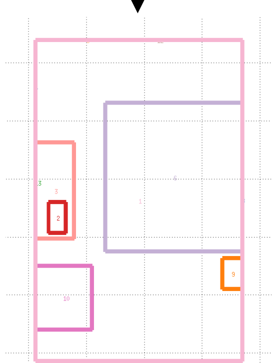
Holistic feature

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



Hypothesis

Dataset



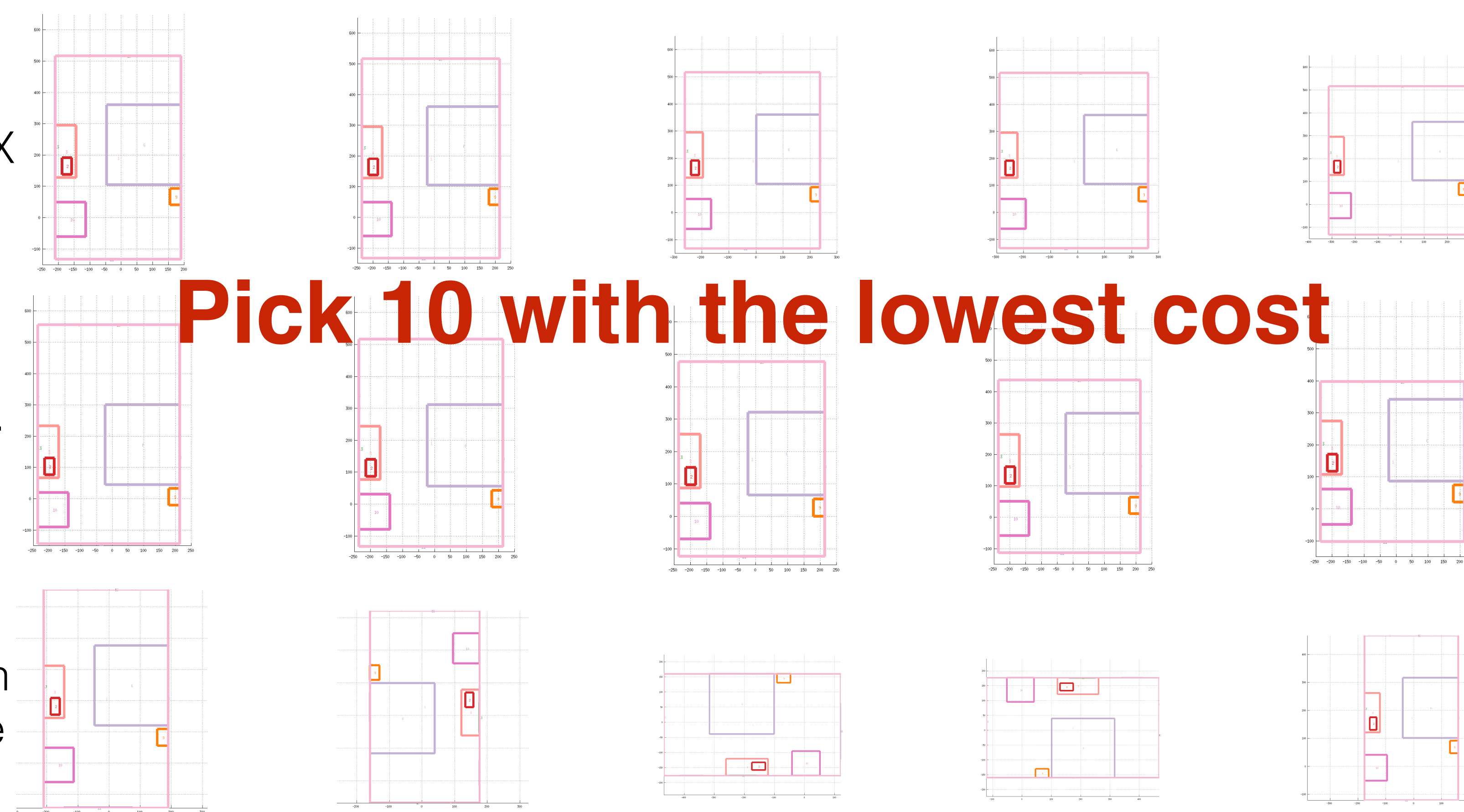
A ground truth room

Transformed ground truth

Resize X

Fix dist. to wall

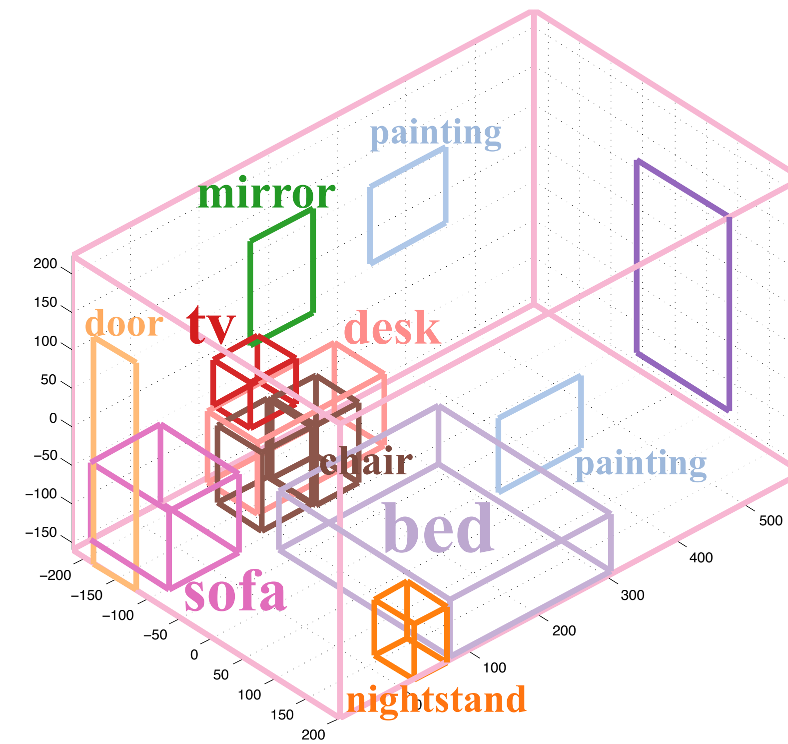
Rotation & Scale



Pick 10 with the lowest cost

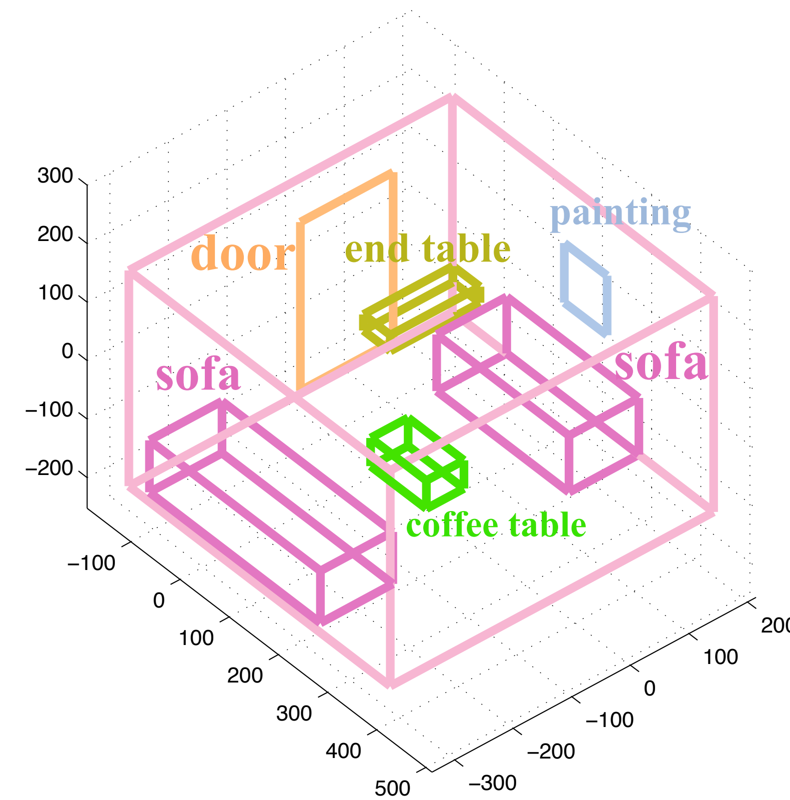
Slide credit: Zhang et al.

Final outputs



Slide credit: Zhang et al.

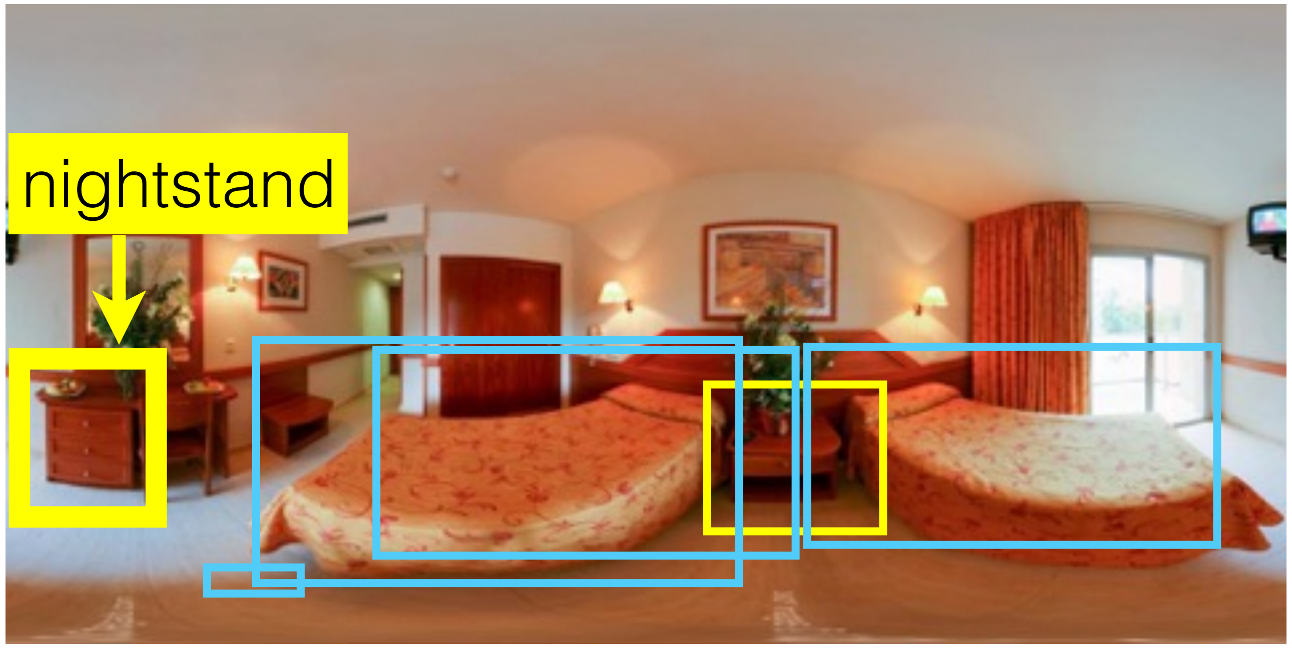
Final outputs



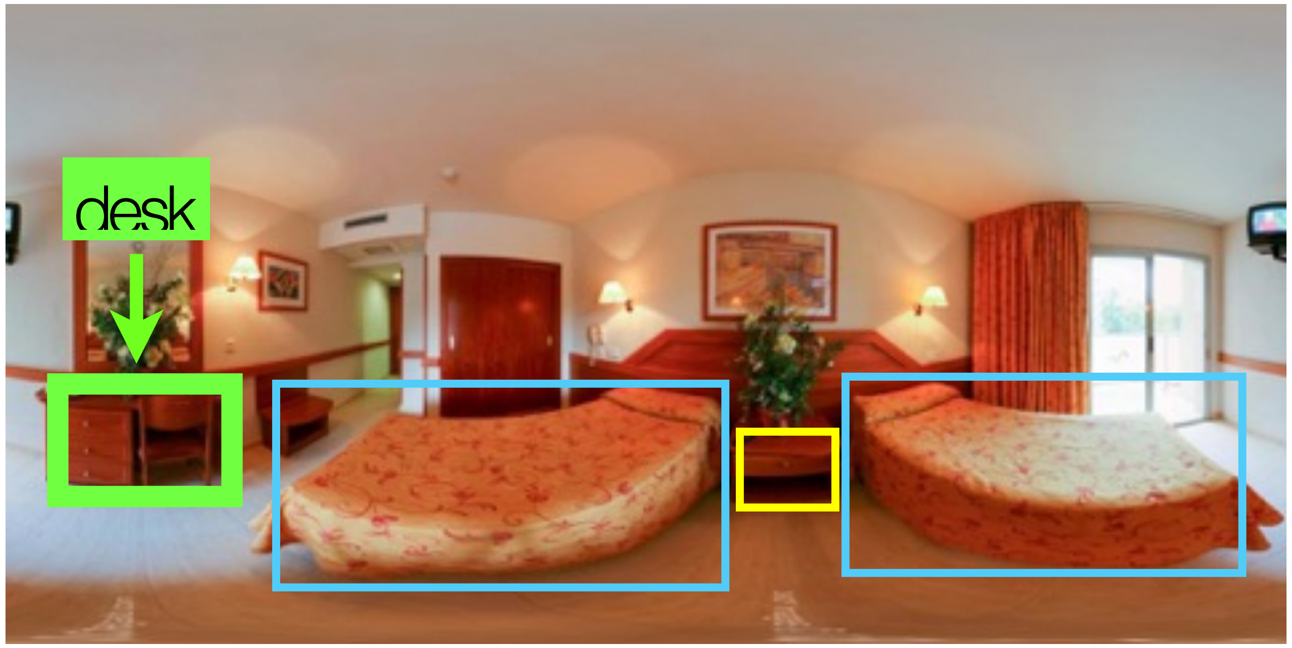
Slide credit: Zhang et al.

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

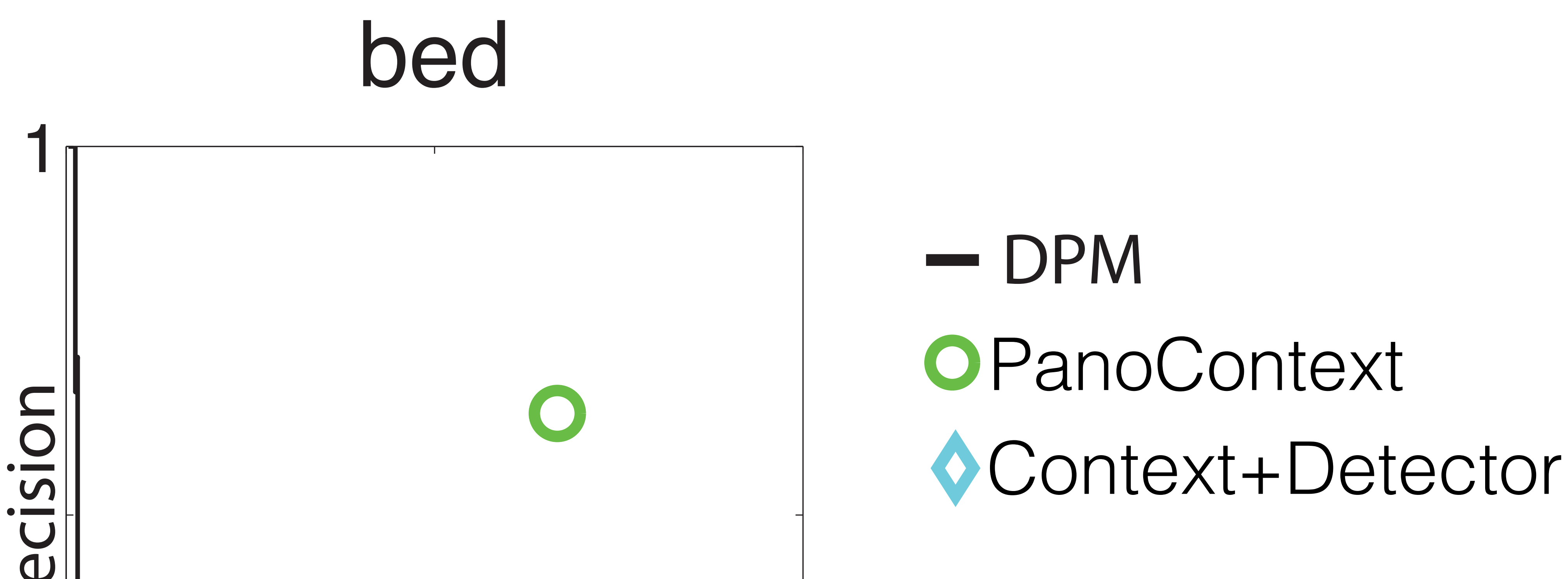


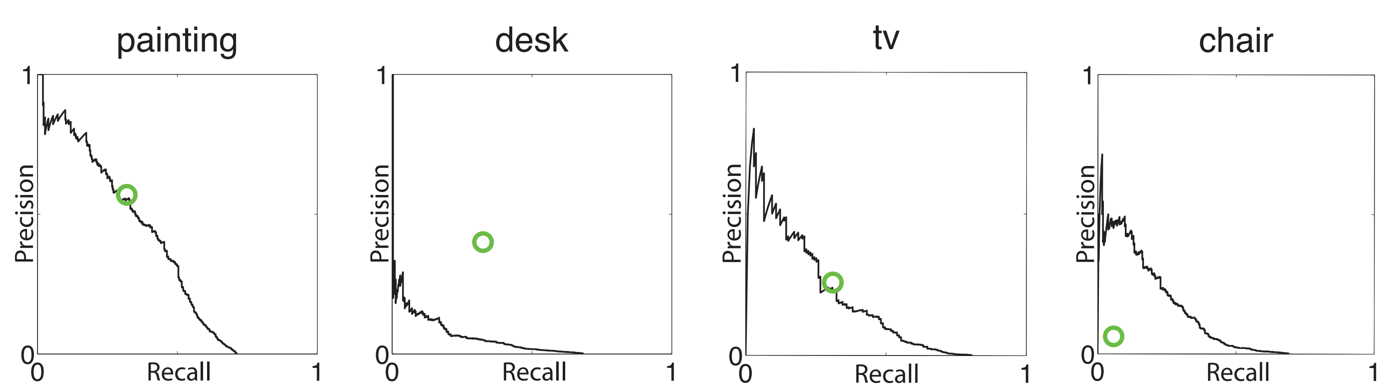
Our detection

Slide credit: Zhang et al.

Context v.s. Appearance

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

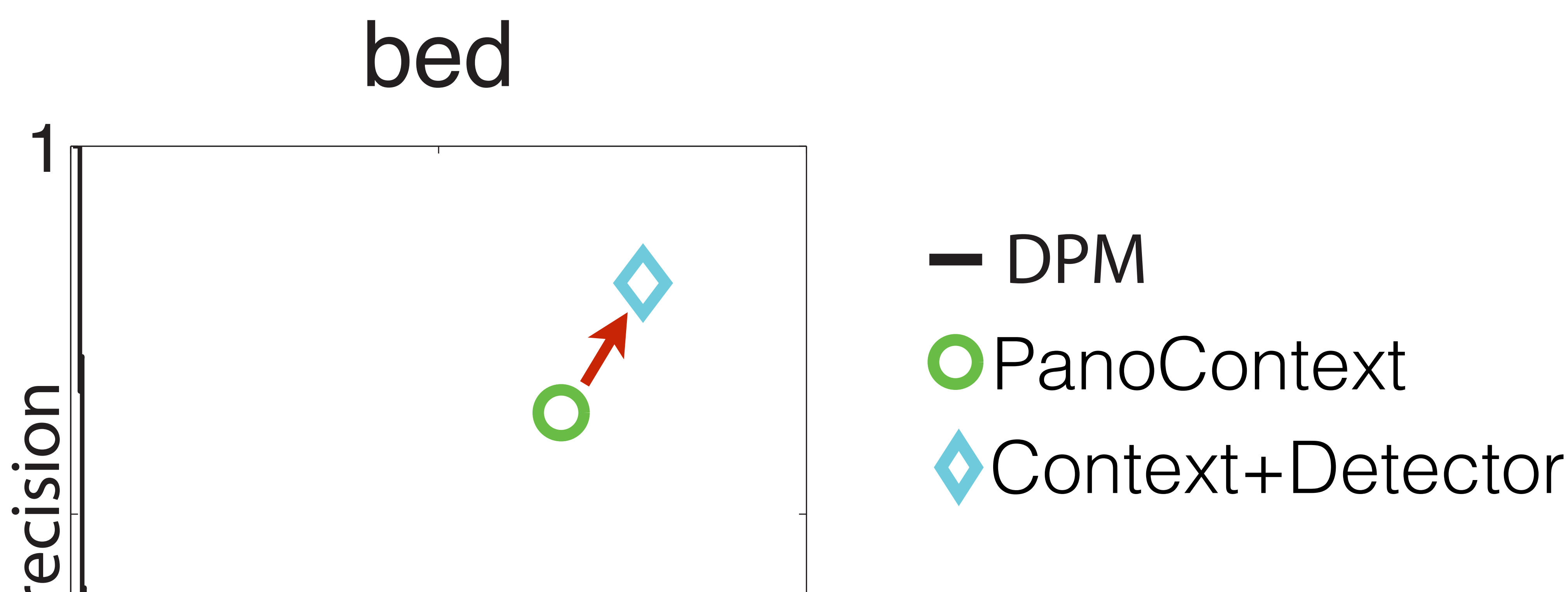


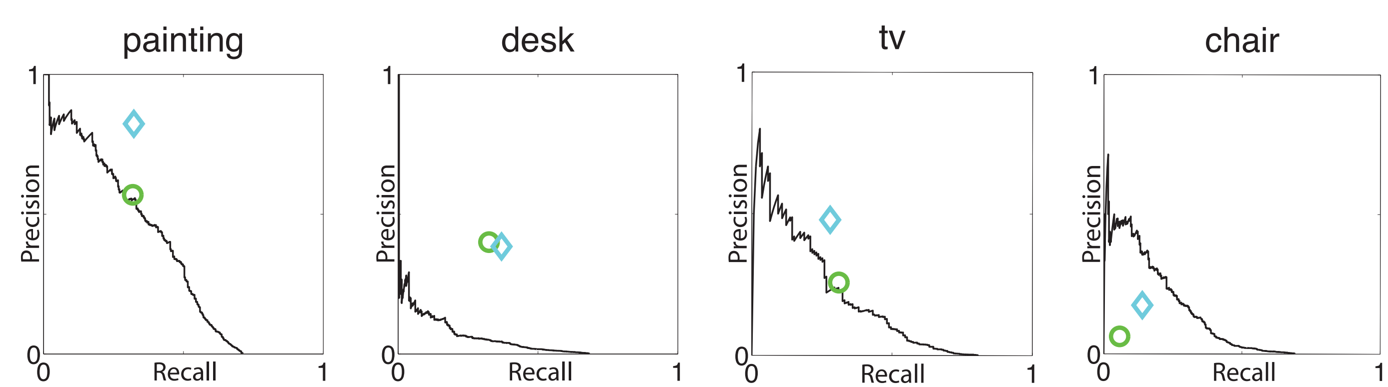


Slide credit: Zhang et al.

Context v.s. Appearance

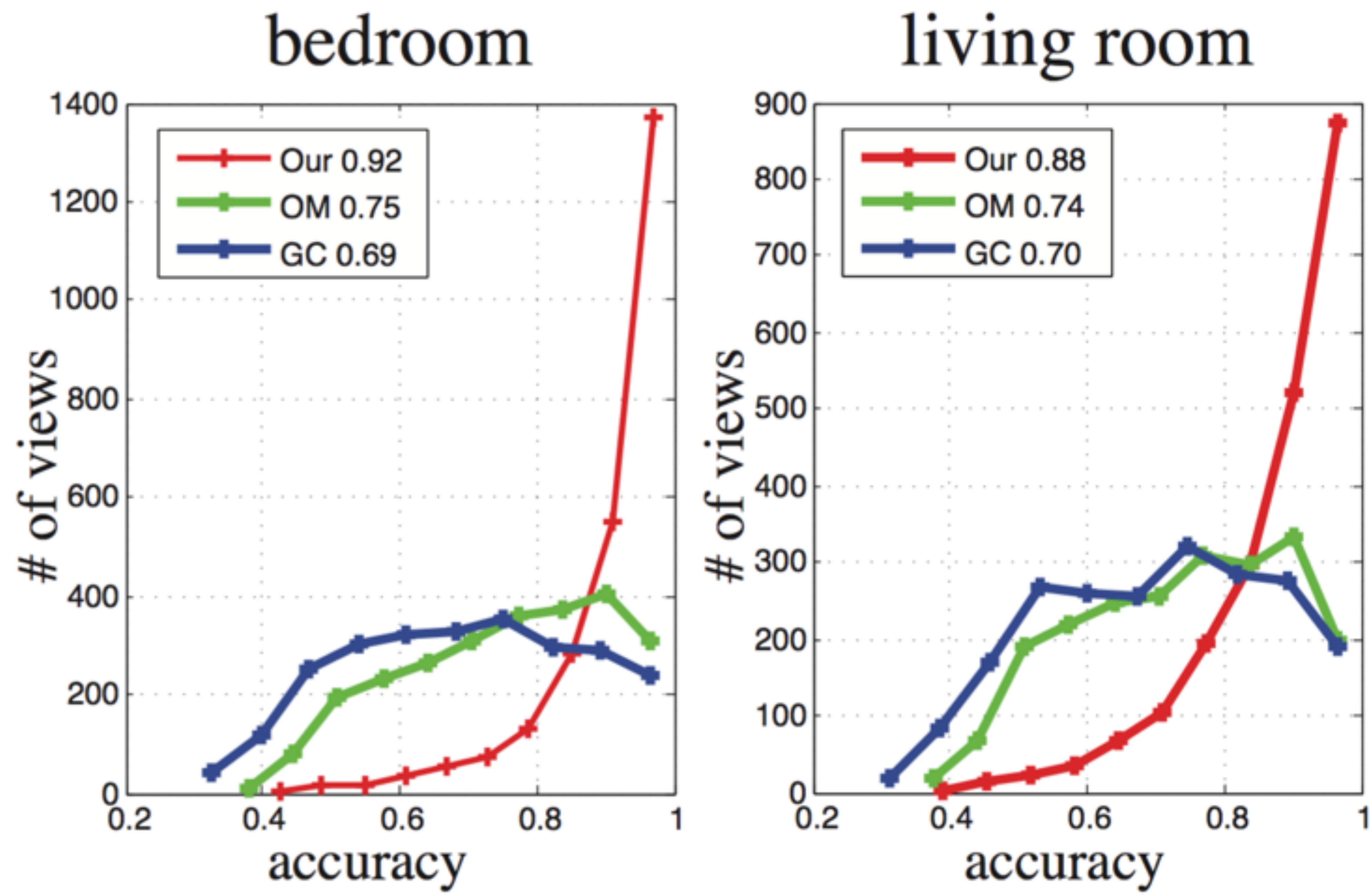
- Context is as powerful as local appearance for detection
- Context is complementary with local appearance



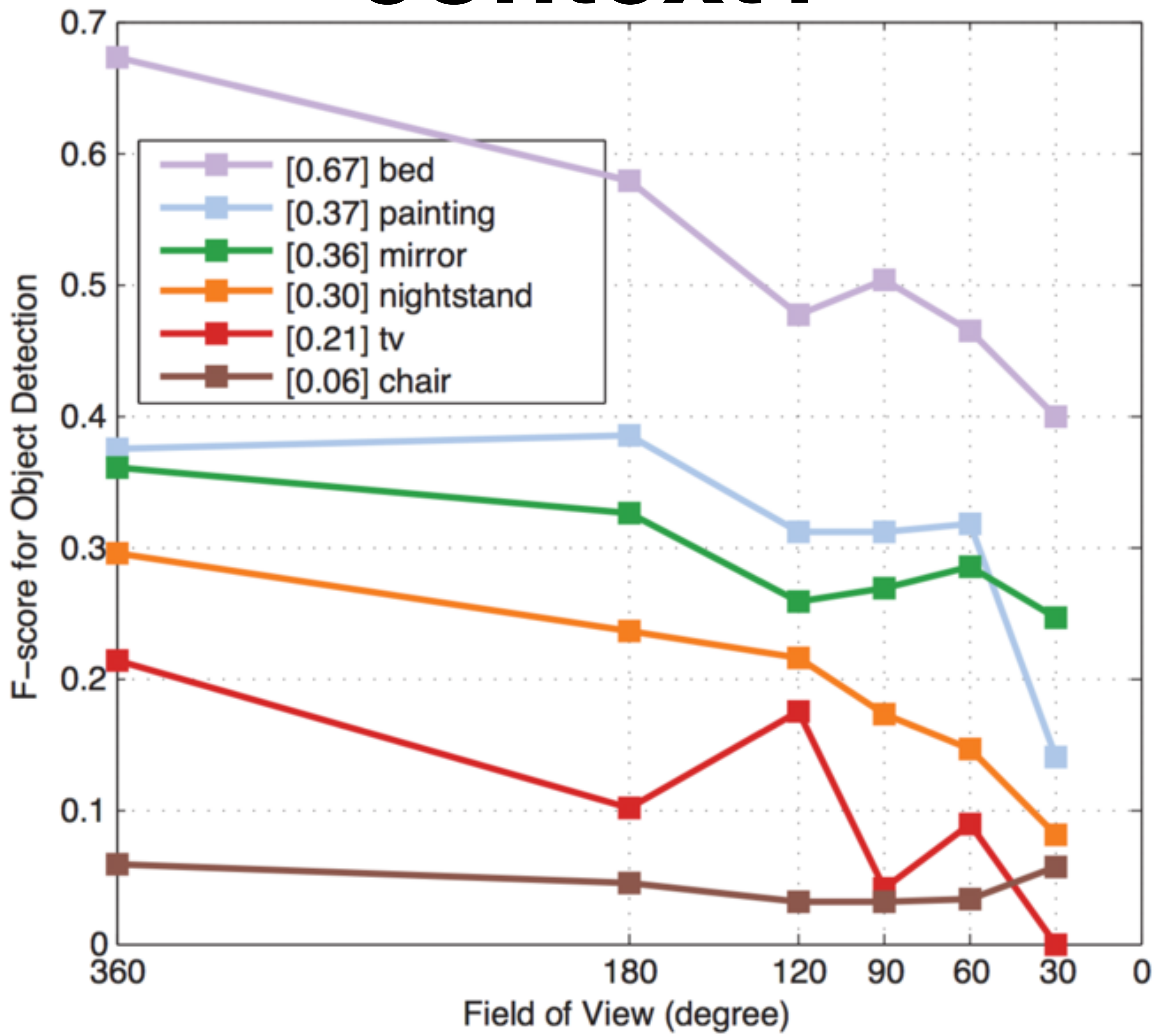


Slide credit: Zhang et al.

Is larger FOV helpful for room layout estimation?



Is larger FOV better for context?



My Take

- Elements of the ensemble could be valuable
- Too data driven, hard to generalize
- Future: relax the cuboid constraints, try other ways to integrate visual recognition in the pipeline

Discussion

- How can the model be generalized to other scene categories (e.g. outdoor)?
- Performance on deformable or non-axis aligned objects?
- Chairs and other non-standard layout objects?
- Indoor understanding and VQA?

Is context important in sampling and ranking?

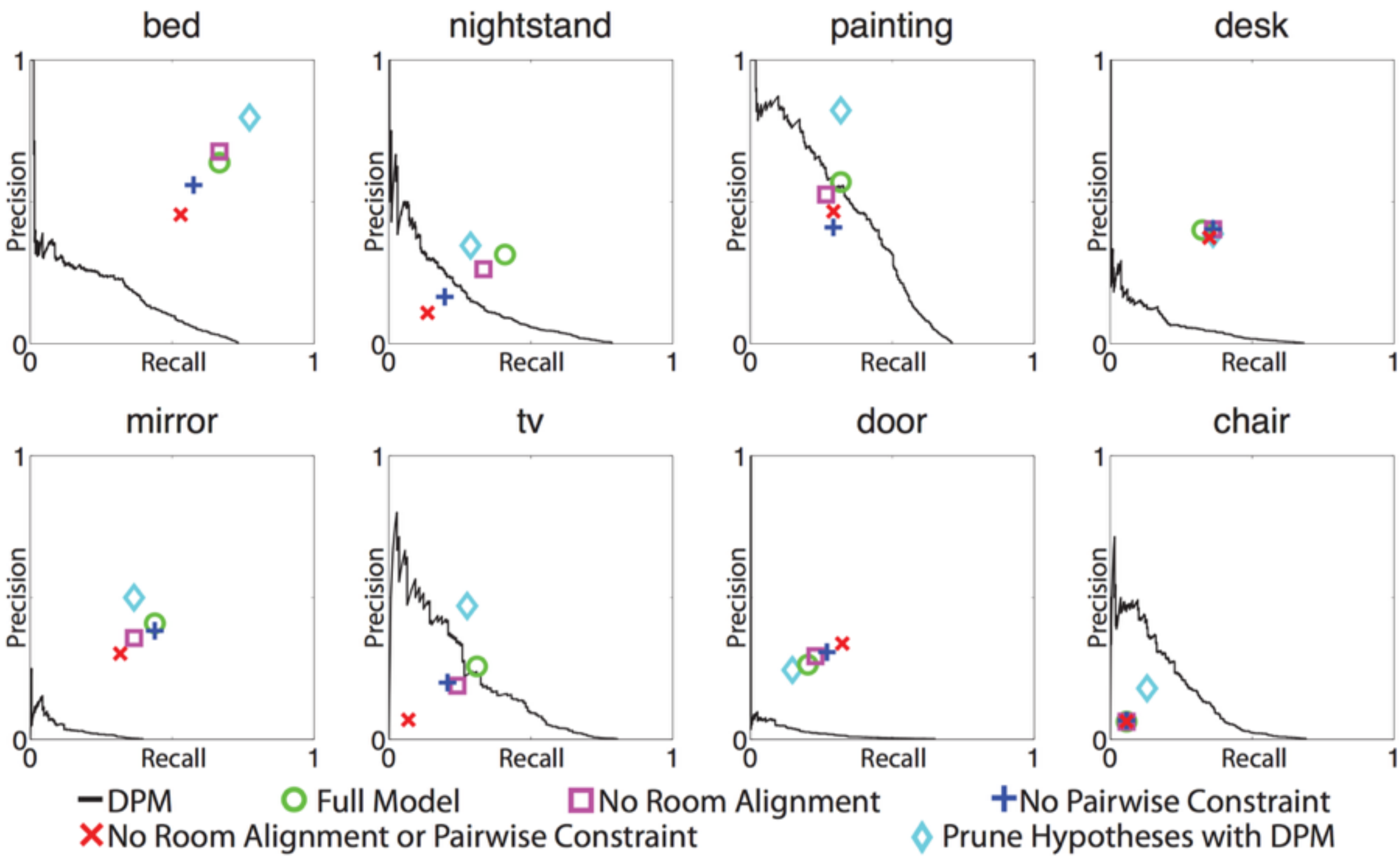


Table 2: Object detection performance

(a) bedroom

object type	bed	desk	window	mirror	door	nightstand	wardrobe	cabinet	painting	tv	chair	sofa
global precision (%)	62.16	40.28	24.00	28.89	30.65	27.50	13.89	0.00	54.79	25.00	6.15	0
global recall (%)	69.70	36.25	22.64	31.71	25.68	33.33	17.86	0.00	34.48	27.59	5.80	0
local precision (%)	63.15	47.89	22.45	34.78	29.23	36.36	16.22	12.50	57.14	27.03	11.59	20.00
local recall (%)	71.21	42.50	20.75	39.02	25.68	48.48	21.43	5.88	37.93	34.48	11.59	3.23

(b) living room

object type	painting	door	cabinet	dining table	window	heater	chair	sofa	coffee table	end table	tv stand
global precision (%)	43.75	30.25	15.00	39.29	16.00	0.00	22.39	44.09	37.84	0.00	6.25
global recall (%)	44.21	27.69	9.38	30.56	8.00	0.00	11.90	39.05	33.33	0.00	4.35
local precision (%)	59.49	45.36	22.73	38.71	30.77	20.00	21.05	59.49	39.39	20.00	22.22
local recall (%)	49.47	33.85	15.63	33.33	16.00	16.67	9.52	44.76	30.95	5.88	8.70

Table 3: Semantic labeling accuracy

(a) bedroom

object type	background	bed	desk	window	mirror	door	nightstand	wardrobe	cabinet	painting	tv	chair	sofa
global (%)	86.90	78.58	29.55	35.58	38.15	19.40	39.66	27.44	0.00	38.70	34.81	9.61	11.10
local (%)	87.13	80.76	33.10	22.78	42.90	25.47	55.67	25.31	5.46	41.58	32.88	17.20	7.74

(b) living room

object type	background	painting	door	cabinet	dining table	window	heater	chair	sofa	coffee table	end table	tv stand
global (%)	91.98	44.66	41.07	7.87	24.24	12.59	0.00	15.46	47.05	42.33	3.87	1.21
local (%)	93.50	47.50	36.75	16.27	21.80	12.37	11.19	14.95	49.47	42.78	3.99	7.66