

# Scene Understanding from RGB-D Images

Object Detection, Semantic and Instance Segmentation

Saurabh Gupta, Ross Girshick, Pablo Arbeláez,  
Jitendra Malik

UC Berkeley

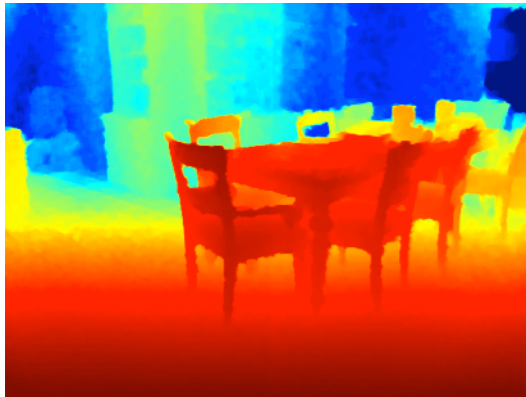
# Learning Rich Features from RGB-D Images for Object Detection and Segmentation

ECCV 2014

Saurabh Gupta, Ross Girshick, Pablo Arbelaez,  
Jitendra Malik

# Overview

## Input

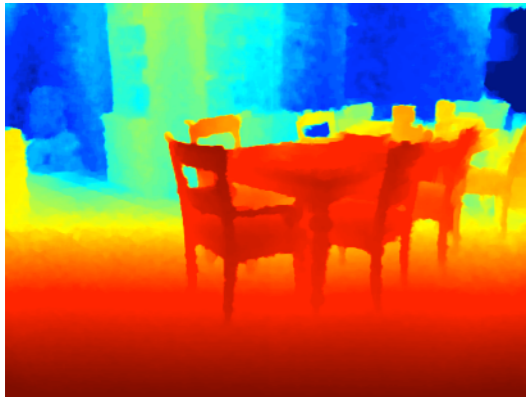


Color and Depth  
Image Pair

# Overview

Input

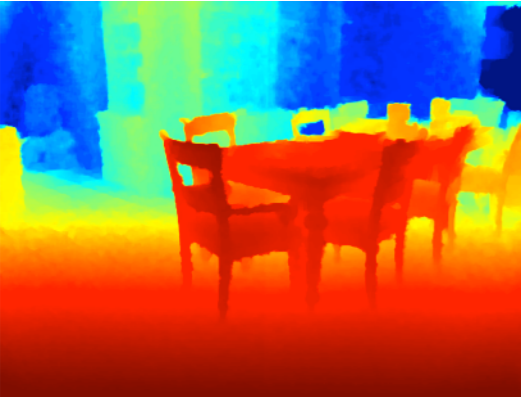
Re-organization



Color and Depth  
Image Pair

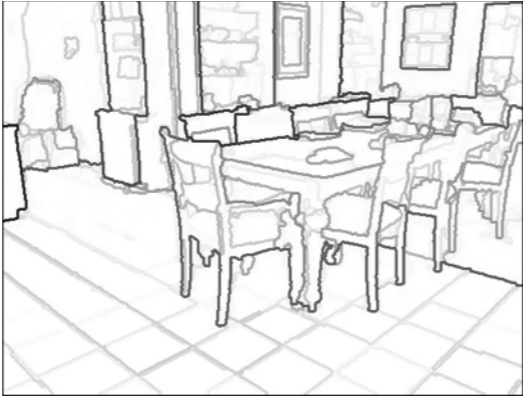
# Overview

## Input



Color and Depth  
Image Pair

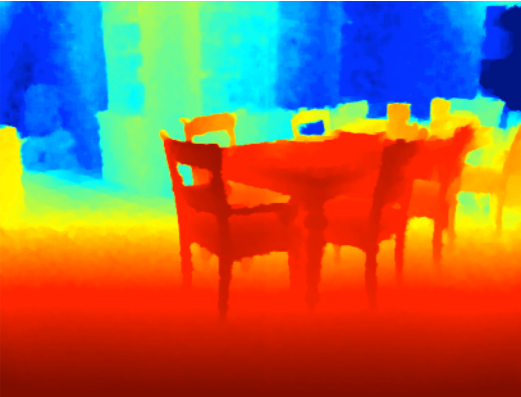
## Re-organization



Contour Detection

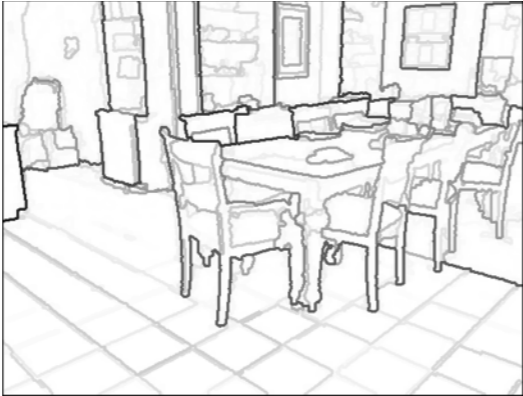
# Overview

## Input



Color and Depth Image Pair

## Re-organization



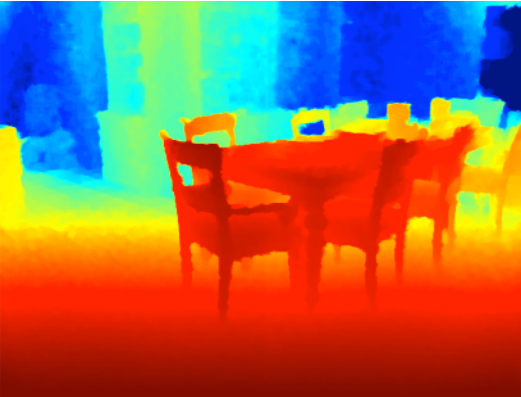
Contour Detection



Region Proposal Generation

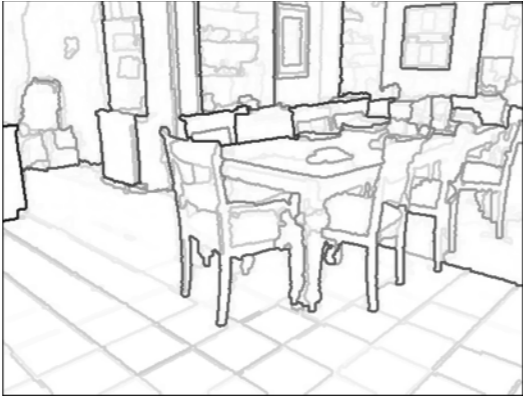
# Overview

## Input



Color and Depth Image Pair

## Re-organization



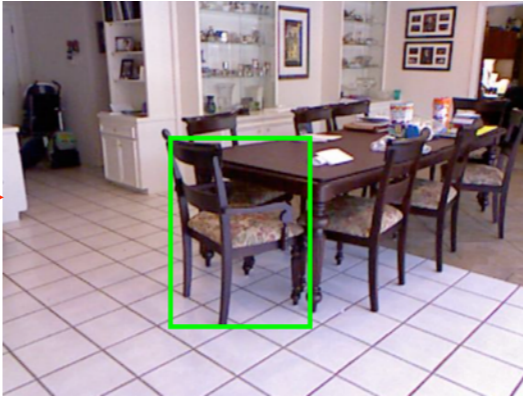
Contour Detection



Region Proposal Generation



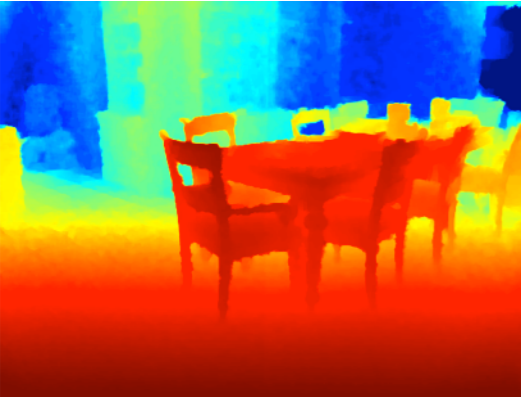
## Recognition



Object Detection

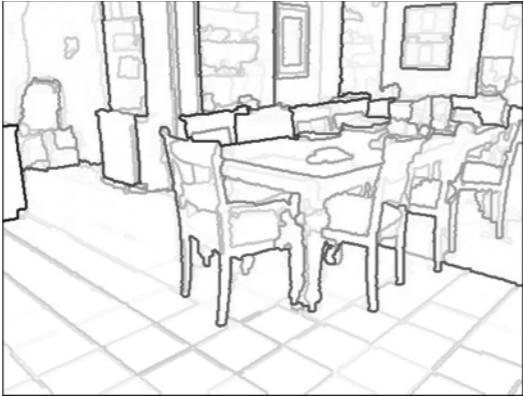
# Overview

## Input

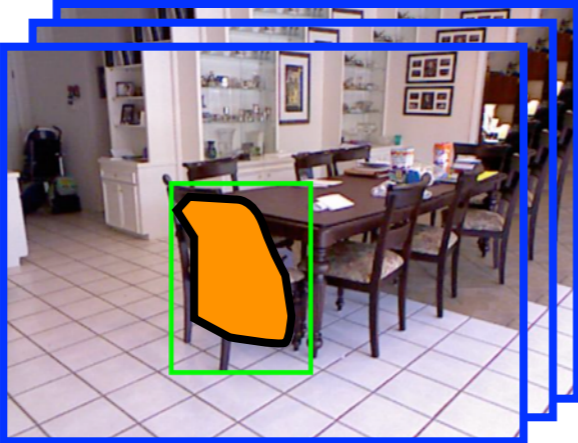


Color and Depth  
Image Pair

## Re-organization

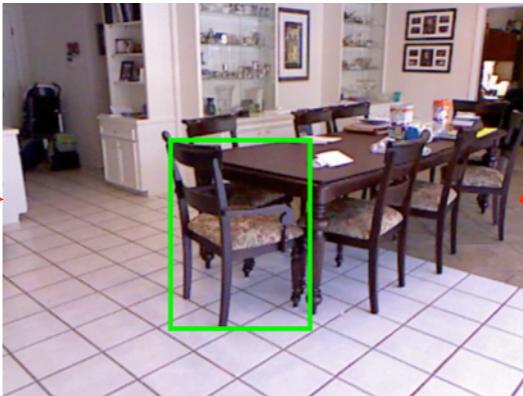


Contour Detection



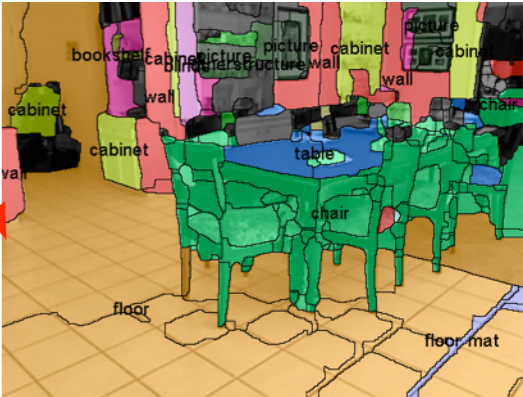
Region Proposal  
Generation

## Recognition



Object Detection

## Extensions



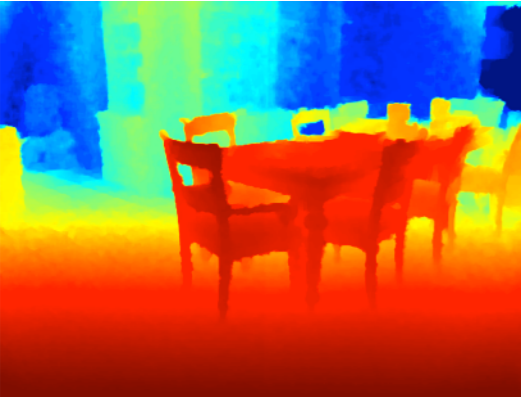
Semantic Segm.





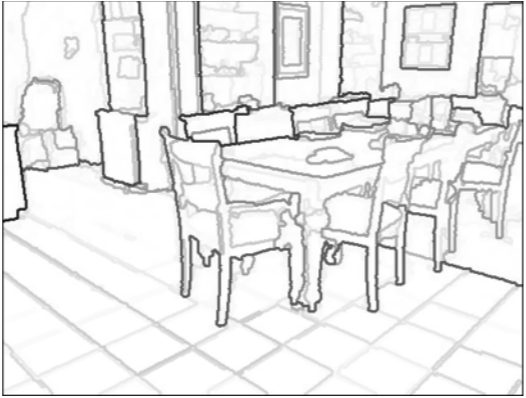
# Overview

## Input

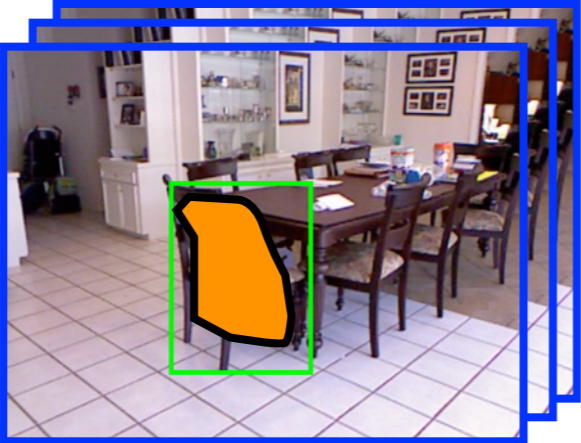


Color and Depth Image Pair

## Re-organization

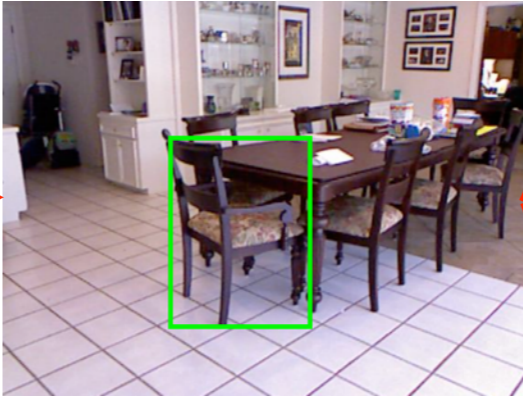


Contour Detection



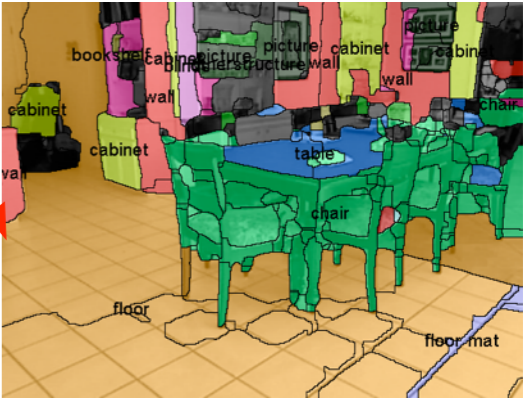
Region Proposal Generation

## Recognition

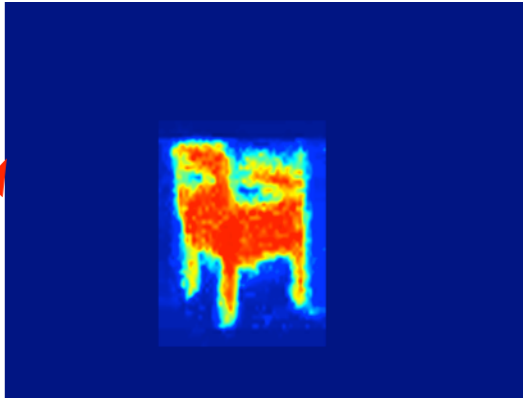


Object Detection

## Extensions



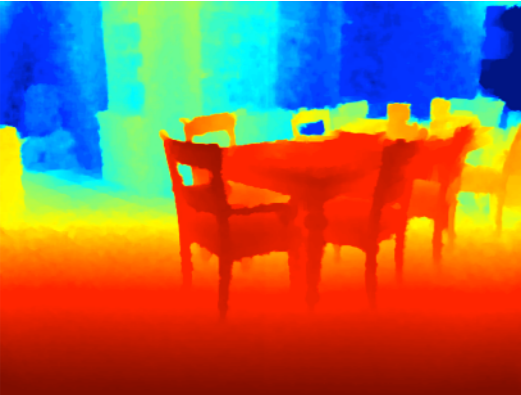
Semantic Segm.



Instance Segm.

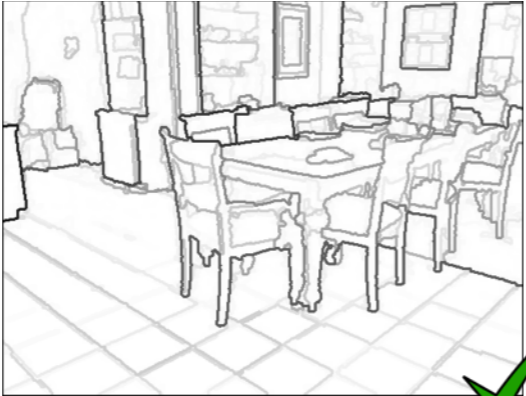
# Overview

## Input

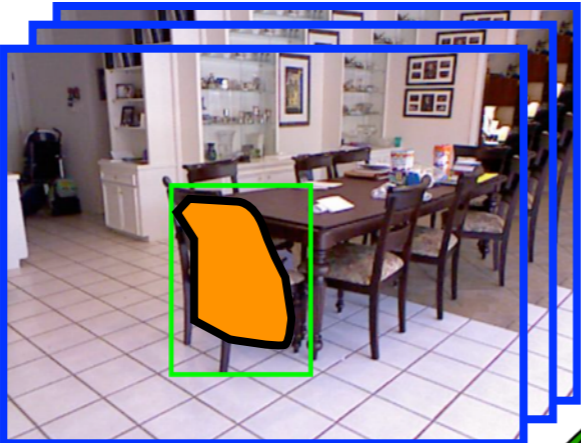


Color and Depth Image Pair

## Re-organization

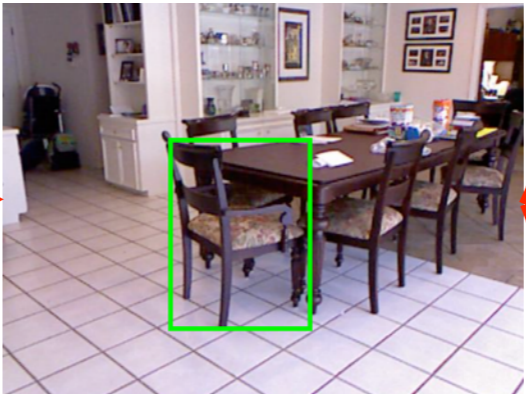


Contour Detection



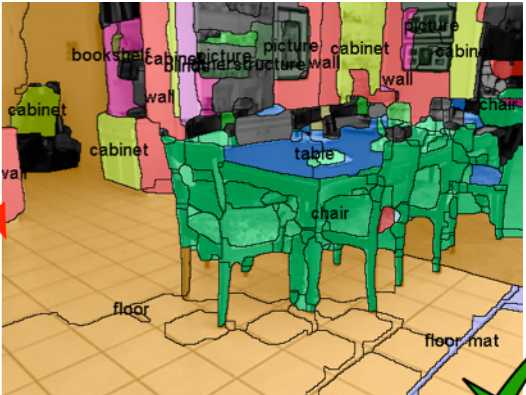
Region Proposal Generation

## Recognition

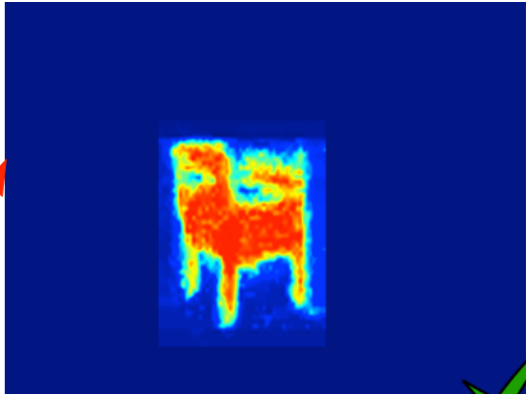


Object Detection

## Extensions



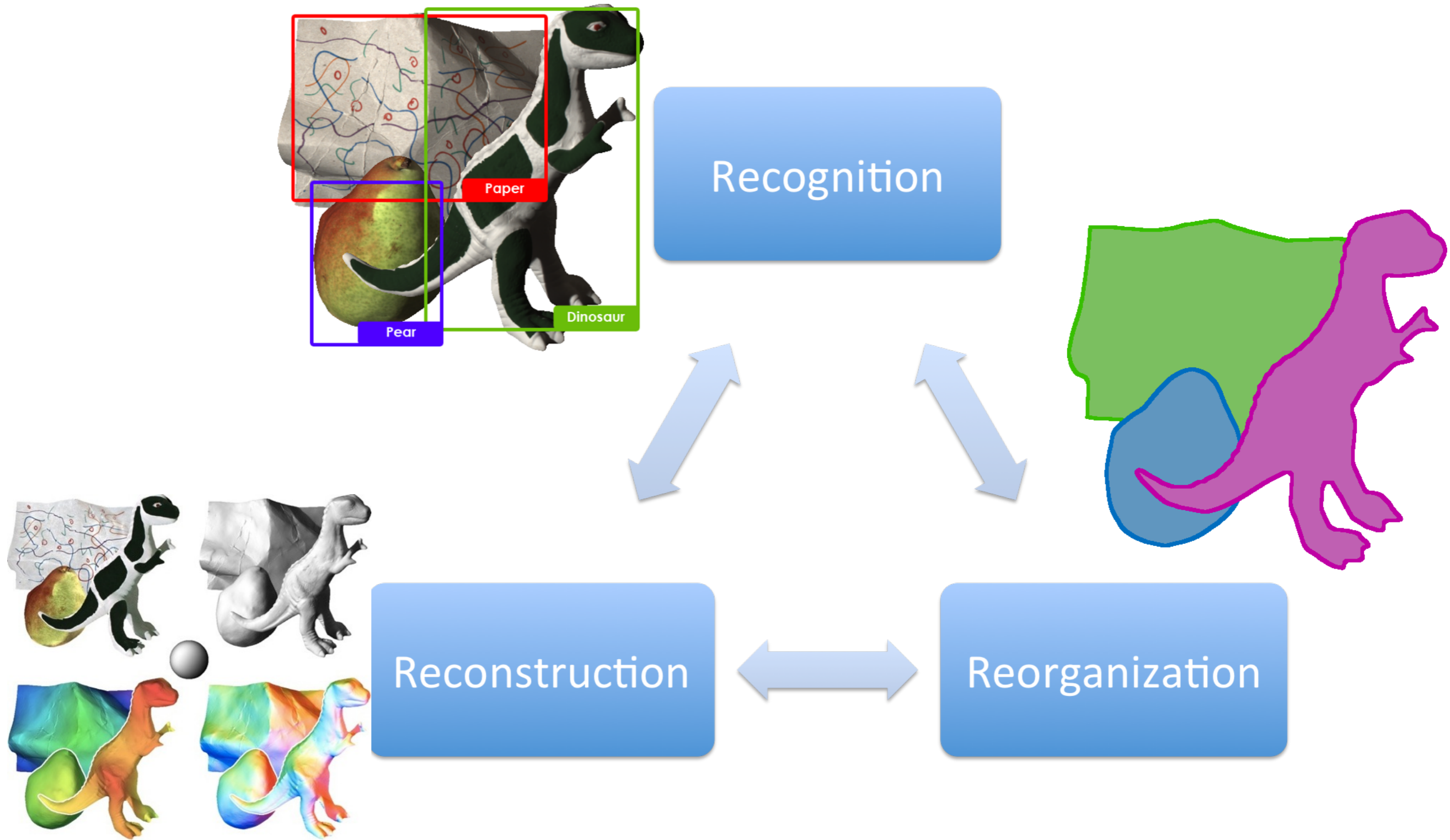
Semantic Segm.



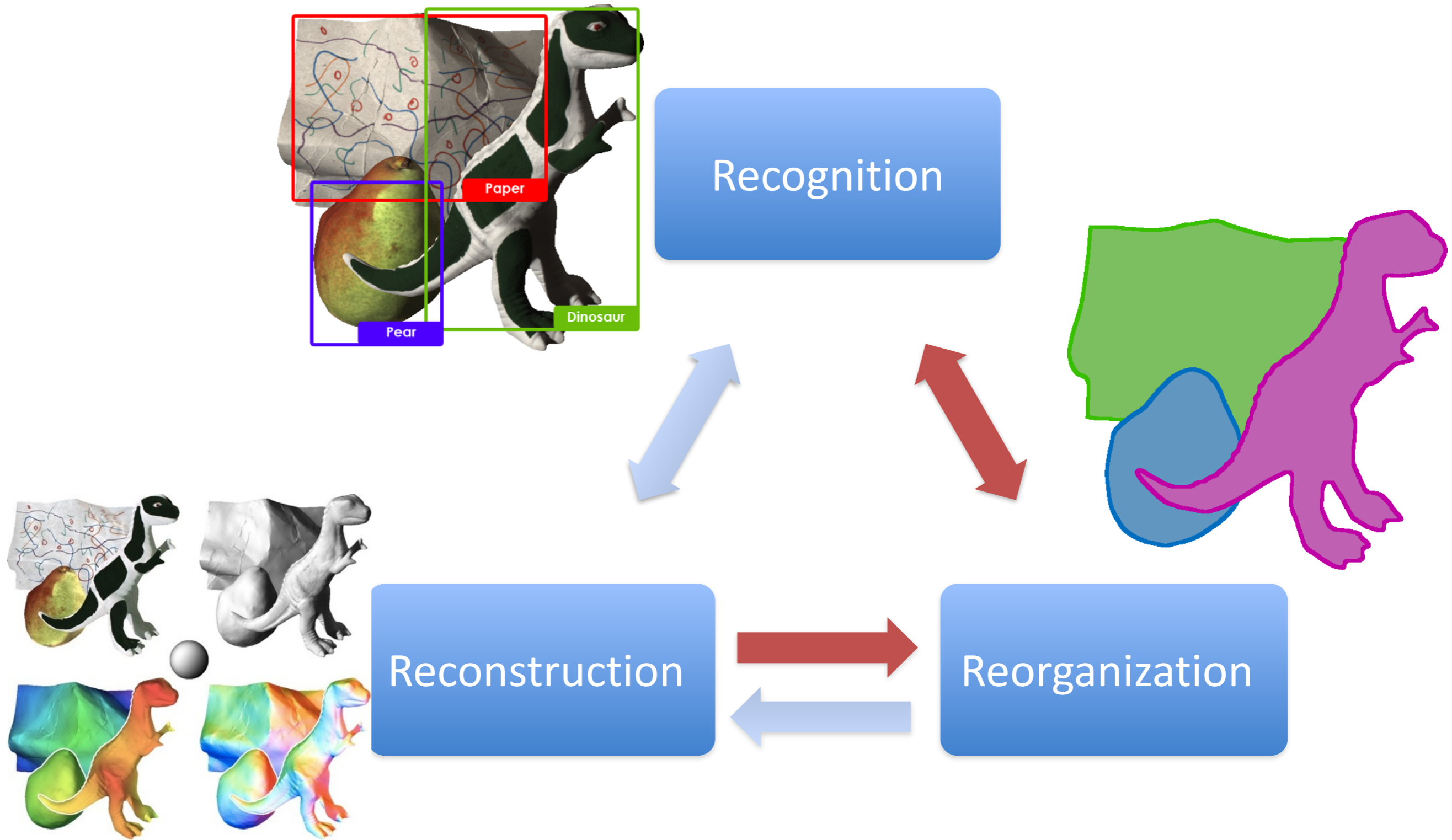
Instance Segm.

✓ State-of-the-art

# Recognition, Reconstruction & Reorganization

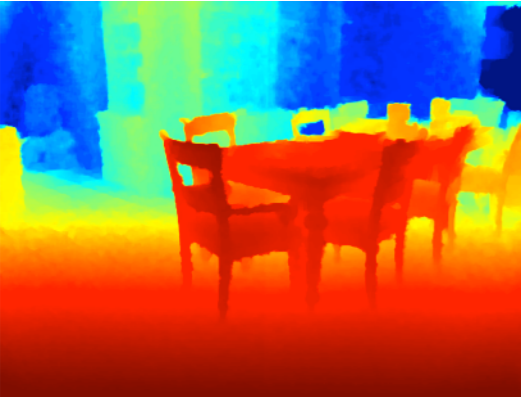


# Recognition, Reconstruction & Reorganization



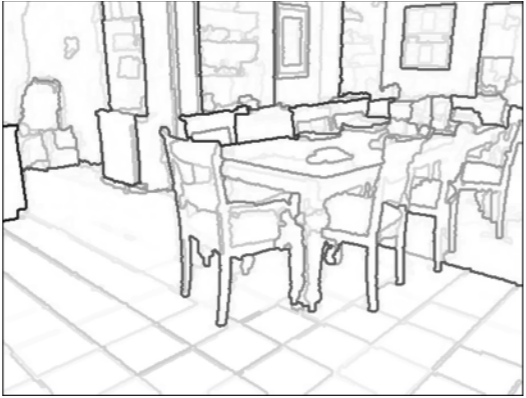
# Overview

## Input

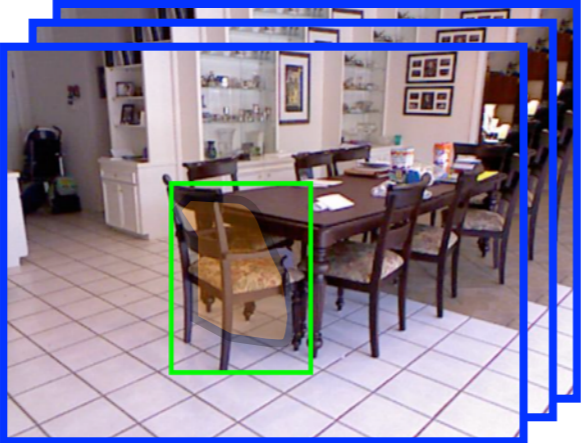


Color and Depth Image Pair

## Re-organization

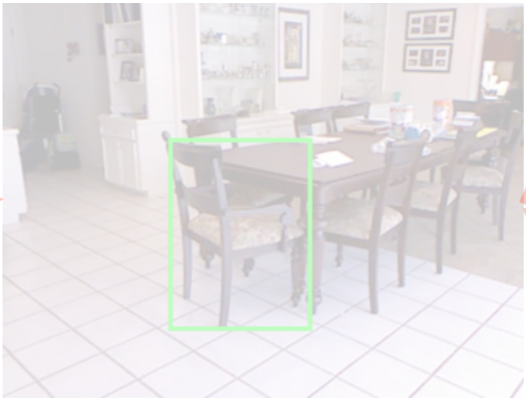


Contour Detection



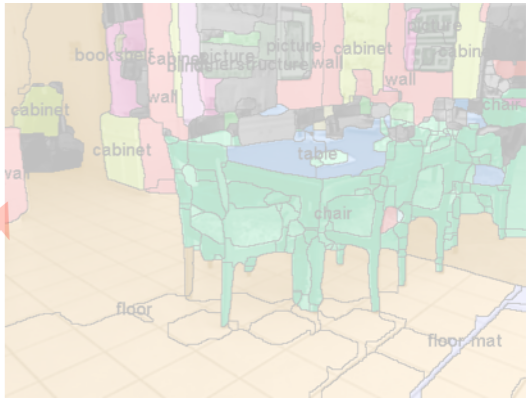
Region Proposal Generation

## Recognition

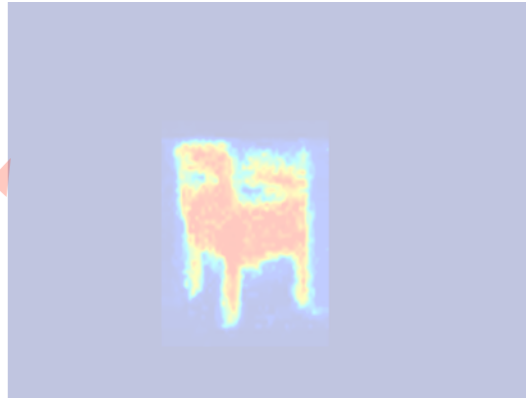


Object Detection

## Extensions



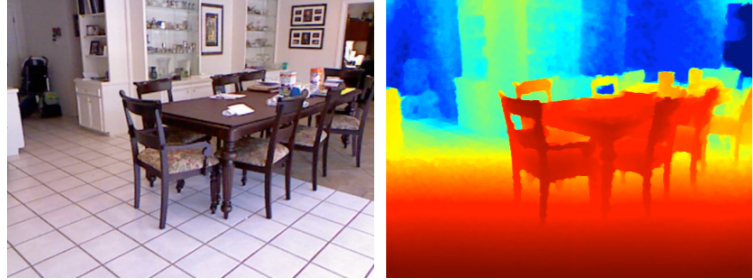
Semantic Segm.



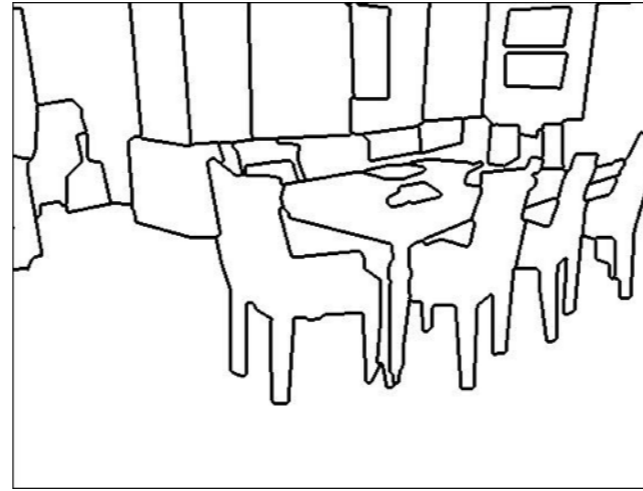
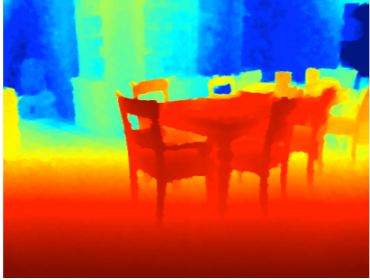
Instance Segm.

# Reorganization

# Reorganization



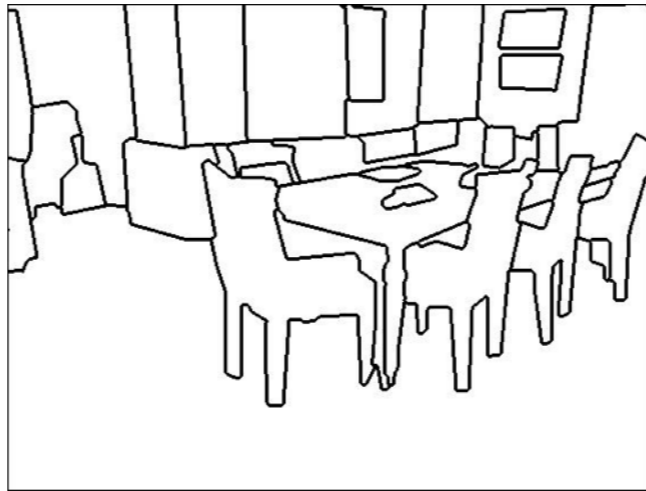
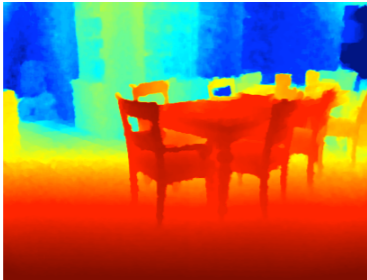
# Reorganization



Contour Detection



# Reorganization

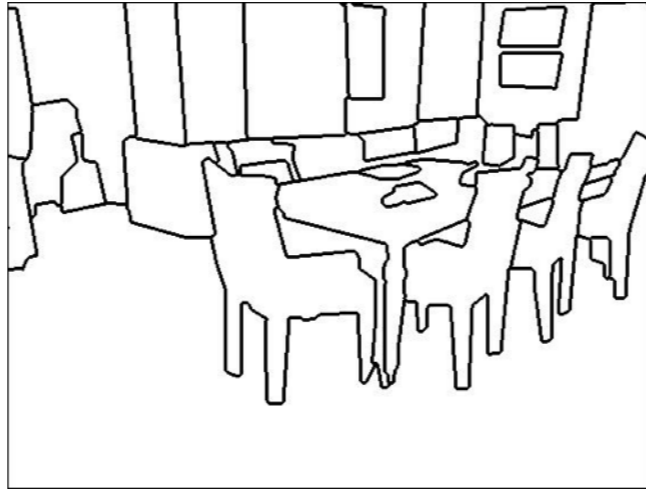
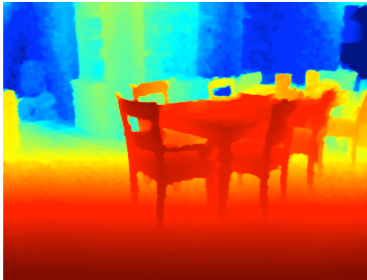


Contour Detection



Region Proposal

# Reorganization



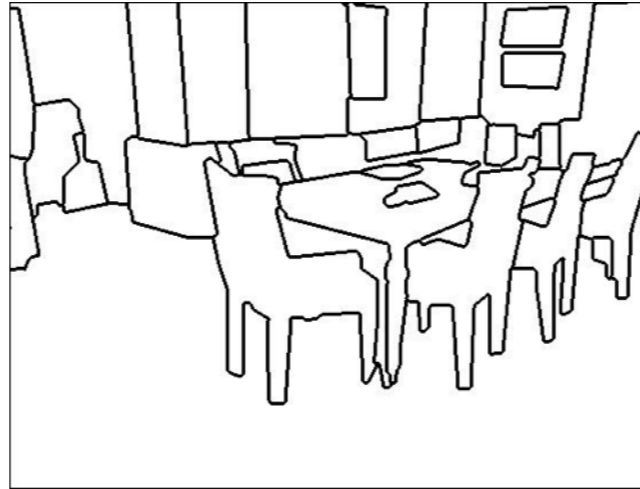
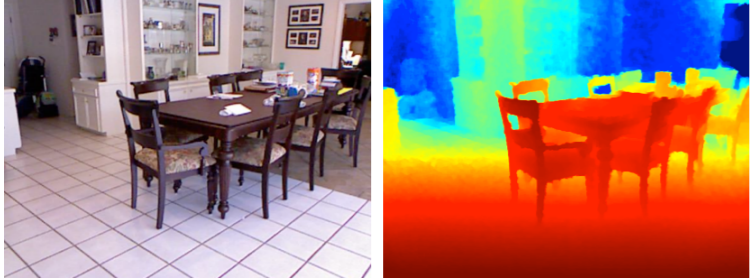
Contour Detection



Region Proposal

Related Work

# Reorganization



Contour Detection



Region Proposal

## Related Work

### RGB Contour Detection

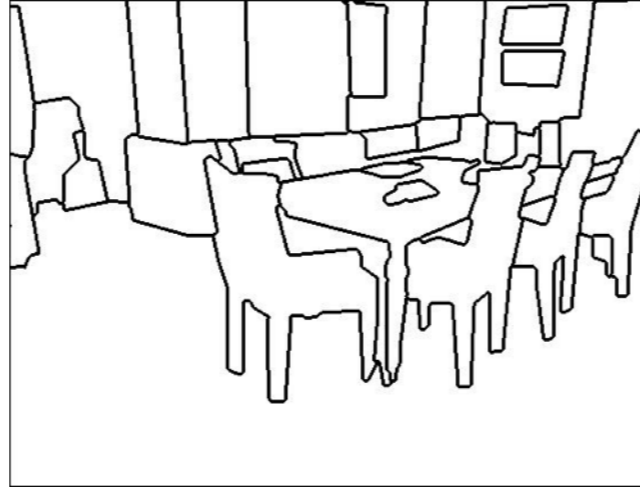
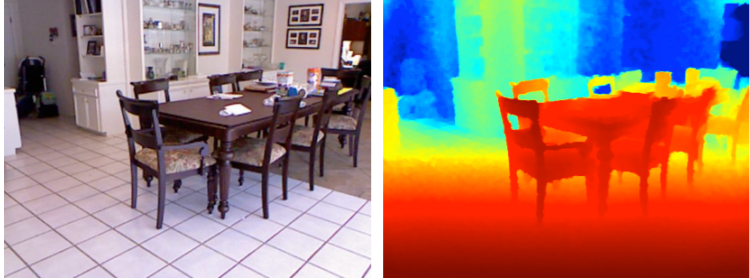
Martin et al. PAMI 2004 [Learning to Detect Natural Image Boundaries Using Local Brightness, Color, and Texture Cues](#)

Arbeláez et al. PAMI 2011 [Contour Detection and Hierarchical Image Segmentation](#)

Hoiem et al. IJCV 2007 [Recovering Surface Layout from an Image](#)

Dollar et al. ICCV 2013 [Structured Forests for Fast Edge Detection](#)

# Reorganization



Contour Detection



Region Proposal

## Related Work

### RGB Contour Detection

Martin et al. PAMI 2004 [Learning to Detect Natural Image Boundaries Using Local Brightness, Color, and Texture Cues](#)

Arbeláez et al. PAMI 2011 [Contour Detection and Hierarchical Image Segmentation](#)

Hoiem et al. IJCV 2007 [Recovering Surface Layout from an Image](#)

Dollar et al. ICCV 2013 [Structured Forests for Fast Edge Detection](#)

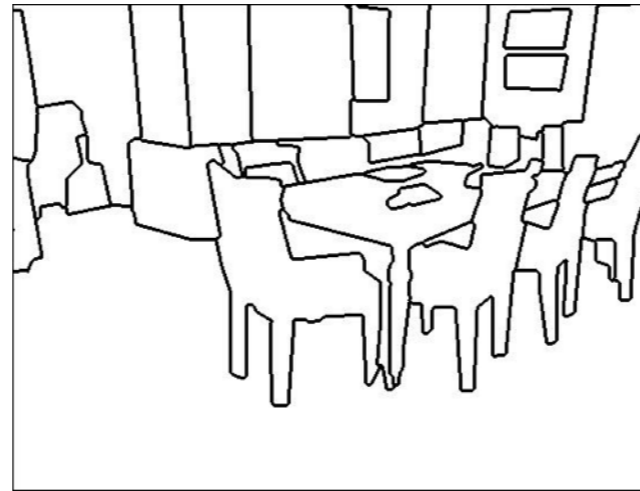
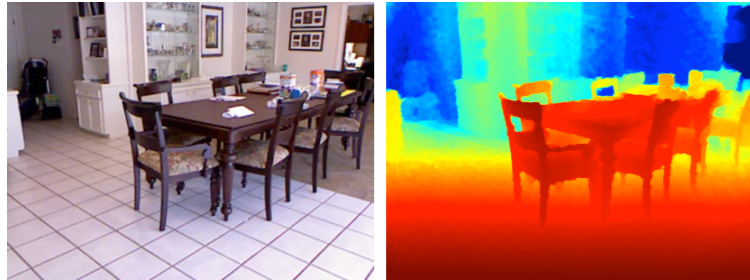
### RGB-D Contour Detection

Ren et al. NIPS 2012 [Discriminatively Trained Sparse Code Gradients for Contour Detection](#)

Ren et al. CVPR 2012 [RGB-\(D\) Scene Labeling: Features and Algorithms](#)

Silberman et al. ECCV 2012 [Indoor Segmentation and Support Inference from RGBD Images](#)

# Reorganization



Contour Detection



Region Proposal

## Related Work

### RGB Contour Detection

Martin et al. PAMI 2004 [Learning to Detect Natural Image Boundaries Using Local Brightness, Color, and Texture Cues](#)

Arbeláez et al. PAMI 2011 [Contour Detection and Hierarchical Image Segmentation](#)

Hoiem et al. IJCV 2007 [Recovering Surface Layout from an Image](#)

Dollar et al. ICCV 2013 [Structured Forests for Fast Edge Detection](#)

### RGB-D Contour Detection

Ren et al. NIPS 2012 [Discriminatively Trained Sparse Code Gradients for Contour Detection](#)

Ren et al. CVPR 2012 [RGB-\(D\) Scene Labeling: Features and Algorithms](#)

Silberman et al. ECCV 2012 [Indoor Segmentation and Support Inference from RGBD Images](#)

### Region/Box Proposal Generation

Carreira et al. PAMI 2012, [CPMC: Automatic Object Segmentation Using Constrained Parametric Min-Cuts.](#)

Uijlings et al. IJCV 2013, [Selective Search for Object Recognition](#)

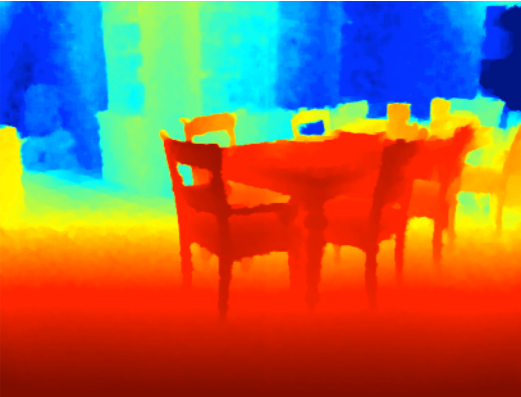
Lin et al. ICCV 2013 [Holistic Scene Understanding for 3D Object Detection with RGBD cameras](#)

Arbeláez et al. CVPR 2014, [Multiscale Combinatorial Grouping](#)

Zitnick et al. ECCV 2014, [Edge Boxes: Locating Object Proposals from Edges](#)

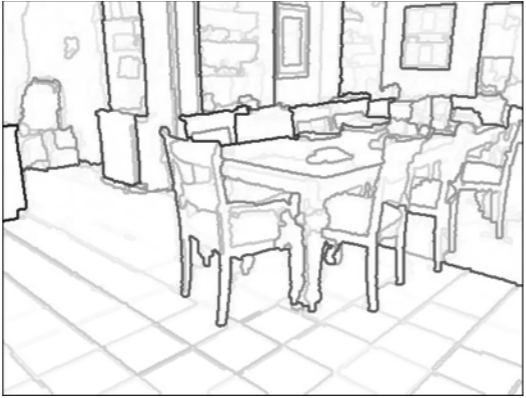
# Overview

## Input

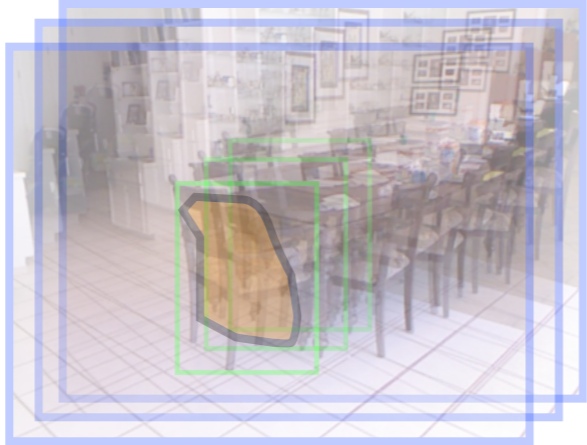


Color and Depth Image Pair

## Re-organization

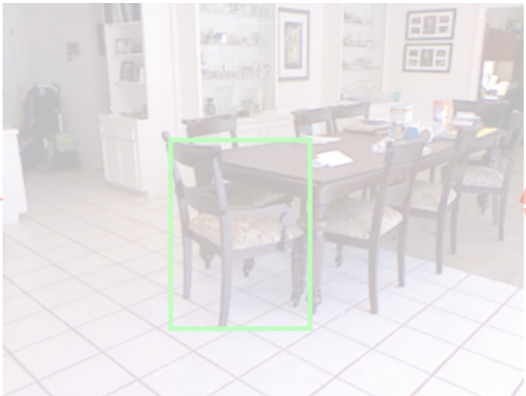


Contour Detection



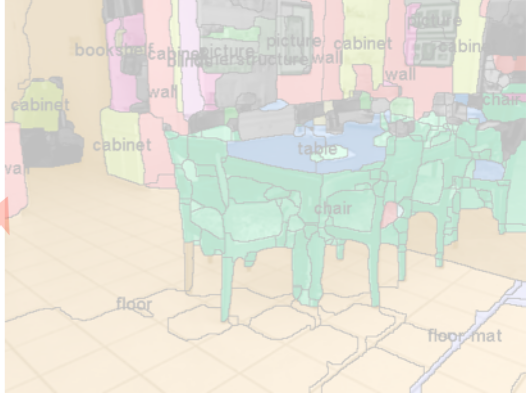
Region Proposal Generation

## Recognition

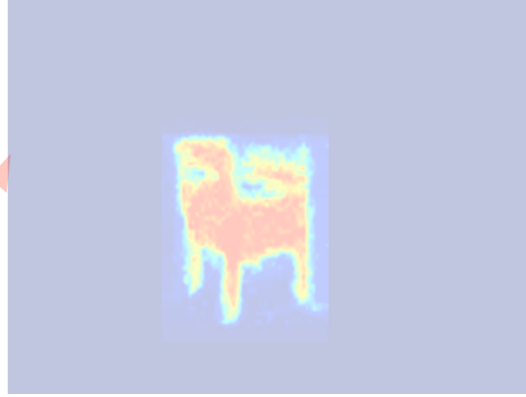


Object Detection

## Extensions



Semantic Segm.

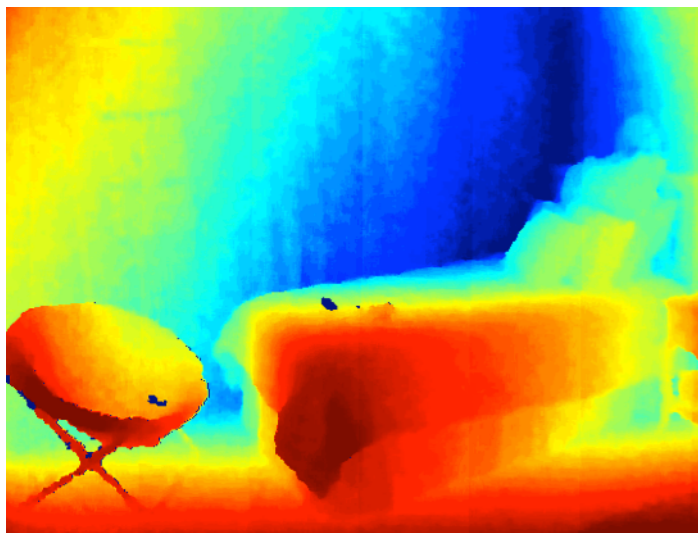
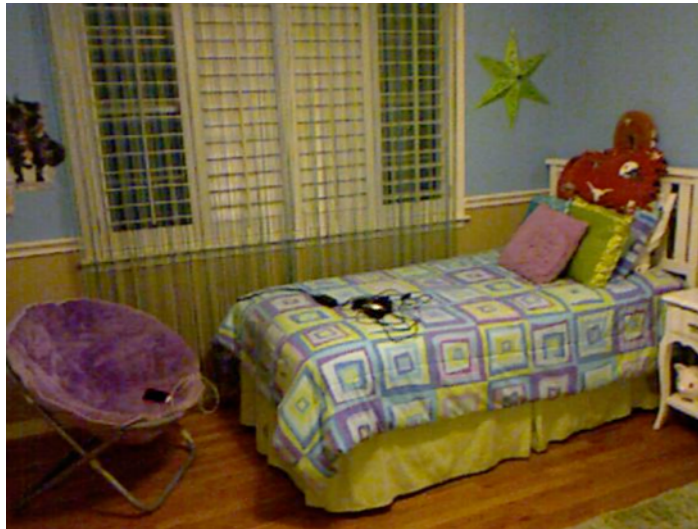


Instance Segm.

# Local Gradients on Depth Images



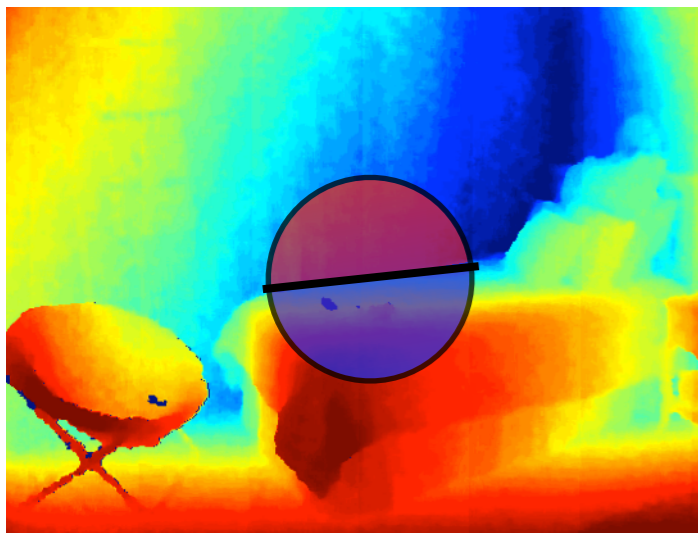
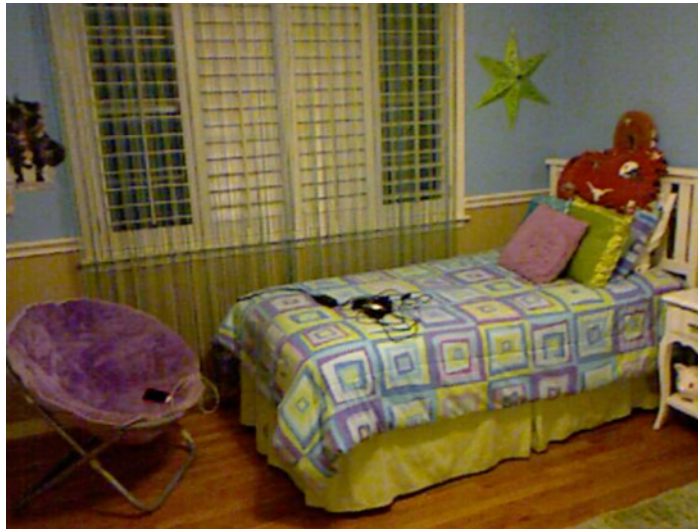
# Local Gradients on Depth Images



Input Depth Image

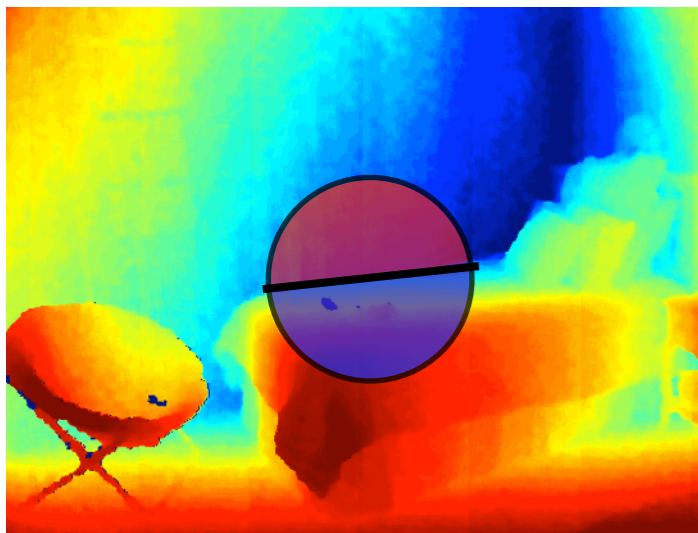
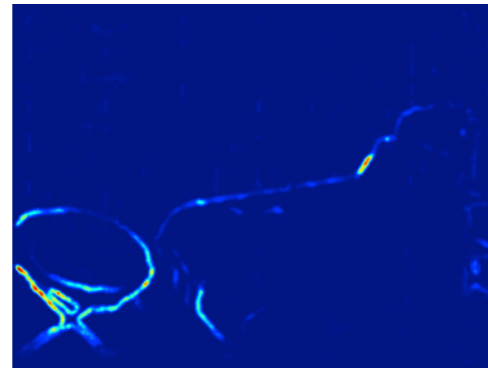
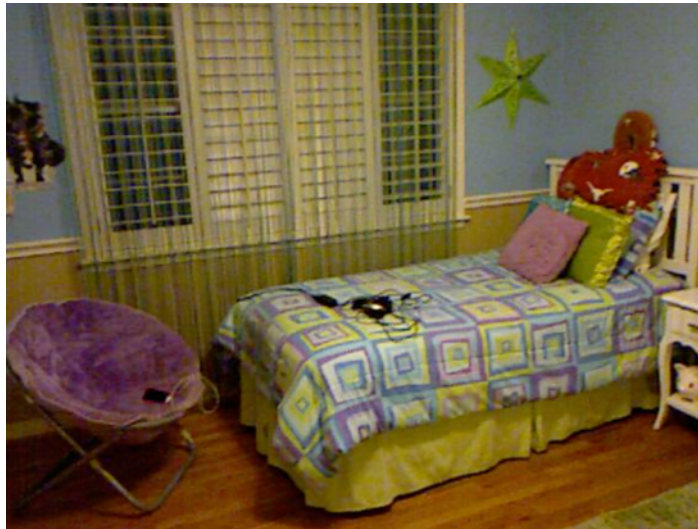


# Local Gradients on Depth Images

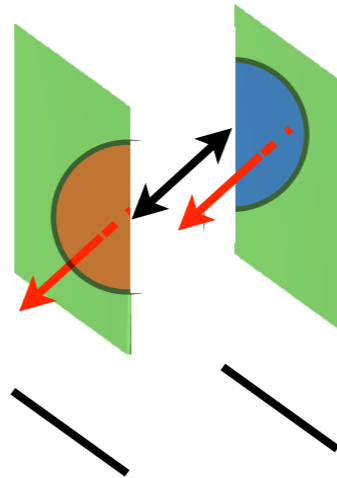


Input Depth Image

# Local Gradients on Depth Images

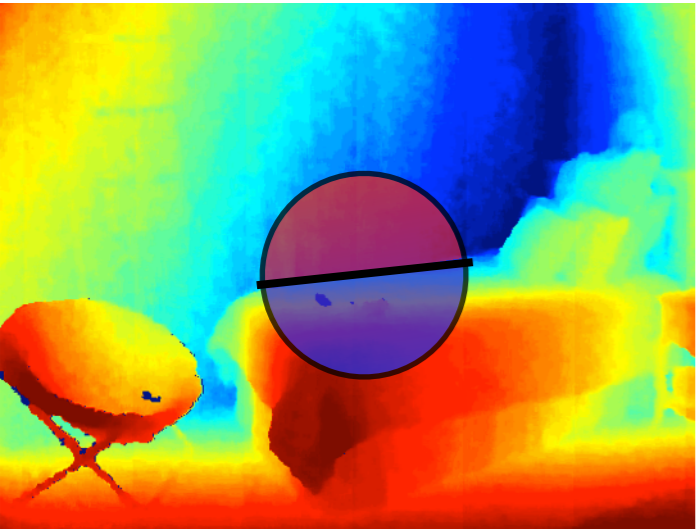
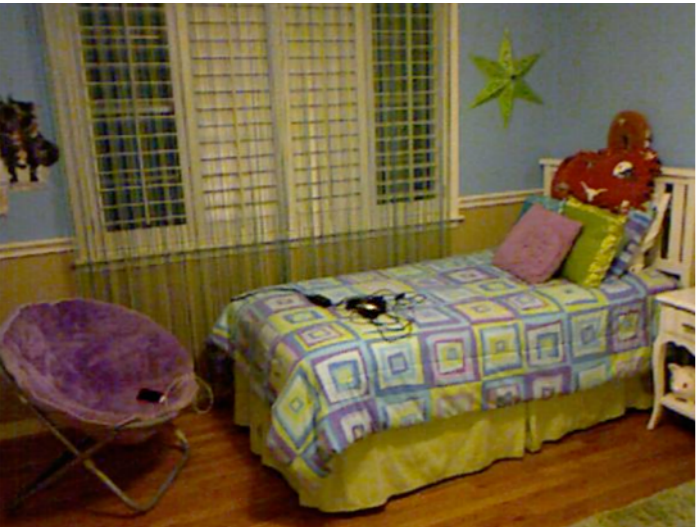


Input Depth Image

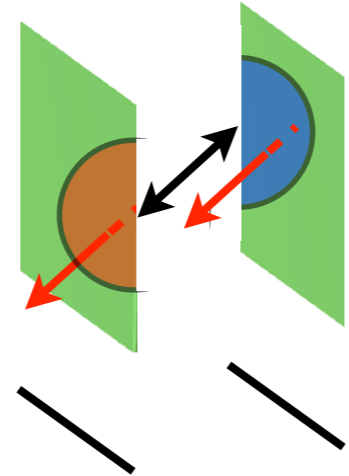
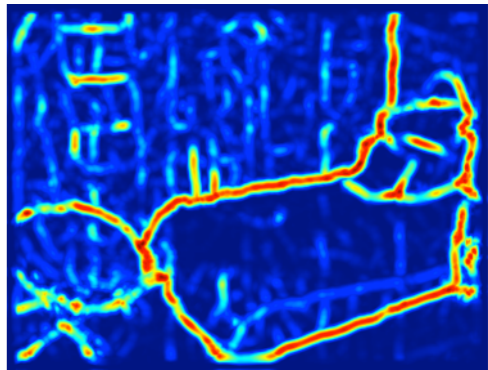
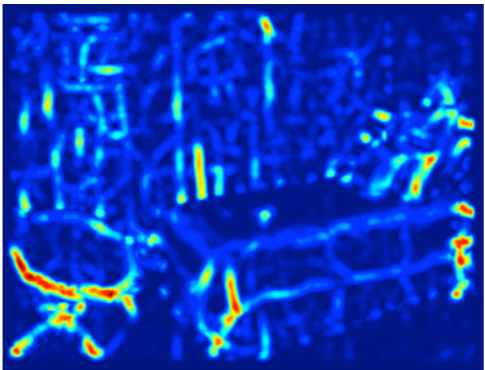
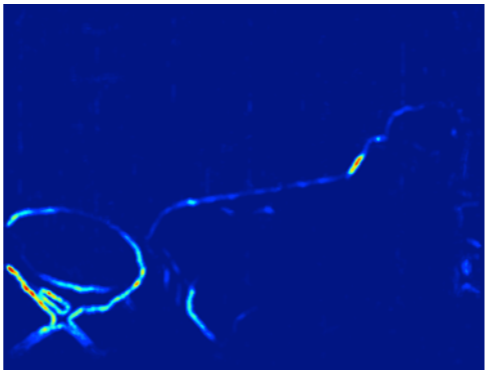


Depth Gradient,  
**DG**

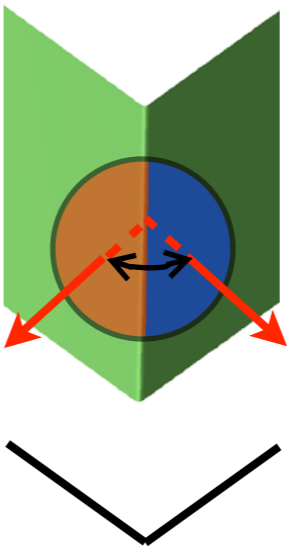
# Local Gradients on Depth Images



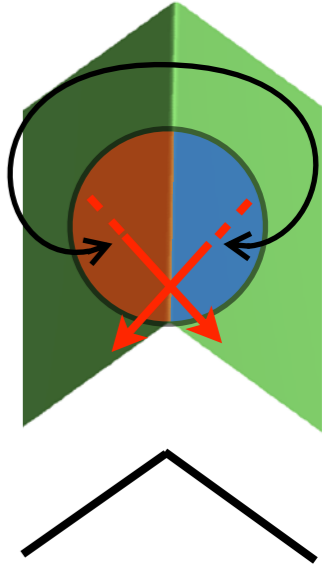
Input Depth Image



Depth Gradient,  
**DG**

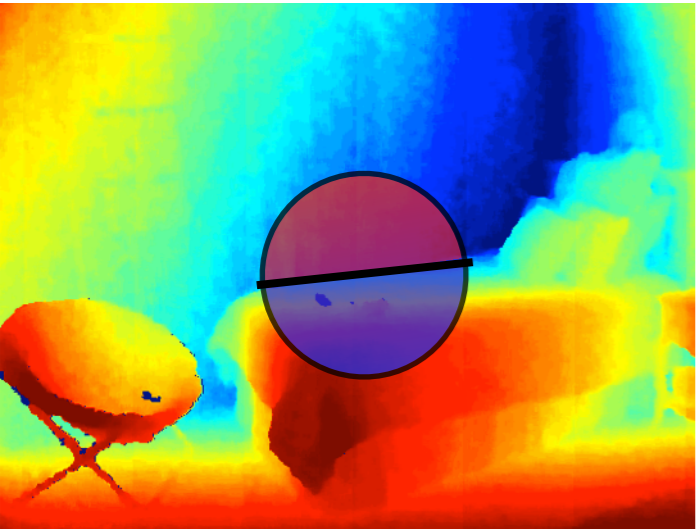
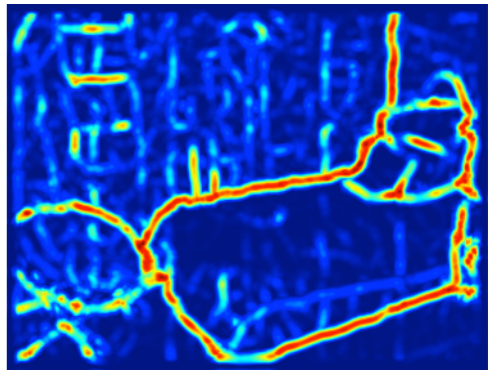
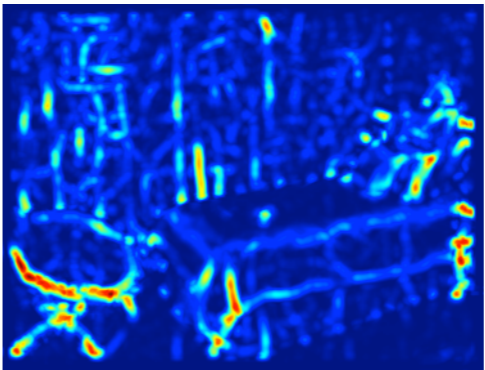
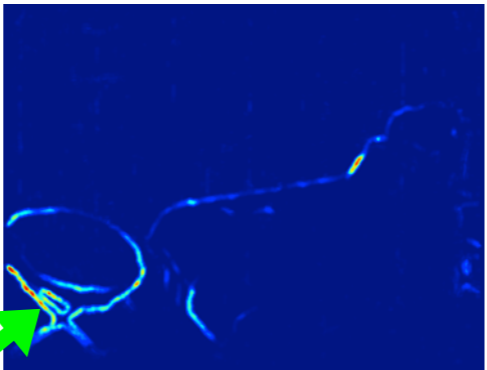
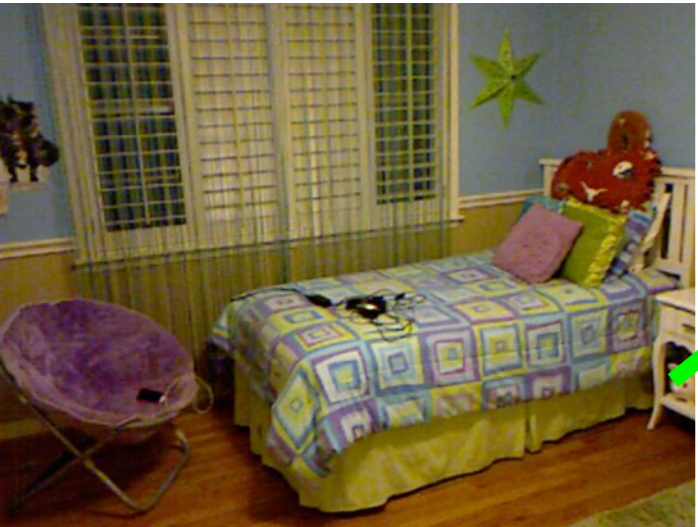


Convex  
Normal Gradient,  
**NG+**

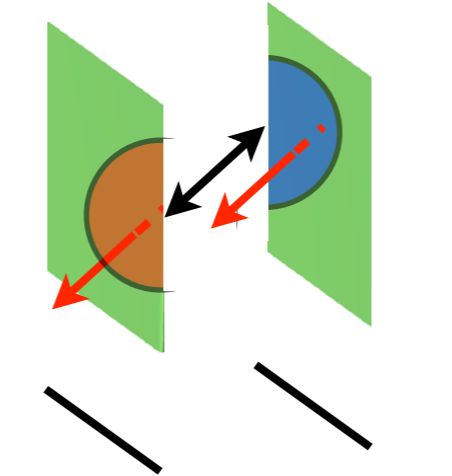


Concave  
Normal Gradient,  
**NG-**

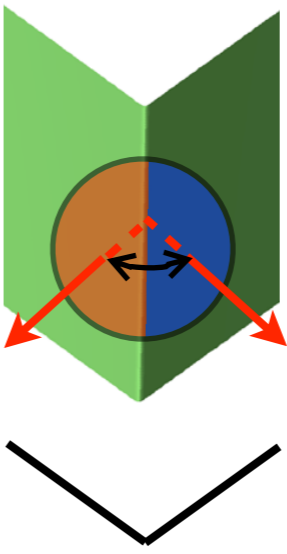
# Local Gradients on Depth Images



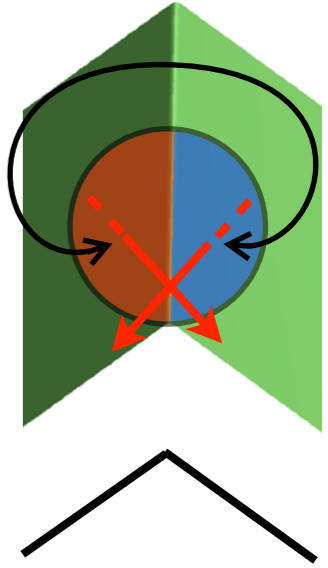
Input Depth Image



Depth Gradient,  
**DG**

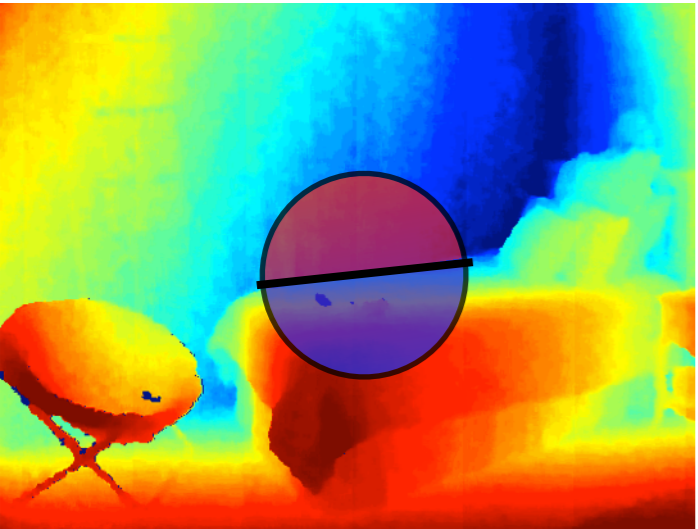
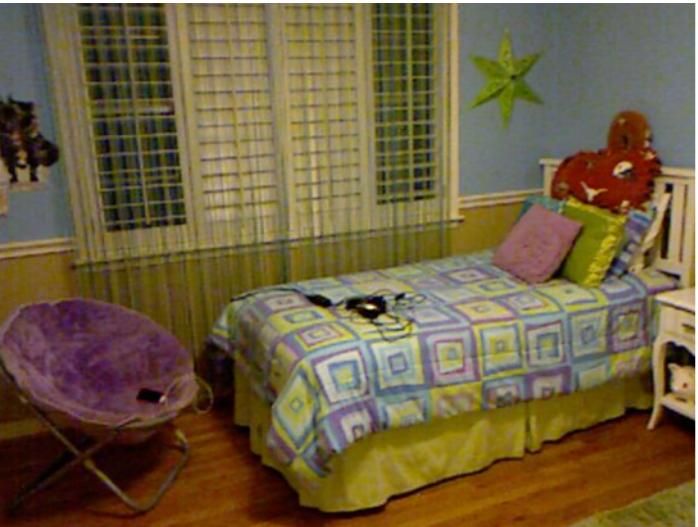


Convex  
Normal Gradient,  
**NG+**

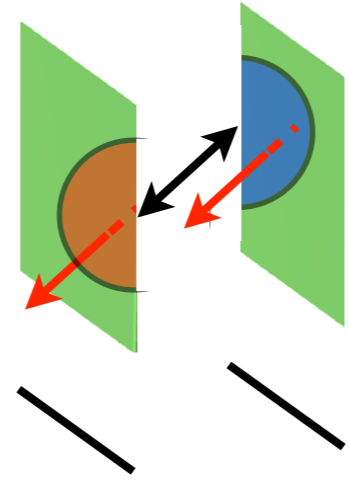
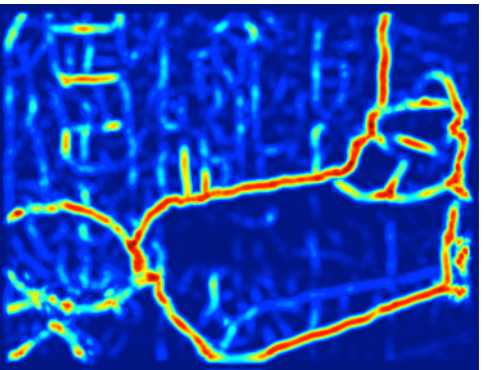
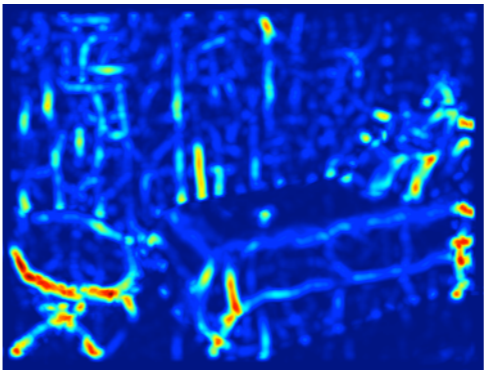
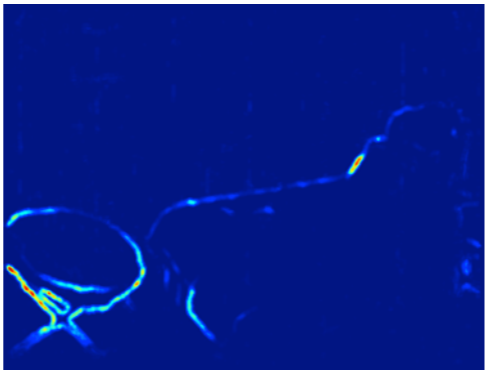


Concave  
Normal Gradient,  
**NG-**

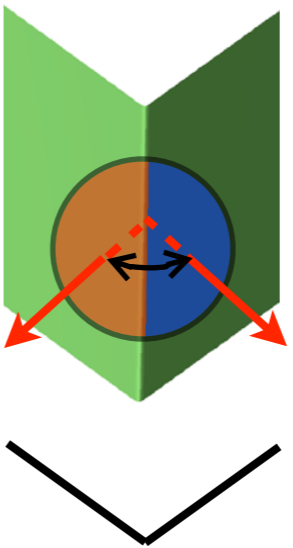
# Local Gradients on Depth Images



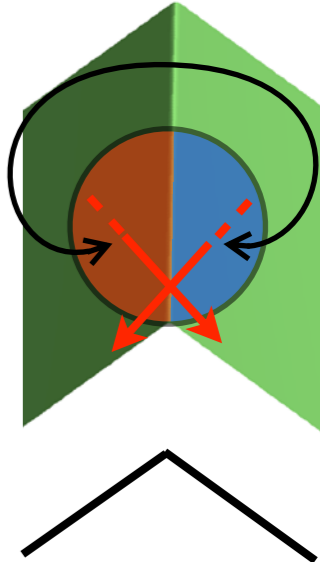
Input Depth Image



Depth Gradient,  
**DG**

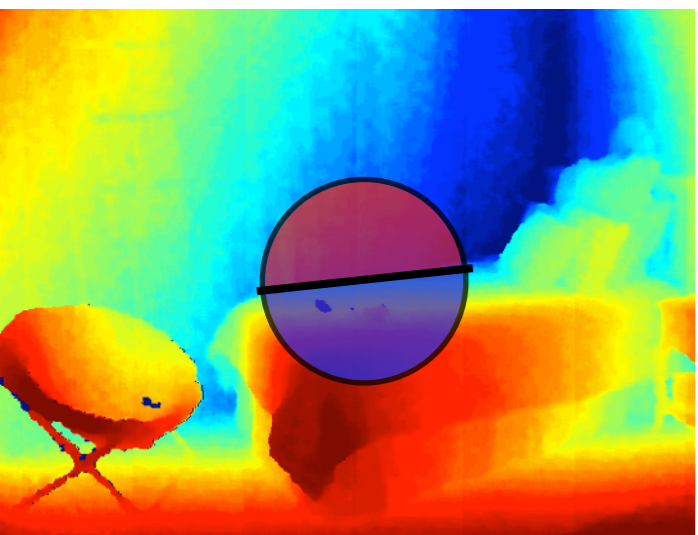
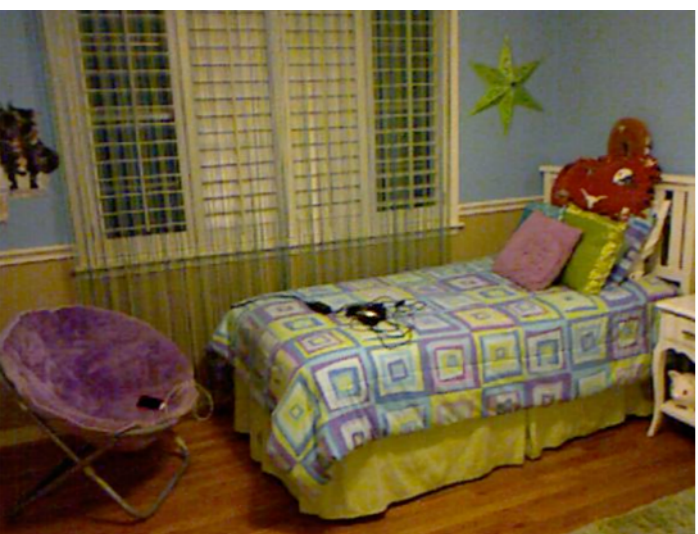


Convex  
Normal Gradient,  
**NG+**

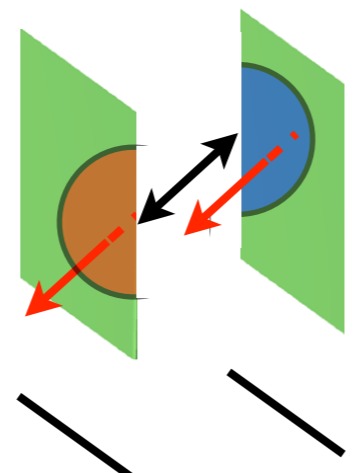
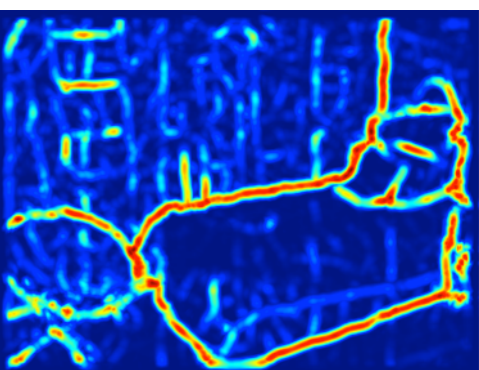
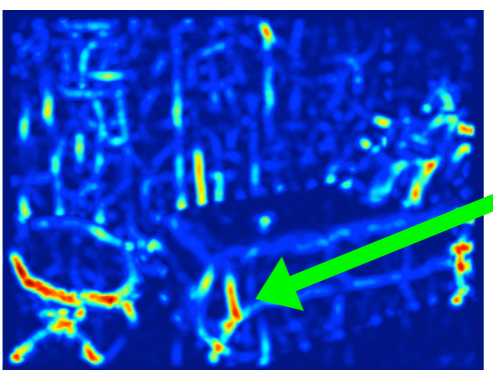
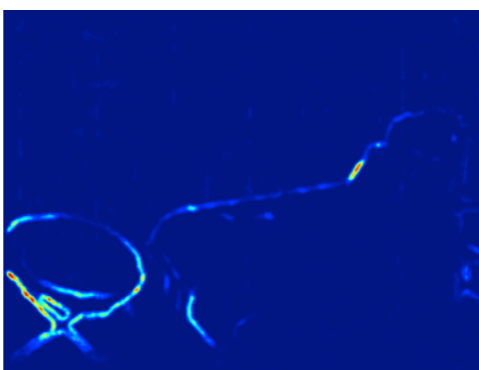


Concave  
Normal Gradient,  
**NG-**

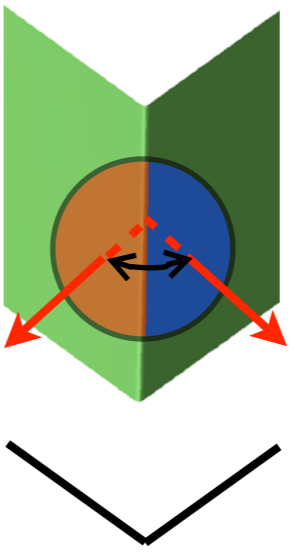
# Local Gradients on Depth Images



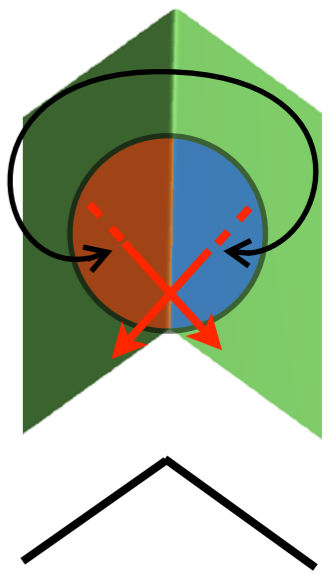
Input Depth Image



Depth Gradient,  
**DG**

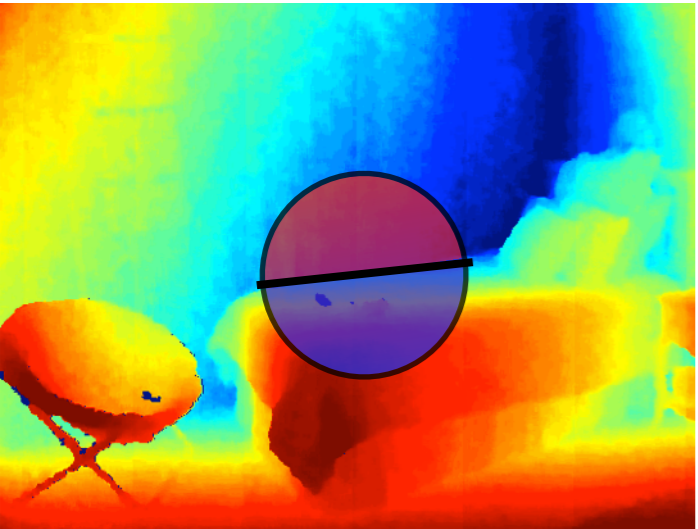
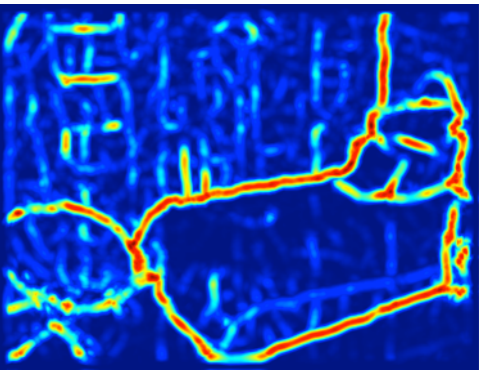
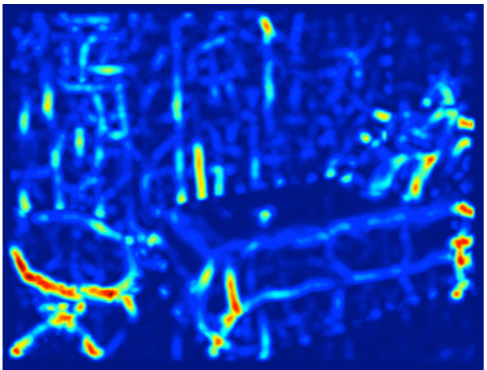
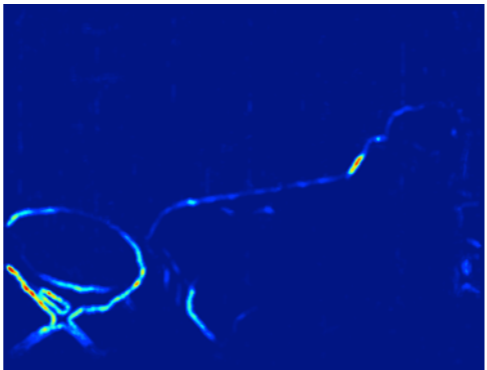
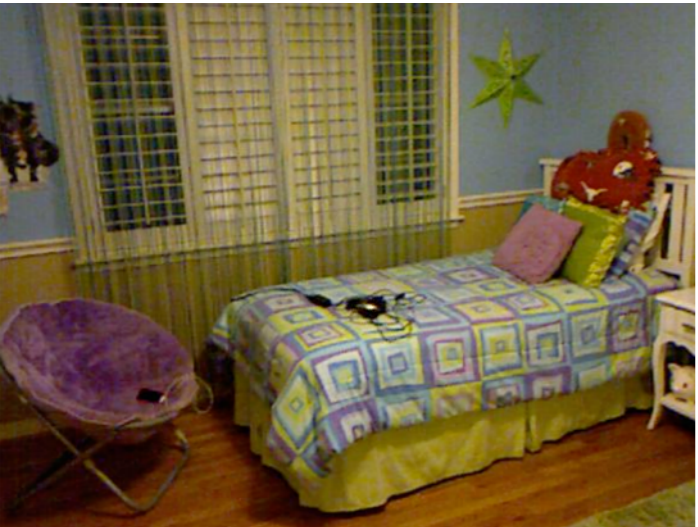


Convex  
Normal Gradient,  
**NG+**

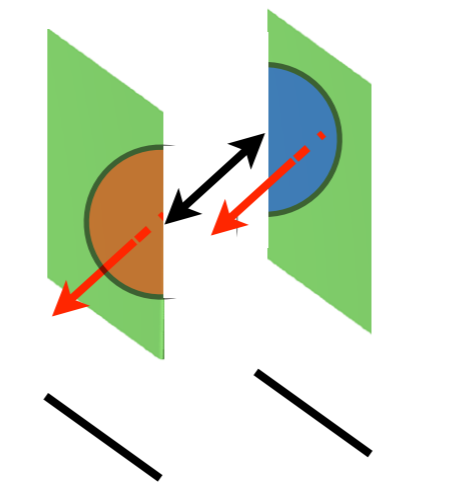


Concave  
Normal Gradient,  
**NG-**

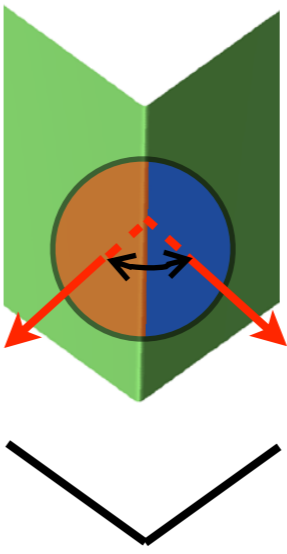
# Local Gradients on Depth Images



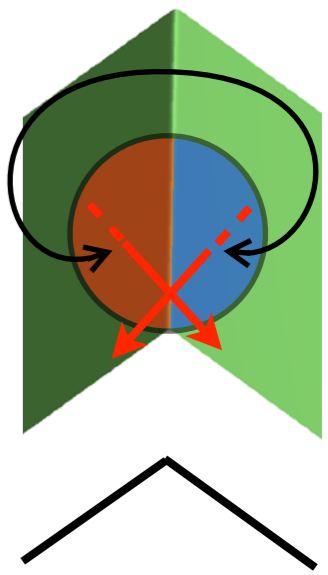
Input Depth Image



Depth Gradient,  
**DG**

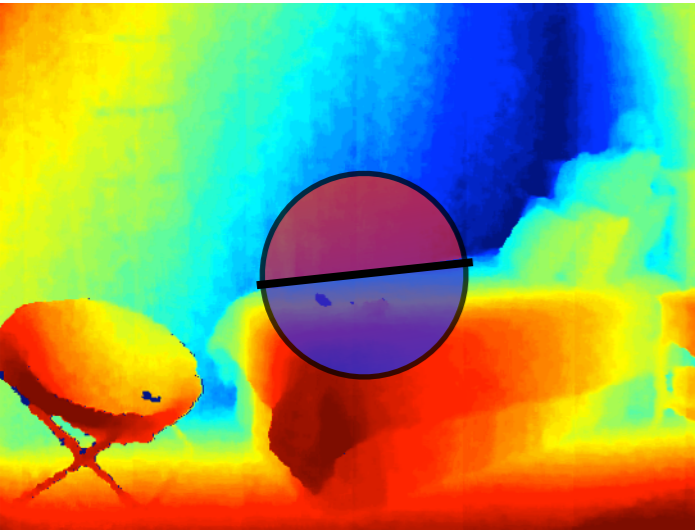


Convex  
Normal Gradient,  
**NG+**

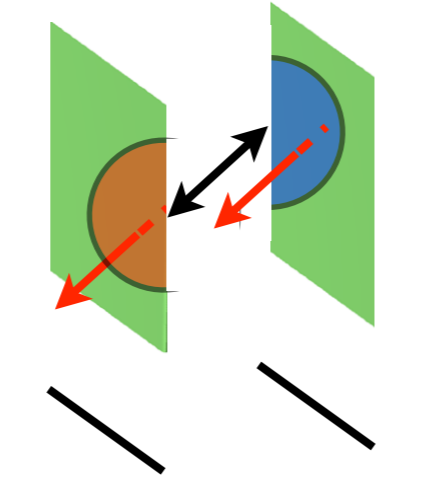
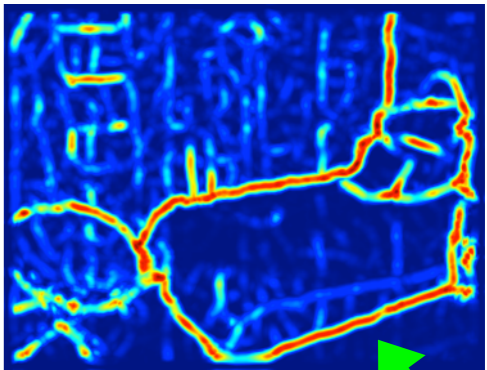
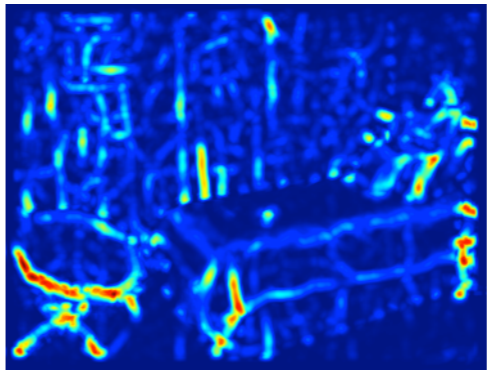
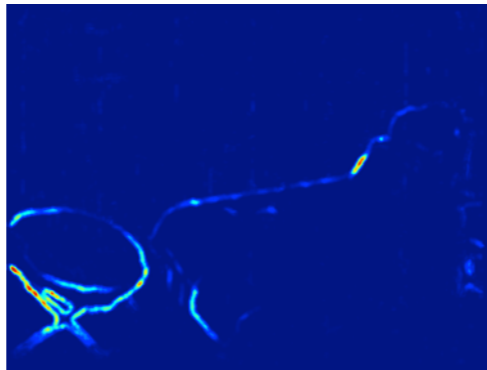


Concave  
Normal Gradient,  
**NG-**

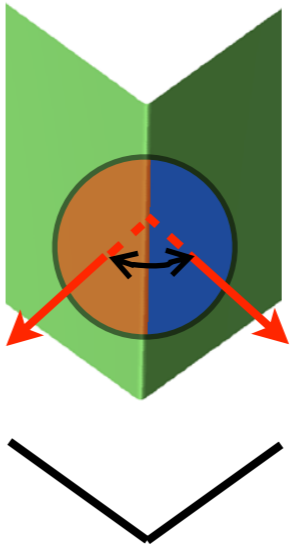
# Local Gradients on Depth Images



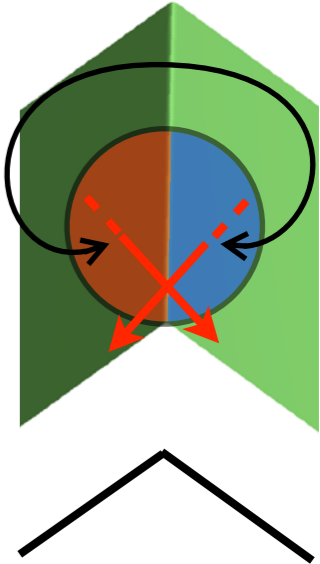
Input Depth Image



Depth Gradient,  
**DG**



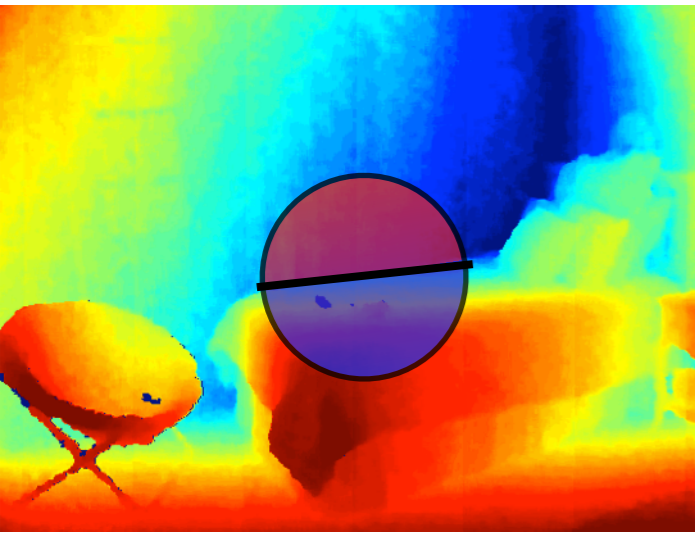
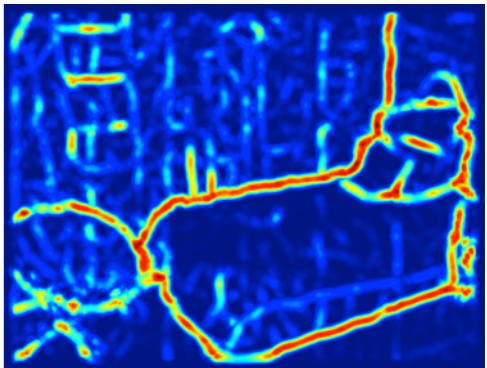
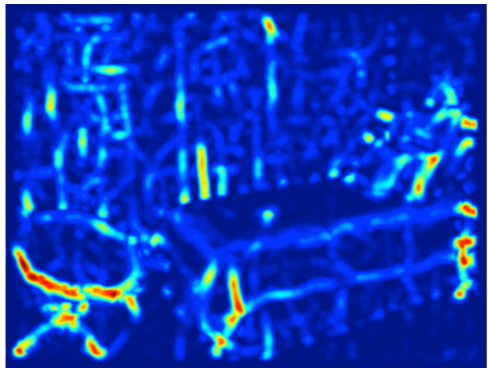
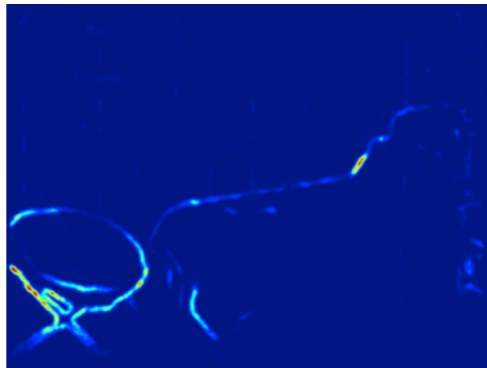
Convex  
Normal Gradient,  
**NG+**



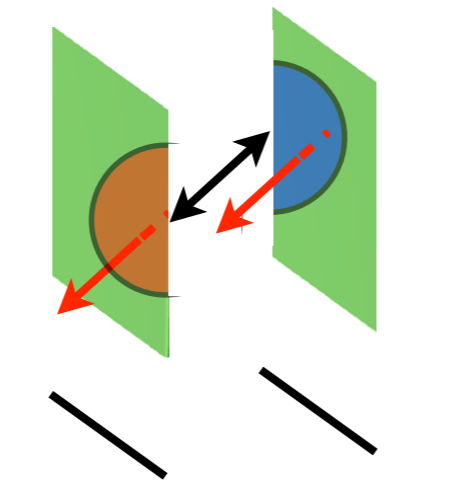
Concave  
Normal Gradient,  
**NG-**



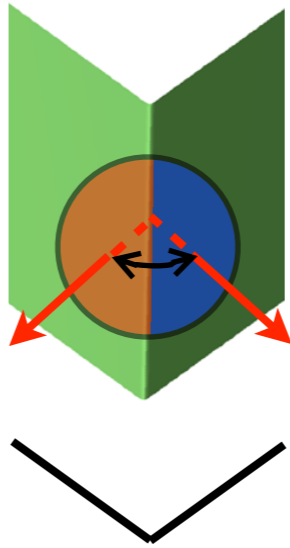
# Local Gradients on Depth Images



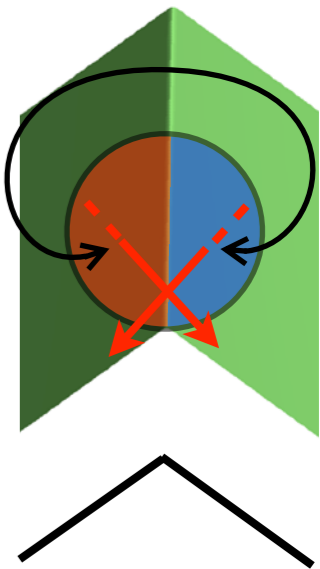
Input Depth Image



Depth Gradient,  
**DG**

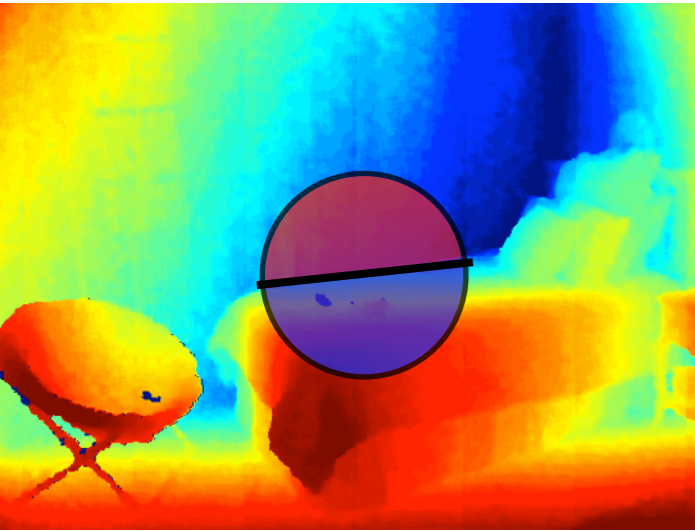


Convex  
Normal Gradient,  
**NG+**

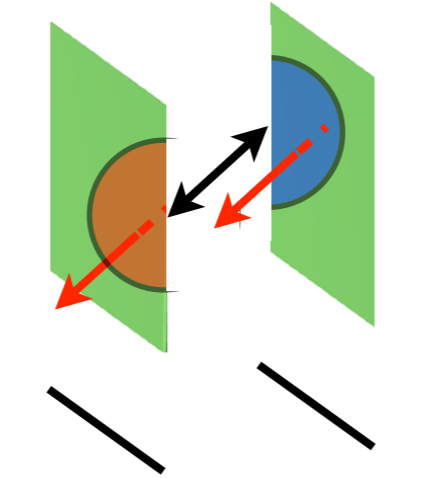
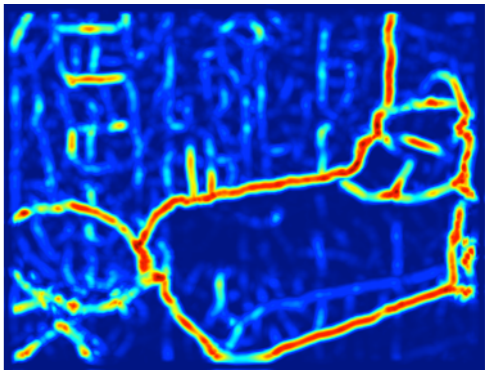
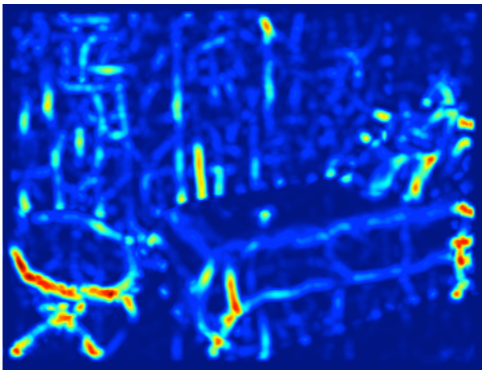
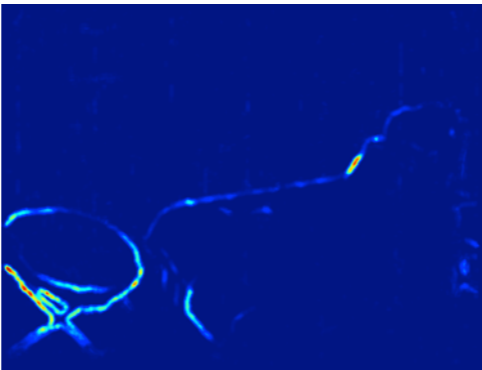


Concave  
Normal Gradient,  
**NG-**

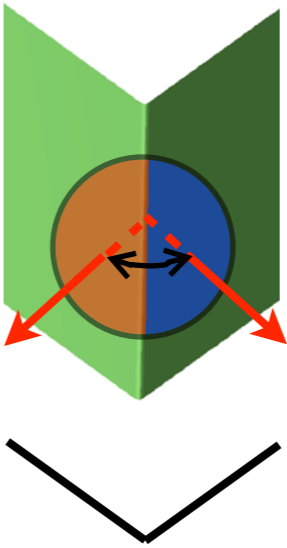
# Local Gradients on Depth Images



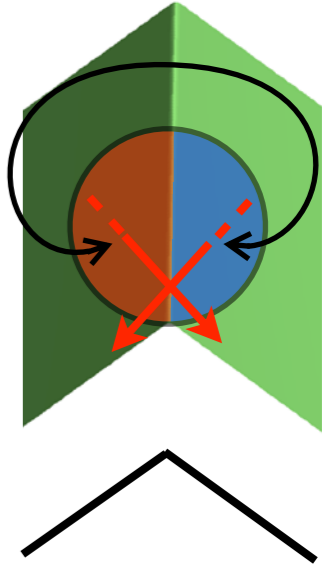
Input Depth Image



Depth Gradient,  
**DG**



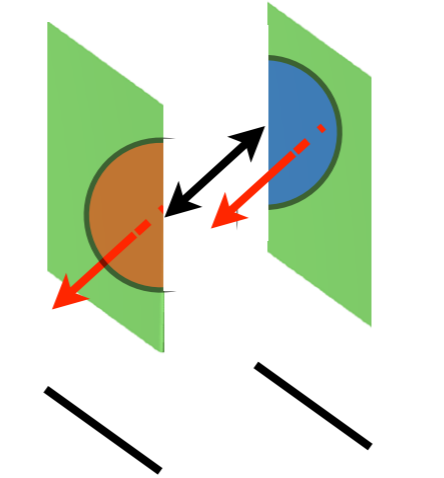
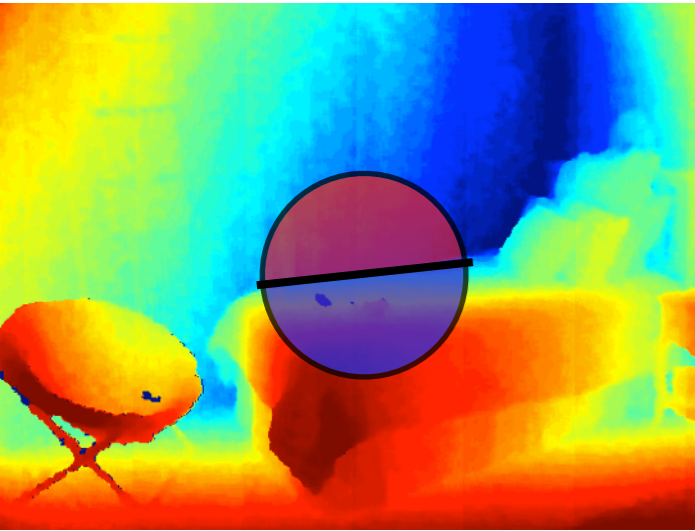
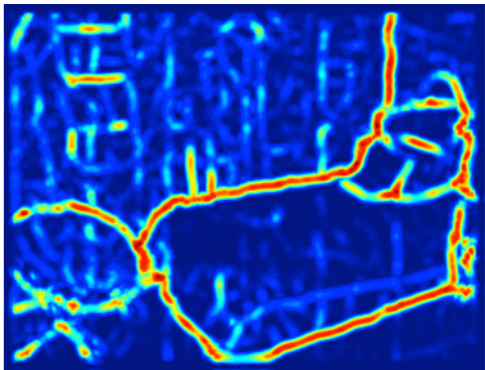
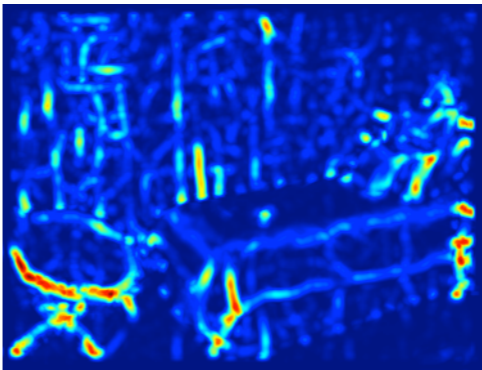
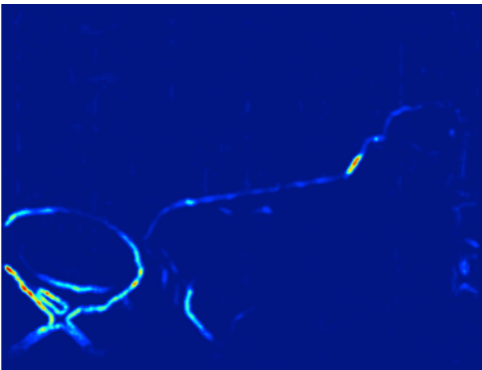
Convex  
Normal Gradient,  
**NG+**



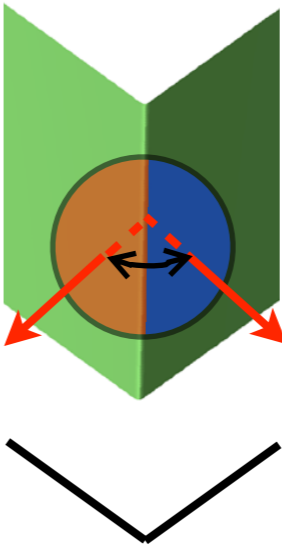
Concave  
Normal Gradient,  
**NG-**

Multi-scale Local Gradients from Depth Images

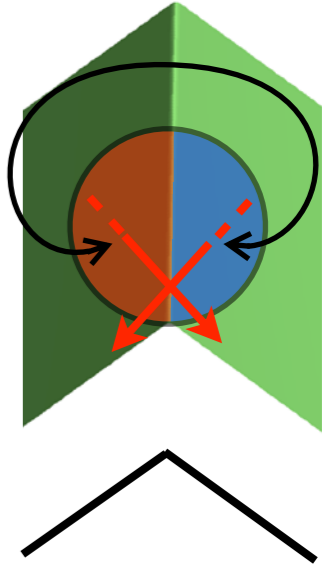
# Local Gradients on Depth Images



Depth Gradient,  
**DG**



Convex  
Normal Gradient,  
**NG+**



Concave  
Normal Gradient,  
**NG-**

Multi-scale Local Gradients from Depth Images

Important to differentiate between convex and concave normal gradients

# Using Local Gradients for Contour Detection

Arbeláez et al., PAMI 2011, Contour Detection and Hierarchical Image Segmentation.

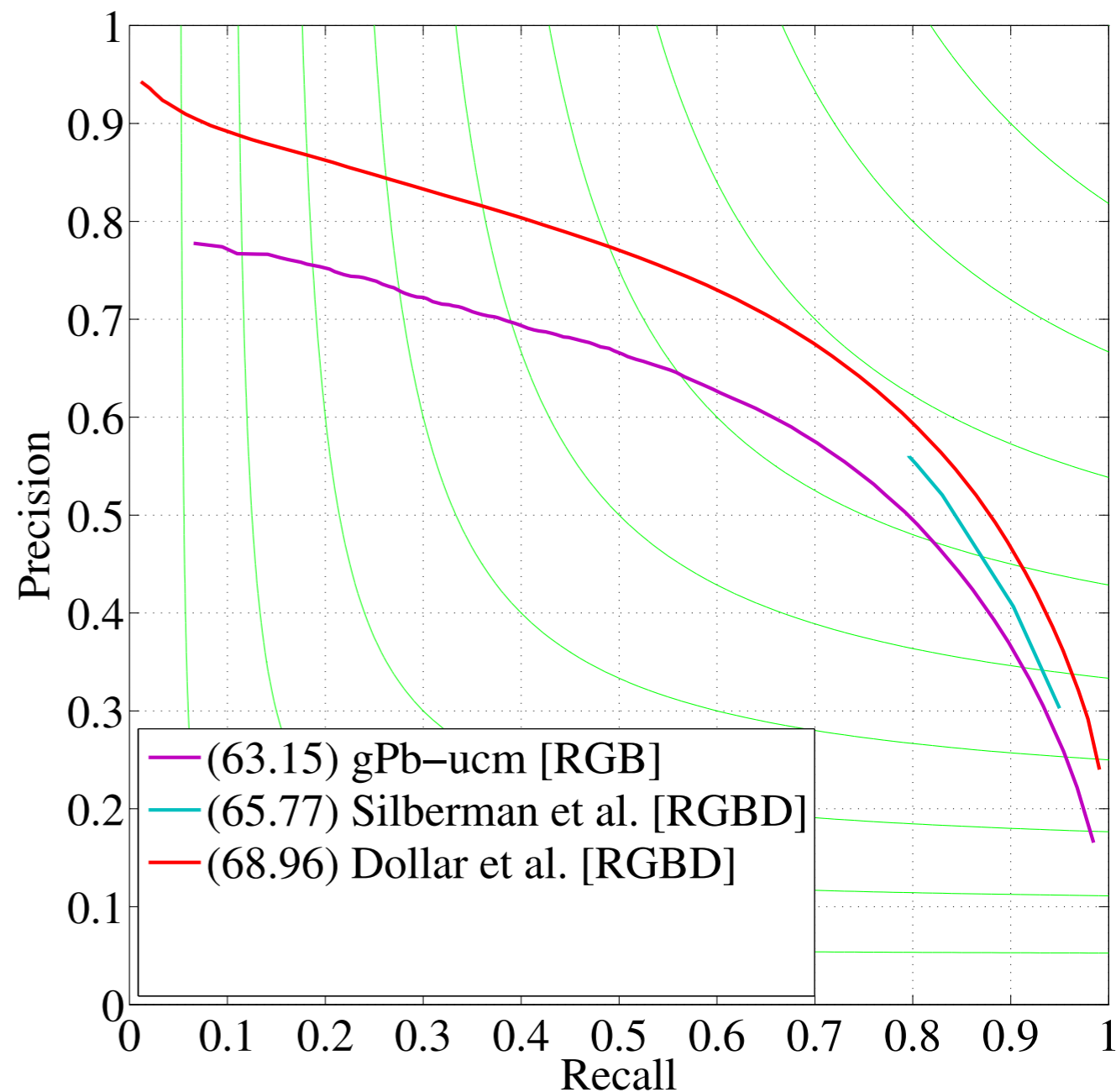
P. Dollar and L. Zitnick Structured Forests for fast edge detection, ICCV 2013

S. Gupta, P Arbeláez, J. Malik Perceptual Organization and Recognition in Indoor RGB-D Images, CVPR 2013

S. Gupta, R. Girshick, P Arbeláez, J. Malik , Object Detection and Segmentation using Semantically Rich Image and Depth Features In submission ECCV 2014

# Using Local Gradients for Contour Detection

Method		max F
gPb-UCM	RGB	63.15
Silberman et al.	RGB-D	65.77
Dollar et al.	RGB-D	68.96
Our (gPb-UCM + our cues)	RGB-D	68.66
Our (Dollar et al. + our cues++)	RGB-D	<b>70.36</b>



Arbeláez et al., PAMI 2011, Contour Detection and Hierarchical Image Segmentation.

P. Dollar and L. Zitnick Structured Forests for fast edge detection, ICCV 2013

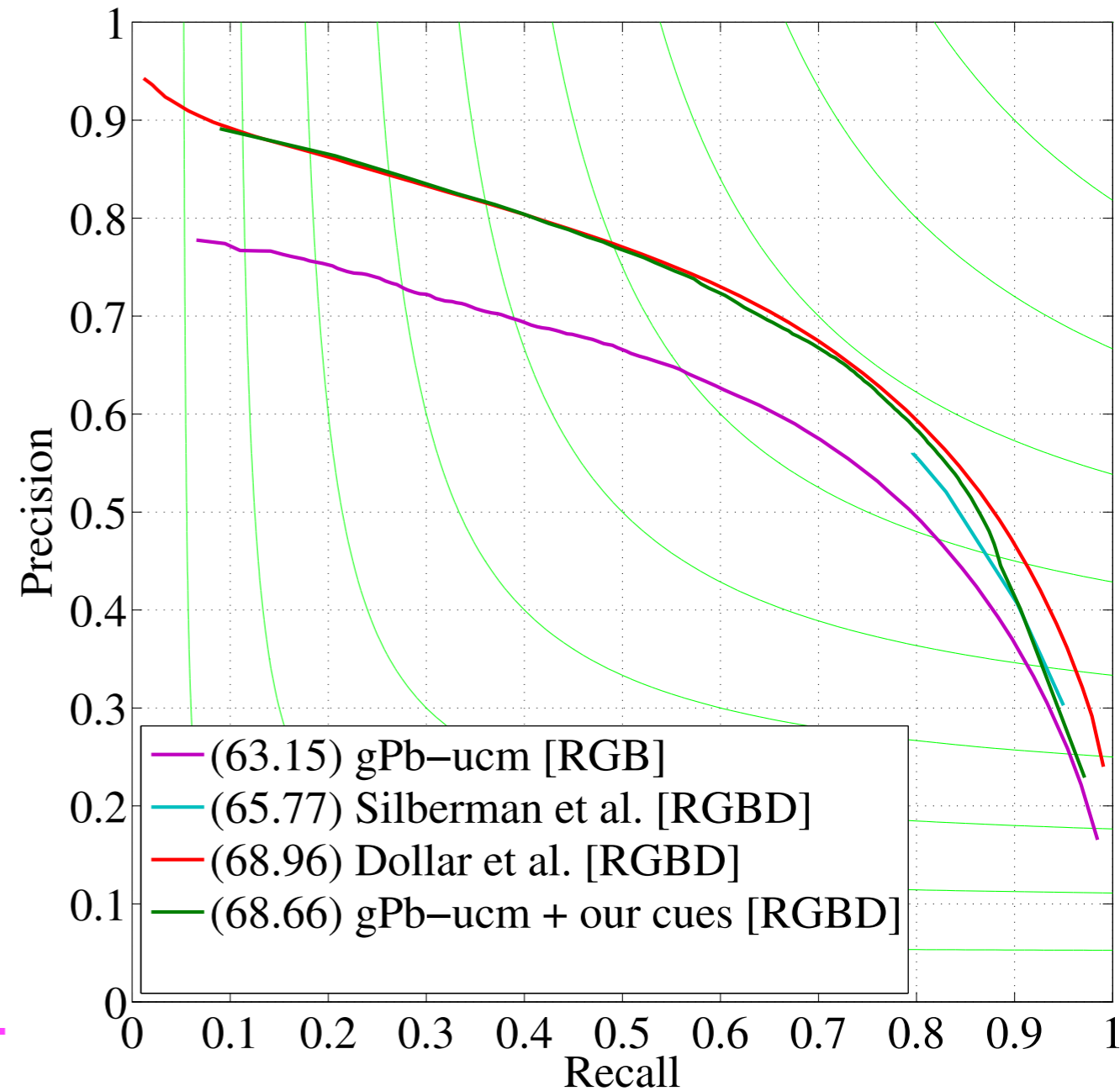
S. Gupta, P Arbeláez, J. Malik Perceptual Organization and Recognition in Indoor RGB-D Images, CVPR 2013

S. Gupta, R. Girshick, P Arbeláez, J. Malik , Object Detection and Segmentation using Semantically Rich Image and Depth Features In submission ECCV 2014

# Using Local Gradients for Contour Detection

Use with gPb-UCM

Method		max F
gPb-UCM	RGB	63.15
Silberman et al.	RGB-D	65.77
Dollar et al.	RGB-D	68.96
Our (gPb-UCM + our cues)	RGB-D	68.66
Our (Dollar et al. + our cues++)	RGB-D	<b>70.36</b>



Arbeláez et al., PAMI 2011, Contour Detection and Hierarchical Image Segmentation.

P. Dollar and L. Zitnick Structured Forests for fast edge detection, ICCV 2013

S. Gupta, P Arbeláez, J. Malik Perceptual Organization and Recognition in Indoor RGB-D Images, CVPR 2013

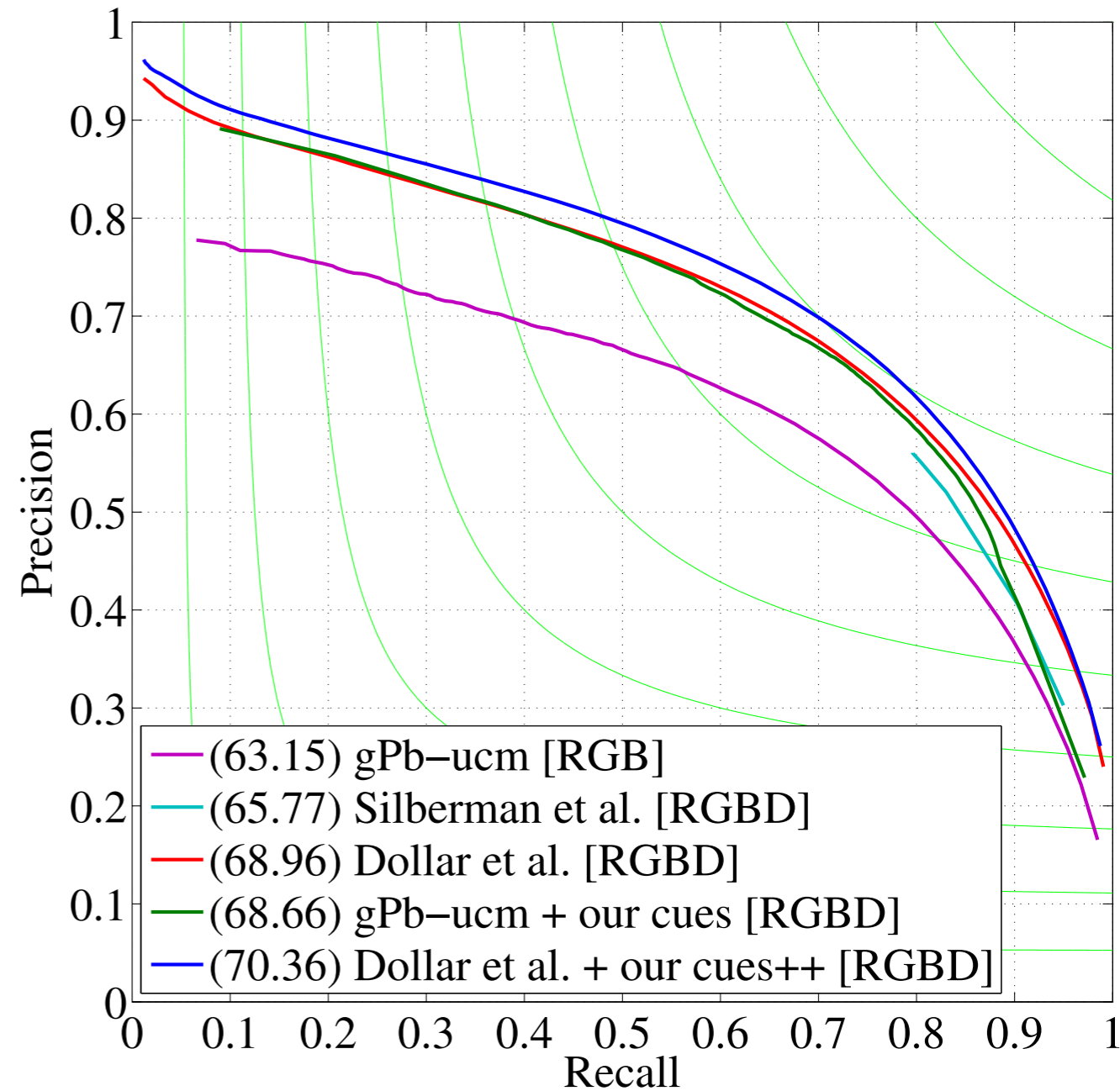
S. Gupta, R. Girshick, P Arbeláez, J. Malik, Object Detection and Segmentation using Semantically Rich Image and Depth Features In submission ECCV 2014

# Using Local Gradients for Contour Detection

Use with gPb-UCM

Use with Dollar et al.'s structured edges

Method		max F
gPb-UCM	RGB	63.15
Silberman et al.	RGB-D	65.77
Dollar et al.	RGB-D	68.96
Our (gPb-UCM + our cues)	RGB-D	68.66
Our (Dollar et al. + our cues++)	RGB-D	<b>70.36</b>



Arbeláez et al., PAMI 2011, Contour Detection and Hierarchical Image Segmentation.

P. Dollar and L. Zitnick Structured Forests for fast edge detection, ICCV 2013

S. Gupta, P Arbeláez, J. Malik Perceptual Organization and Recognition in Indoor RGB-D Images, CVPR 2013

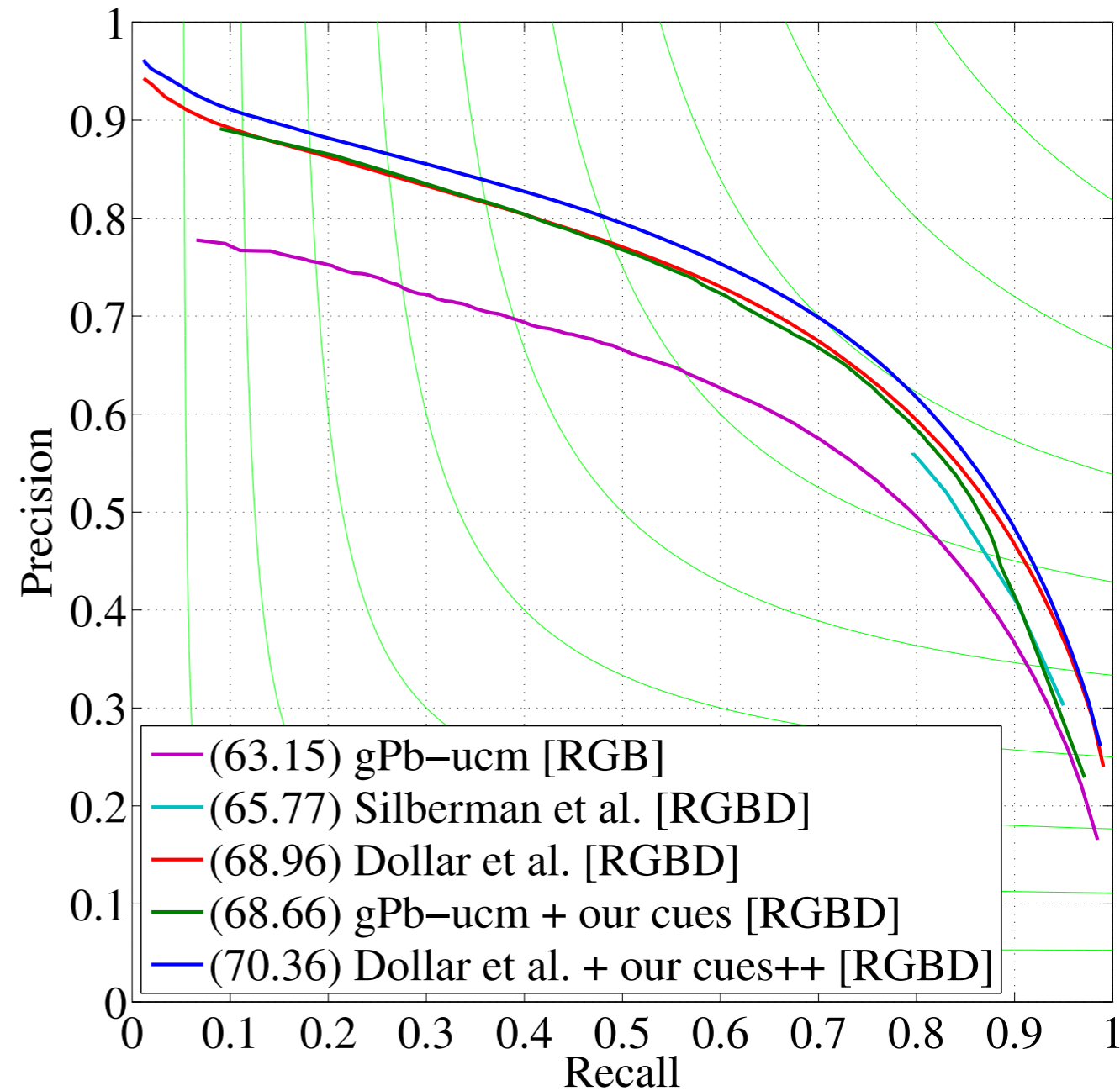
S. Gupta, R. Girshick, P Arbeláez, J. Malik, Object Detection and Segmentation using Semantically Rich Image and Depth Features In submission ECCV 2014

# Using Local Gradients for Contour Detection

Use with gPb-UCM

Use with Dollar et al.'s structured edges

Method		max F
gPb-UCM	RGB	63.15
Silberman et al.	RGB-D	65.77
Dollar et al.	RGB-D	68.96
<b>Our (gPb-UCM + our cues)</b>	RGB-D	68.66
<b>Our (Dollar et al. + our cues++)</b>	RGB-D	<b>70.36</b>



Arbeláez et al., PAMI 2011, Contour Detection and Hierarchical Image Segmentation.

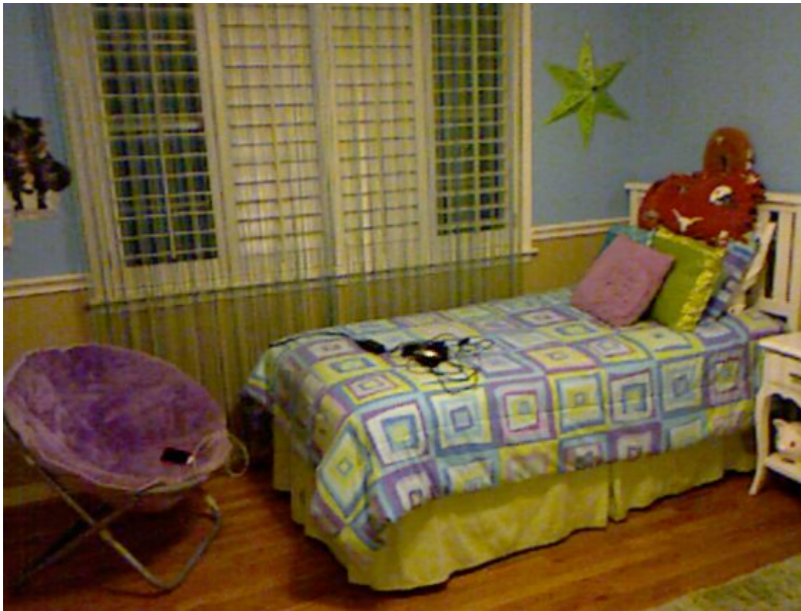
P. Dollar and L. Zitnick Structured Forests for fast edge detection, ICCV 2013

S. Gupta, P Arbeláez, J. Malik Perceptual Organization and Recognition in Indoor RGB-D Images, CVPR 2013

S. Gupta, R. Girshick, P Arbeláez, J. Malik, Object Detection and Segmentation using Semantically Rich Image and Depth Features In submission ECCV 2014



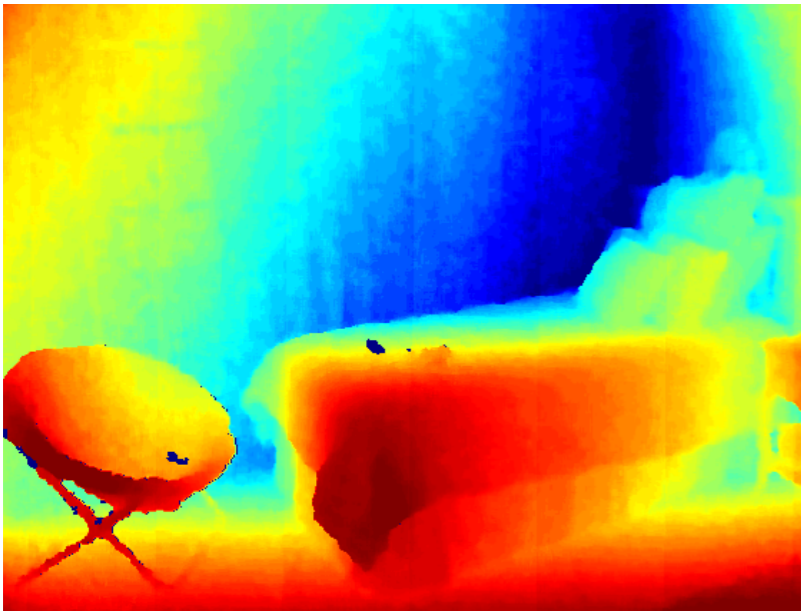
# Results



RGB



gPb-UCM(RGB)

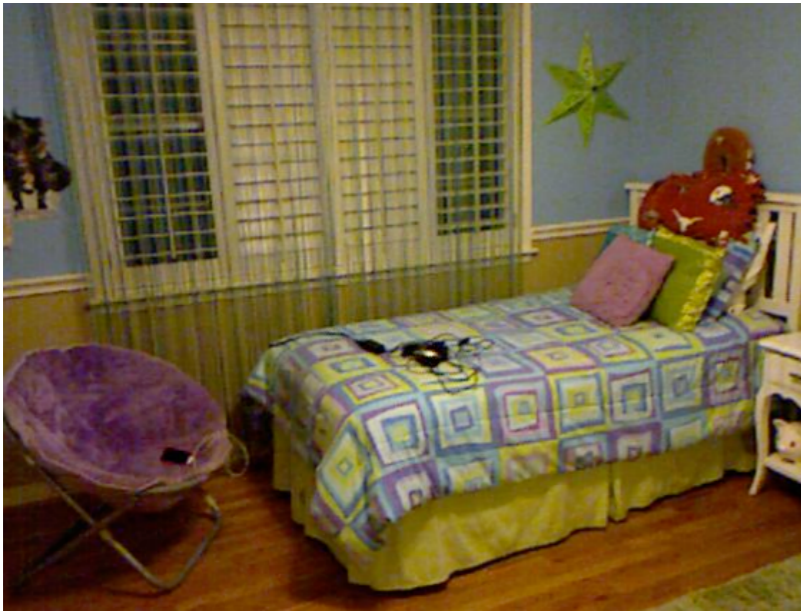


D

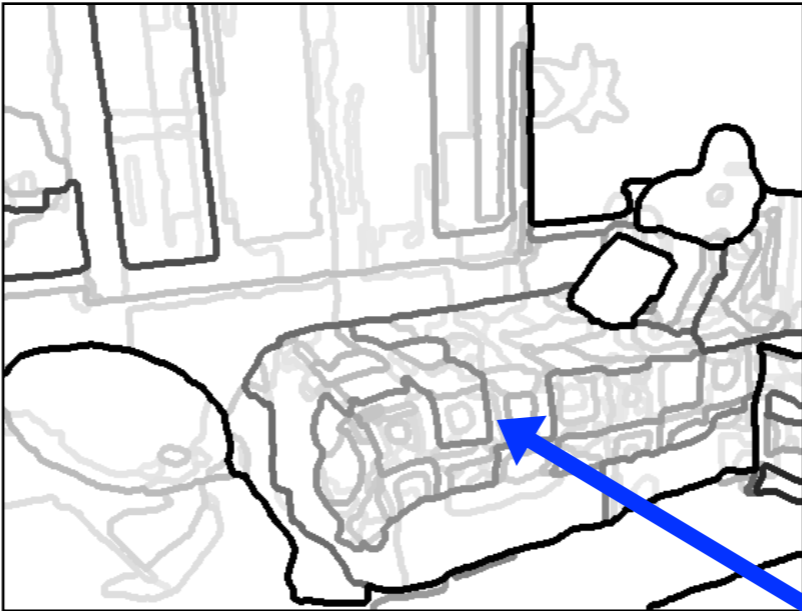


This Work (RGB-D)

# Results

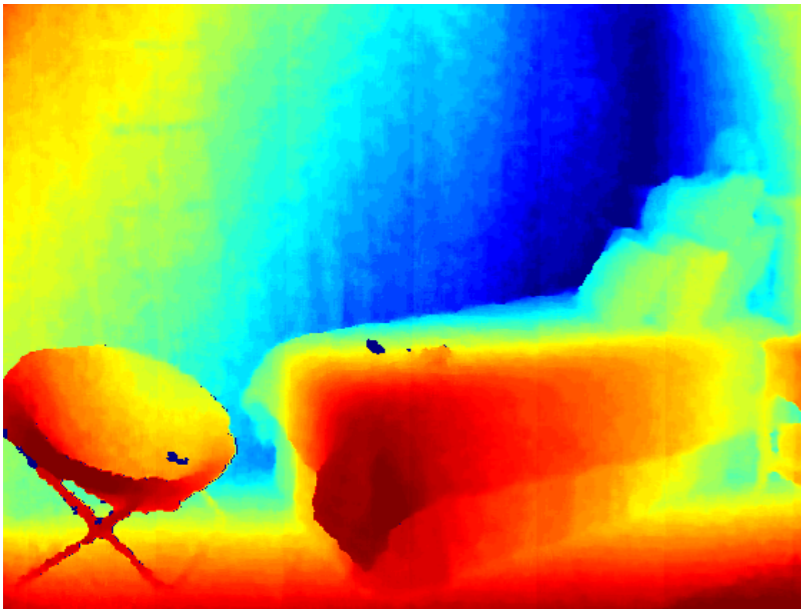


RGB

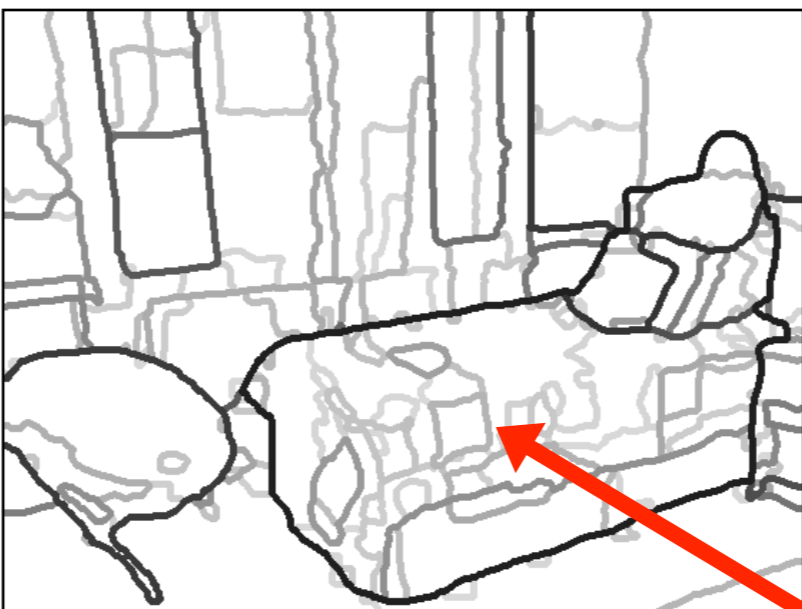


gPb-UCM(RGB)

Less distracted by albedo

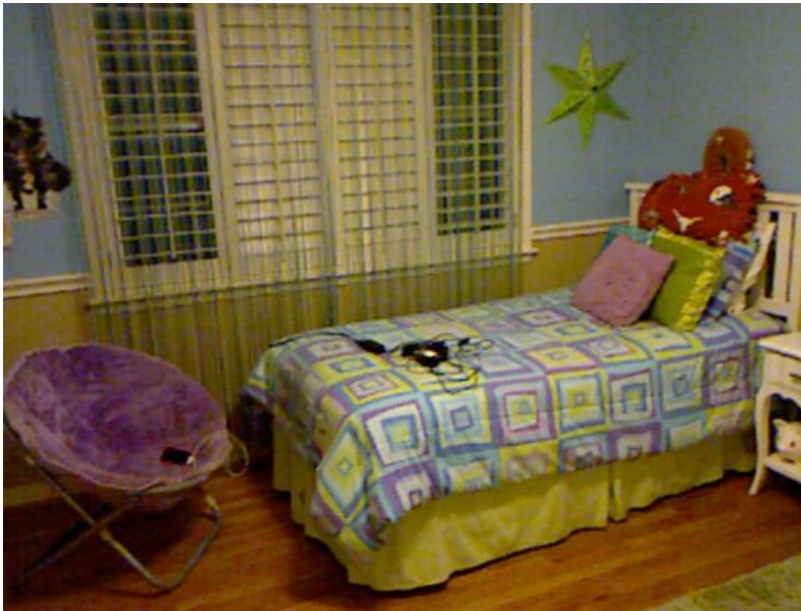


D



This Work (RGB-D)

# Results



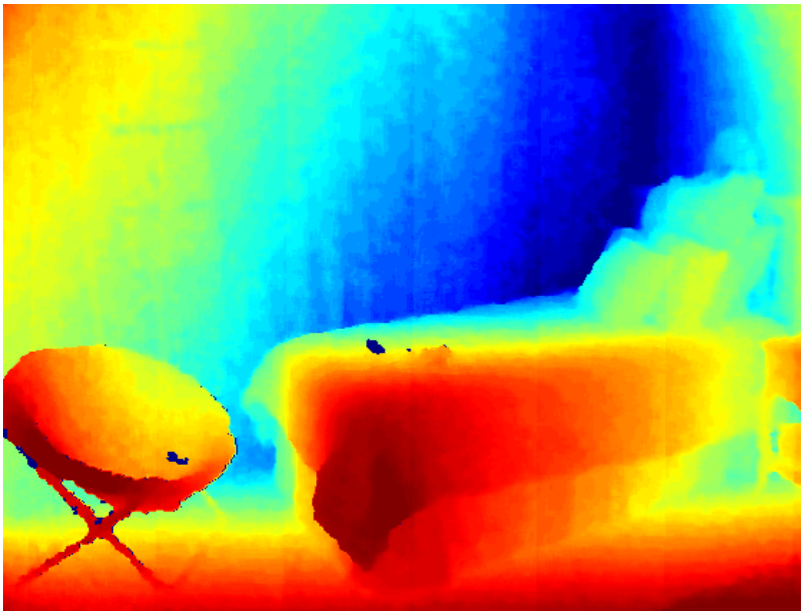
RGB



gPb-UCM(RGB)

Less distracted by albedo

Higher Recall



D

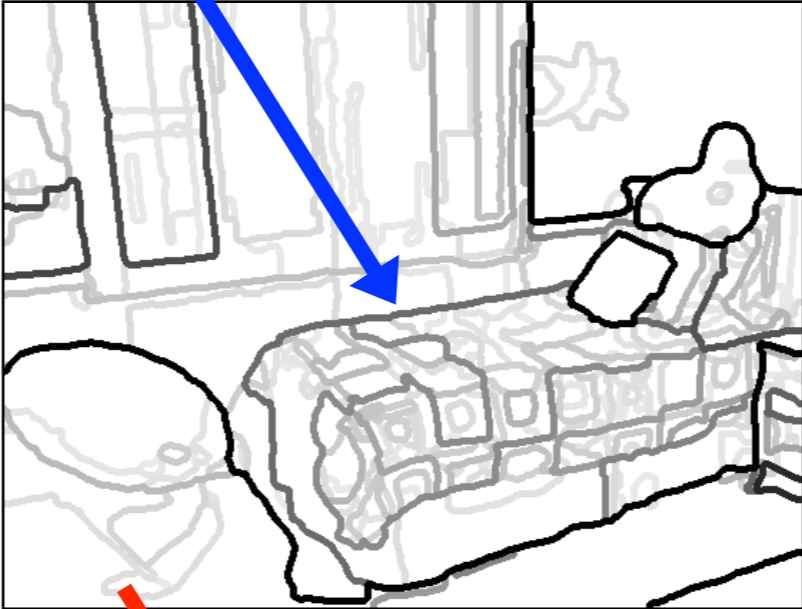


This Work (RGB-D)

# Results



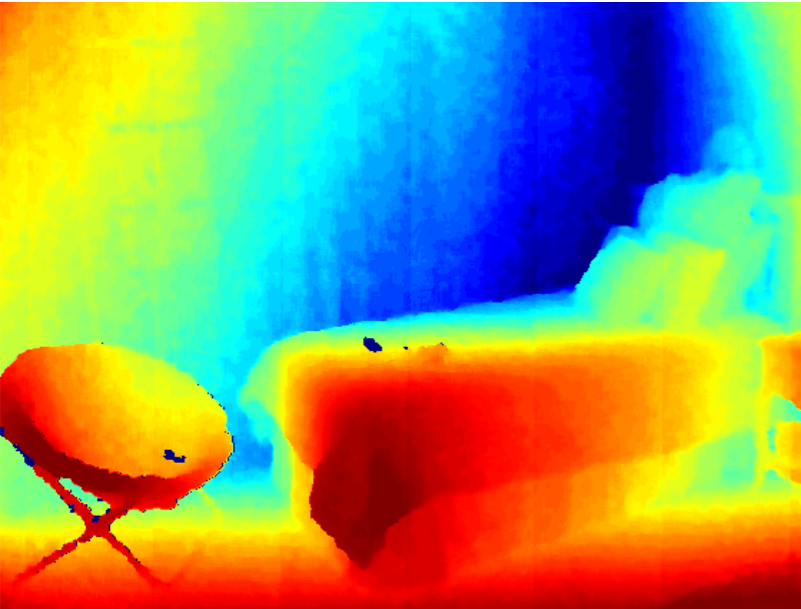
RGB



gPb-UCM(RGB)

Less distracted by albedo

Higher Recall



D



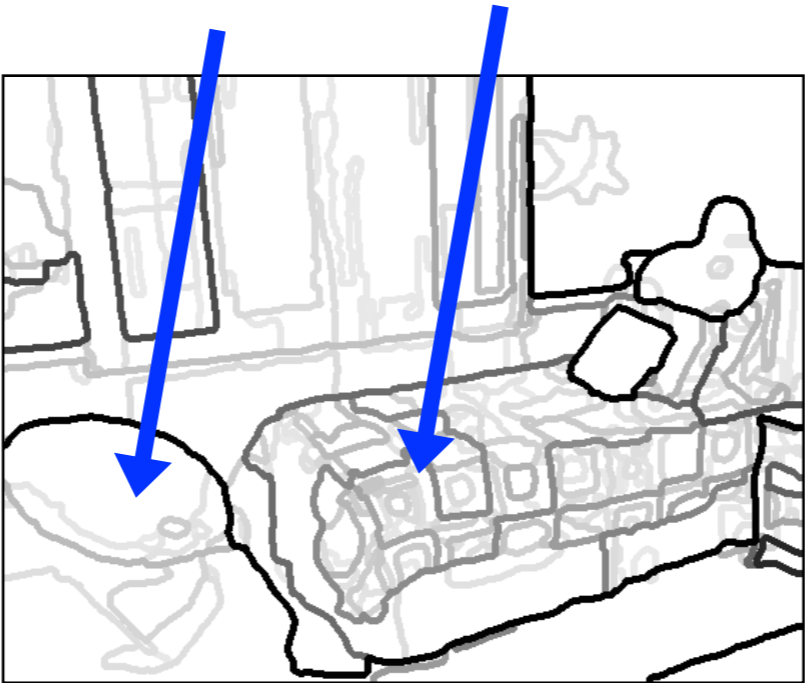
This Work (RGB-D)

Higher Precision

# Results



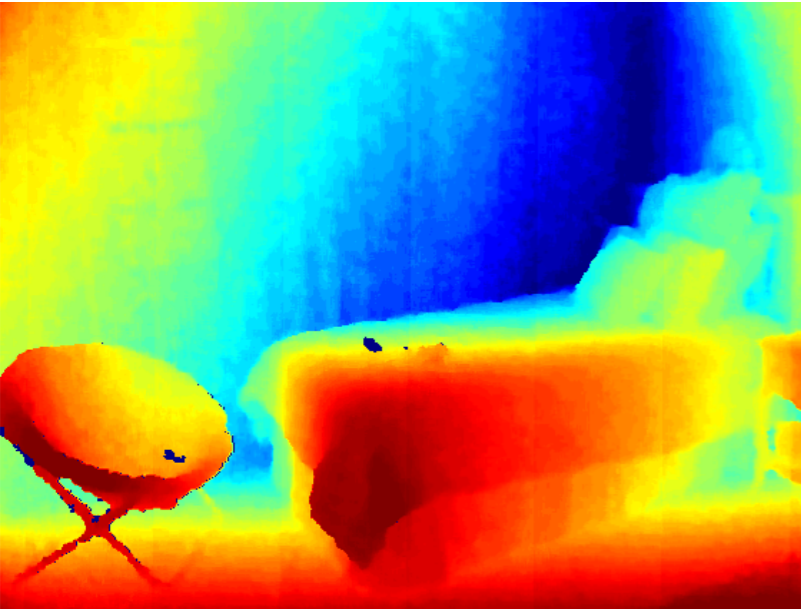
RGB



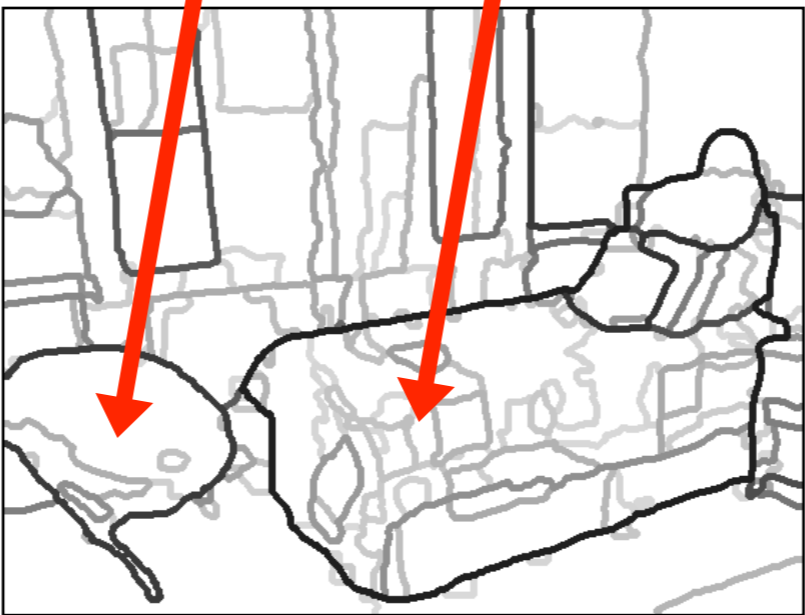
gPb-UCM(RGB)

Less distracted by albedo

Higher Recall



D



This Work (RGB-D)

Higher Precision

More Complete Objects

# Results



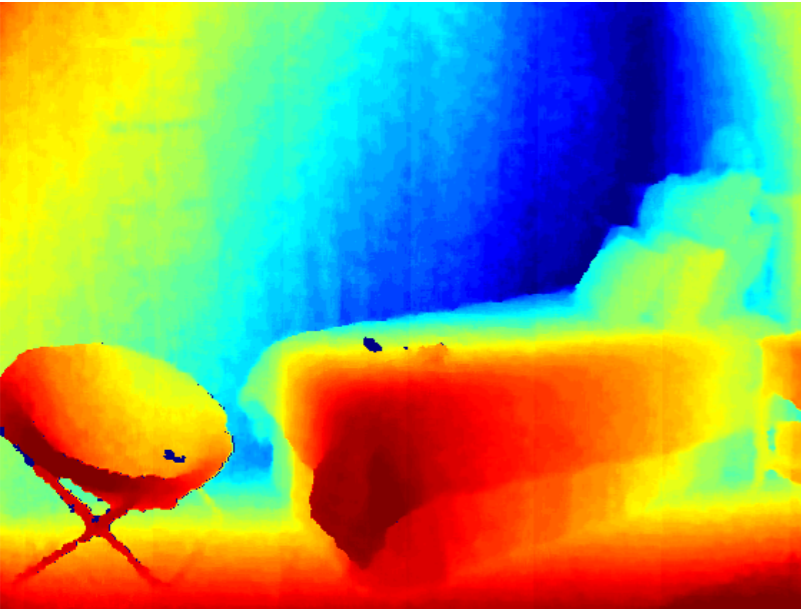
RGB



gPb-UCM(RGB)

Less distracted by albedo

Higher Recall



D



This Work (RGB-D)

Higher Precision

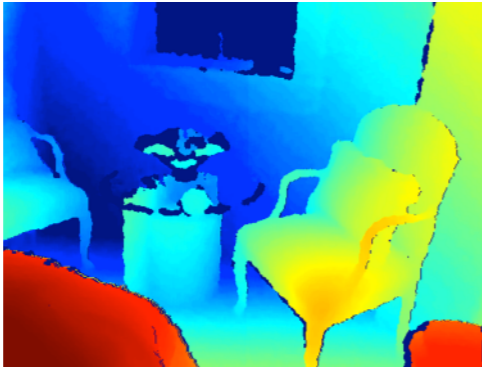
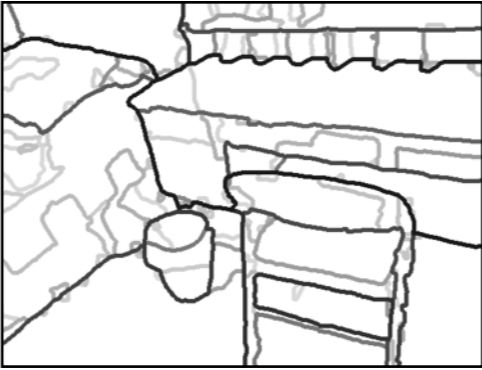
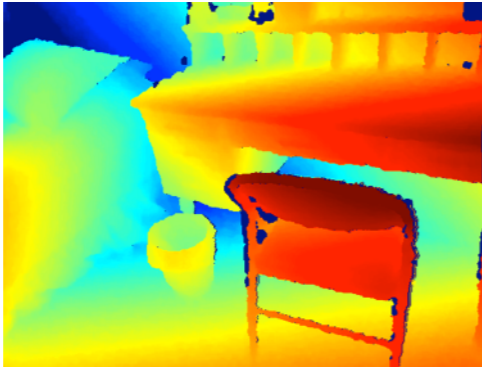
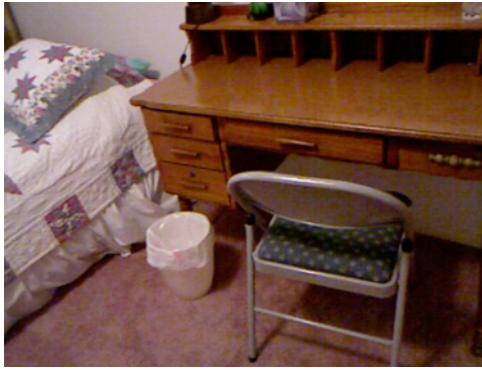
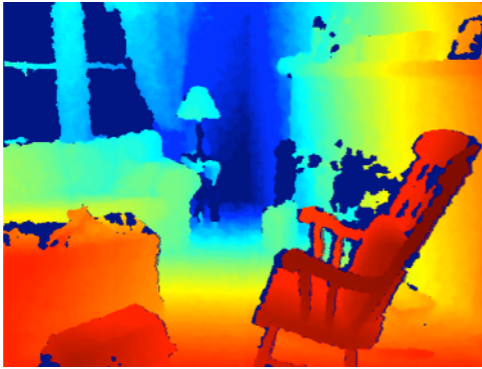
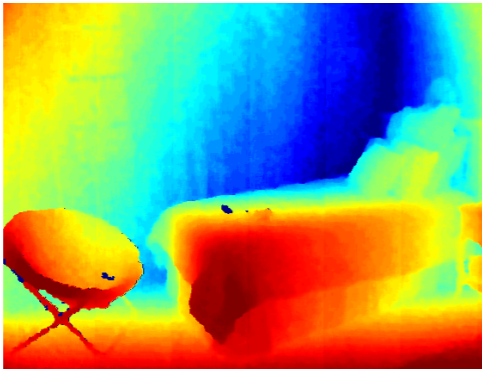
More Complete Objects

# Results

RGB

Depth

Contours



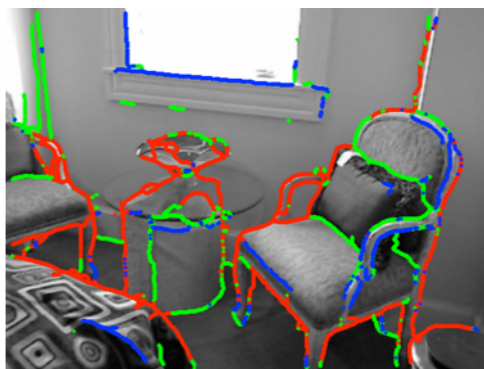
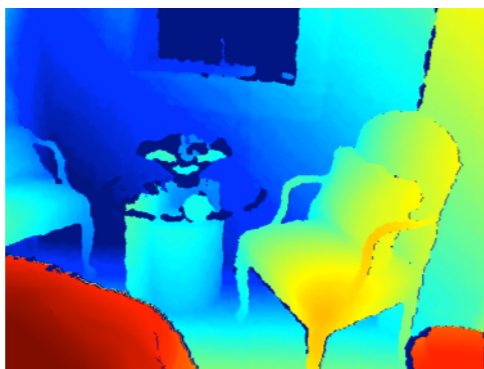
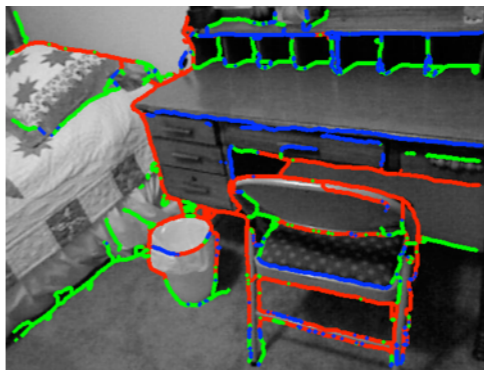
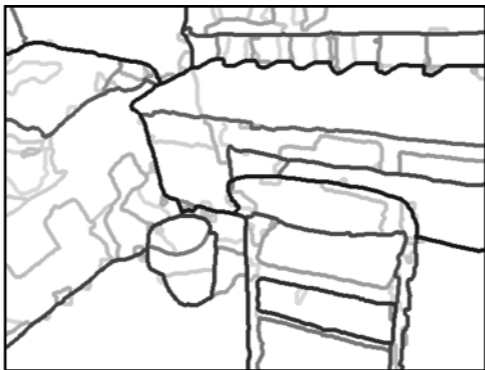
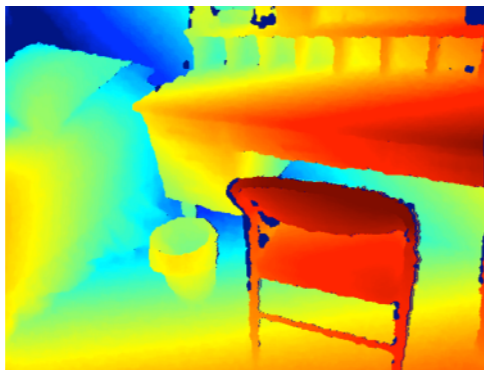
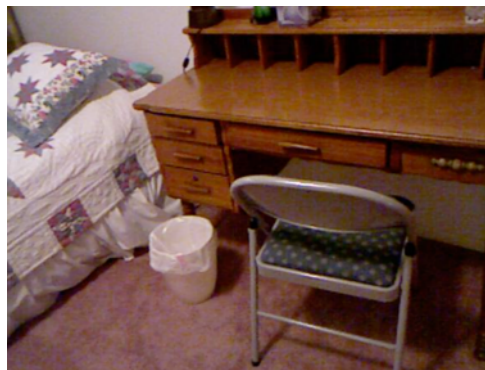
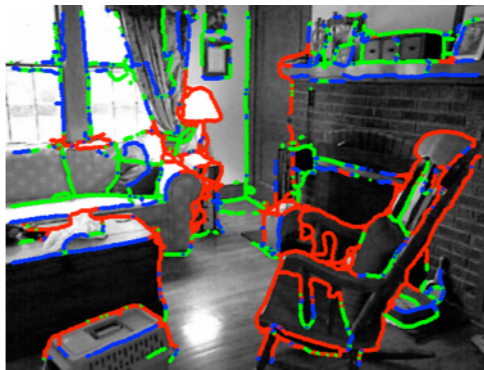
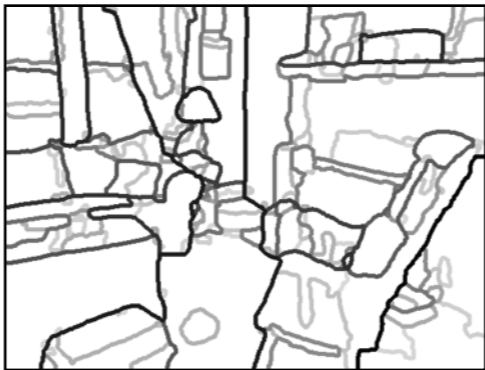
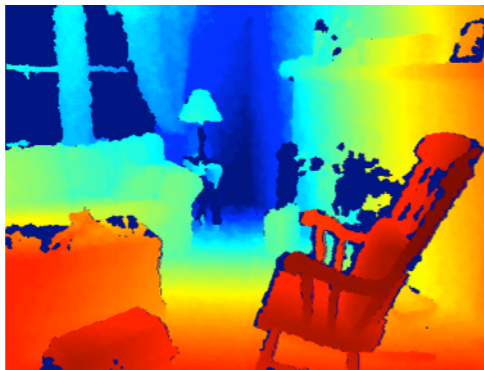
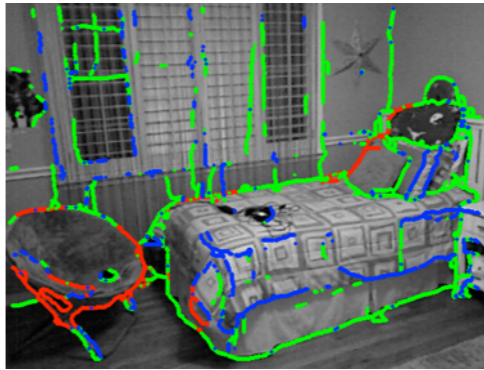
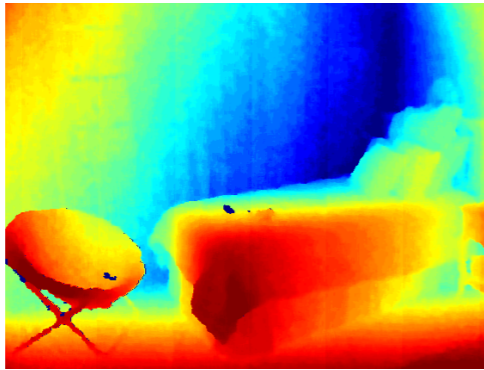
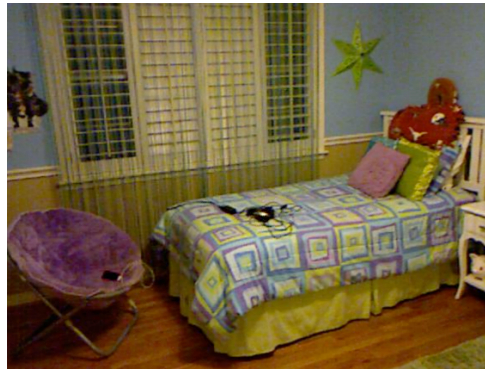
# Results

RGB

Depth

Contours

Contour Labels



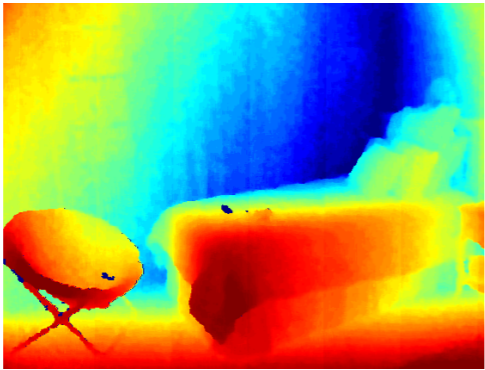


# Results

RGB



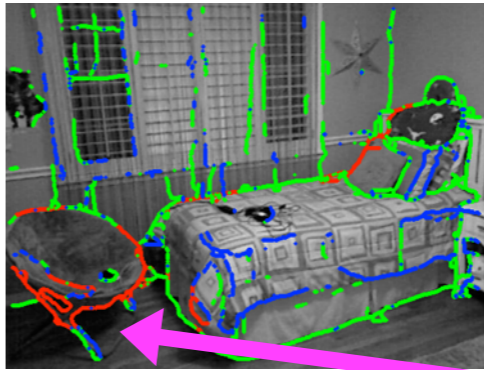
Depth



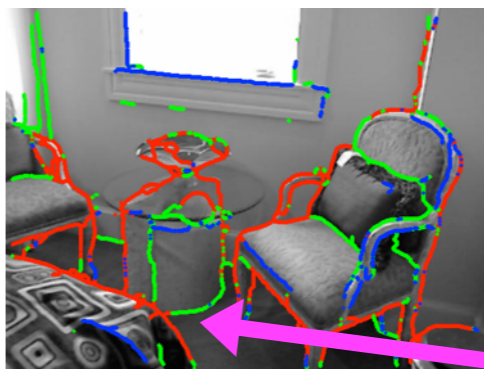
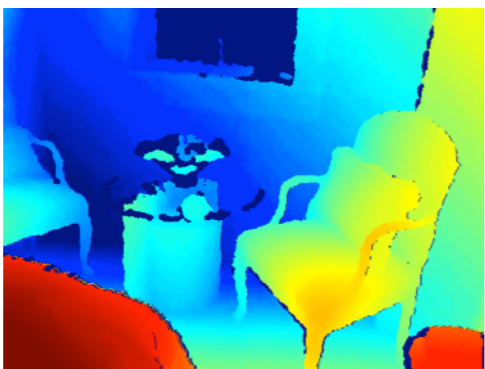
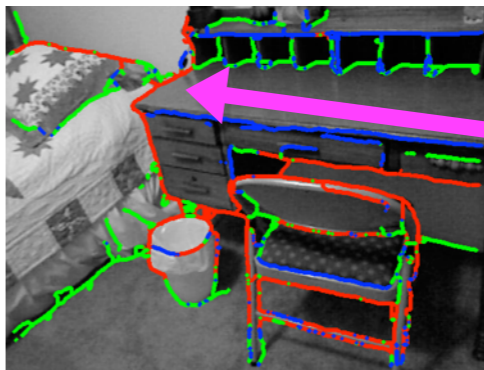
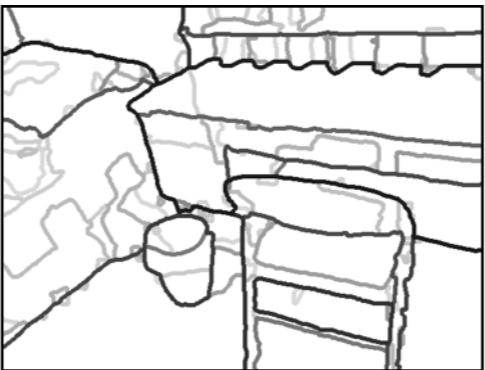
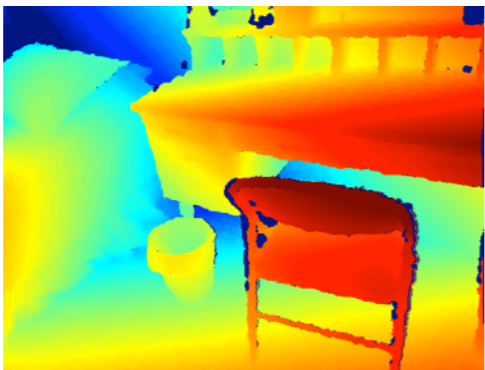
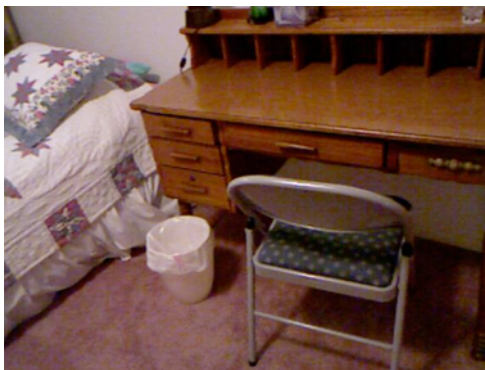
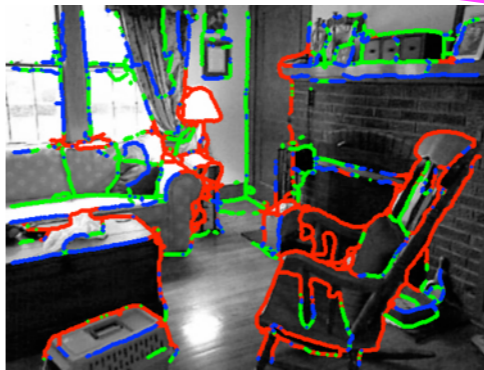
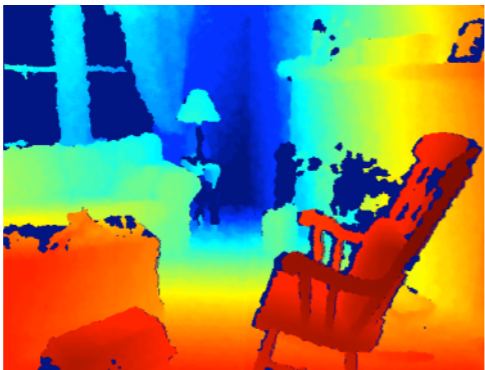
Contours



Contour Labels



Depth Discontinuities (Red)



# Results

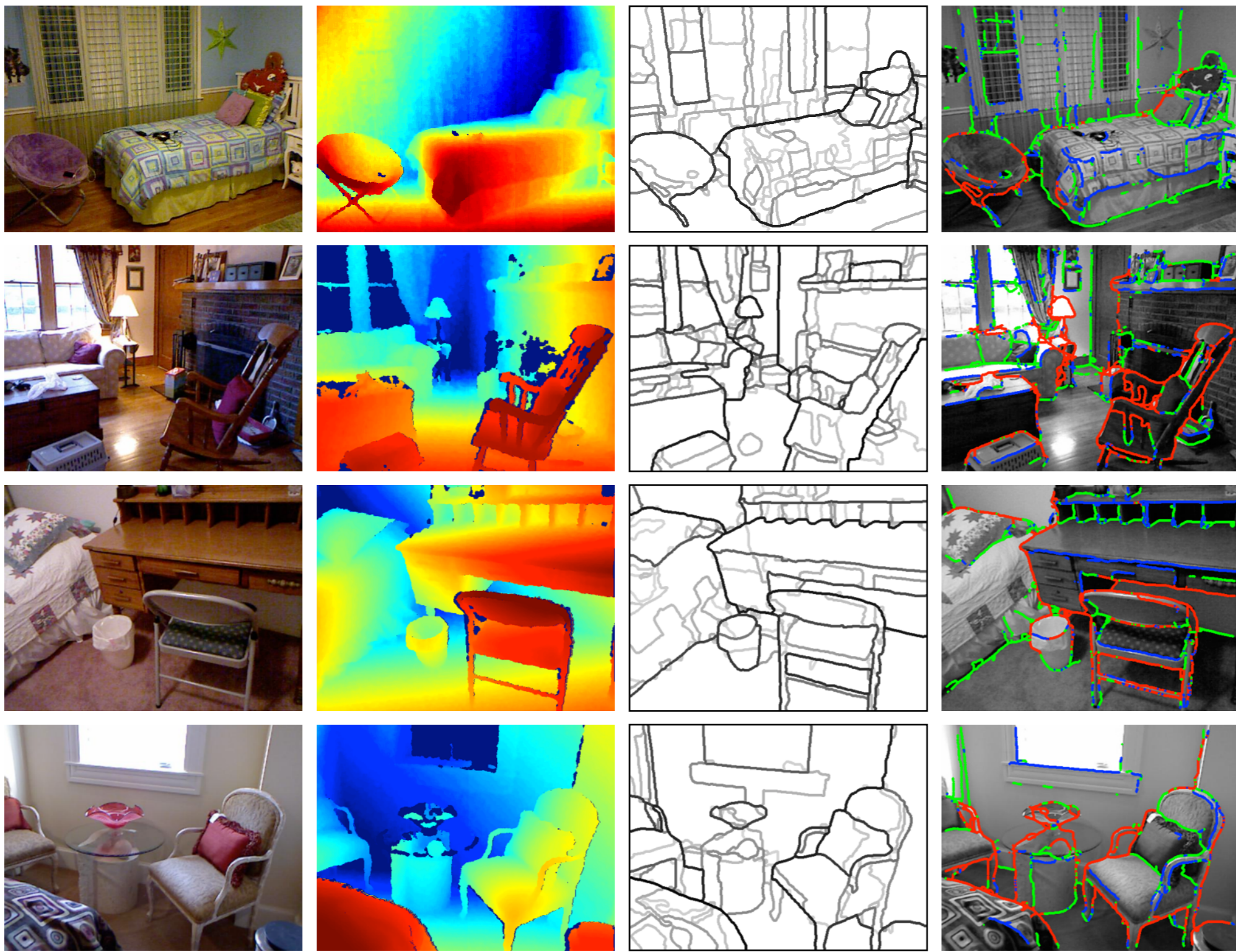
RGB

Depth

Contours

Contour Labels

Depth  
Discontinuities  
(Red)



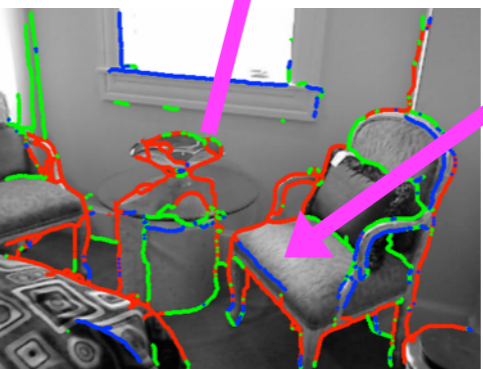
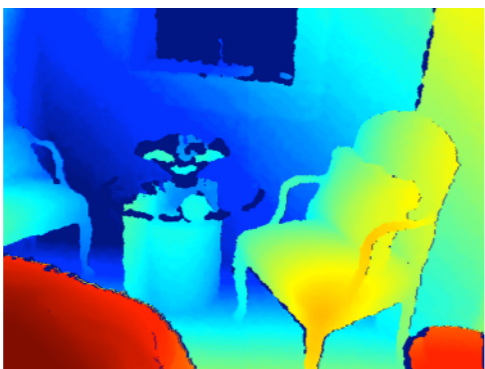
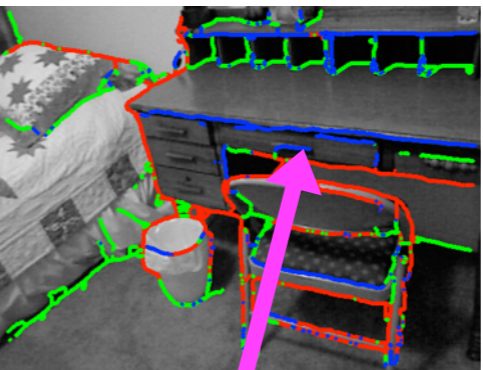
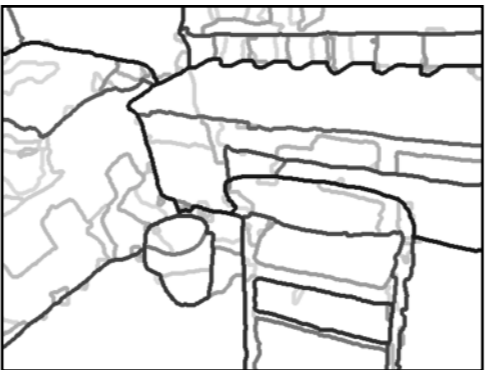
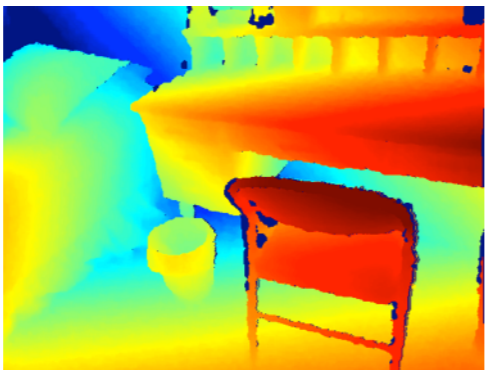
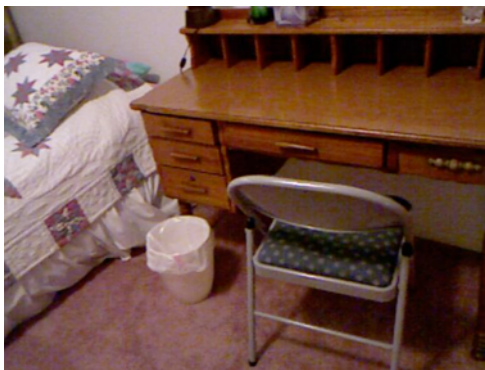
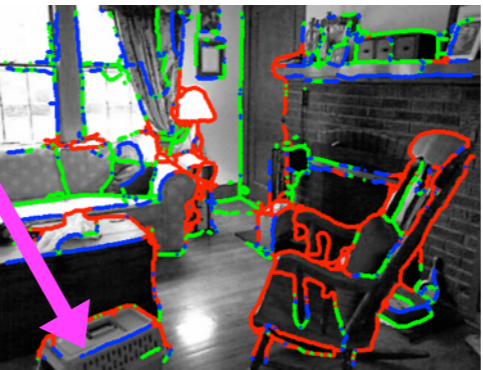
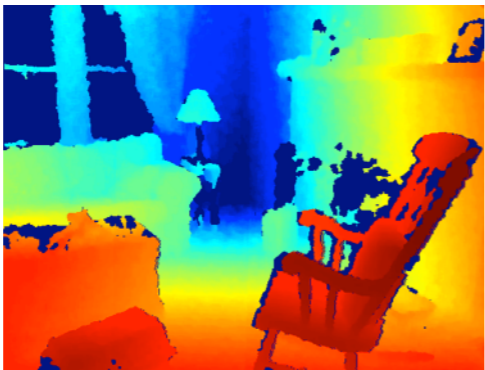
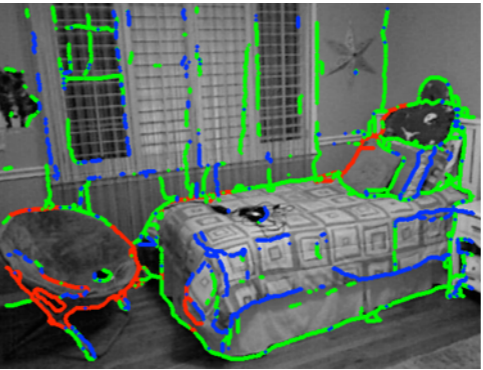
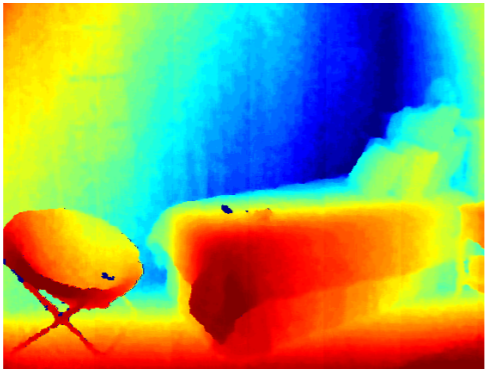
# Results

RGB

Depth

Contours

Contour Labels



Depth  
Discontinuities  
(Red)

Convex  
Normal  
Discontinuities  
(Blue)

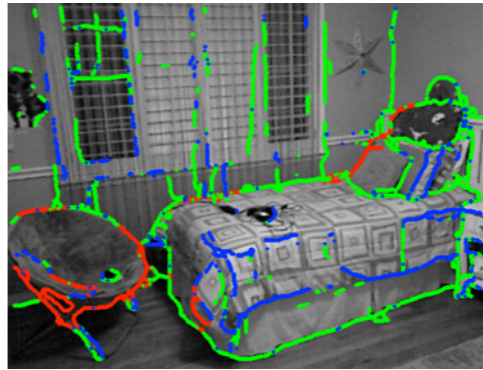
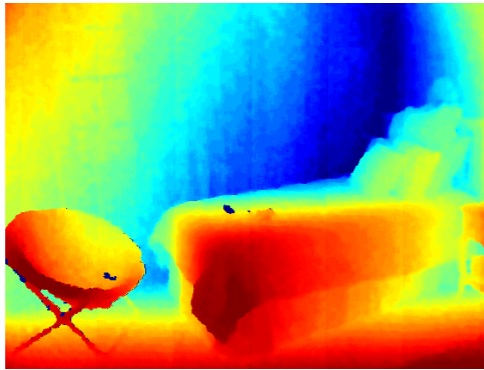
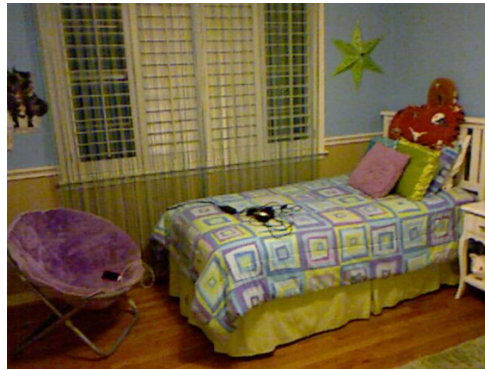
# Results

RGB

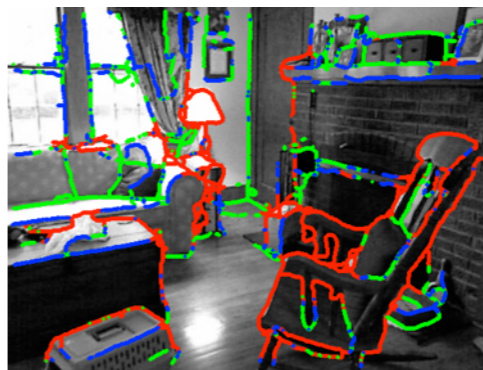
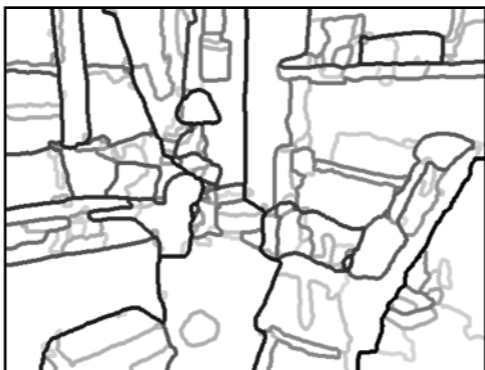
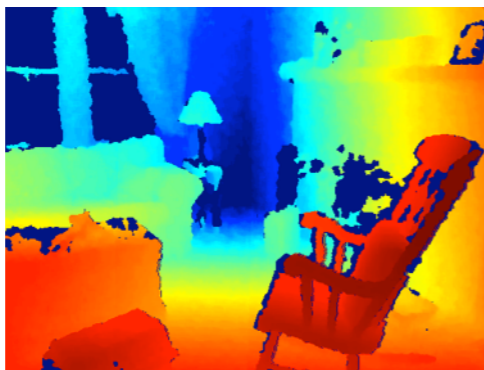
Depth

Contours

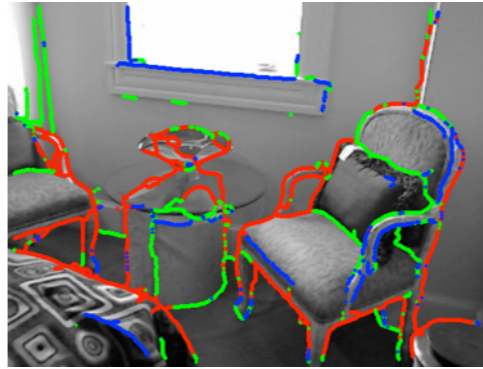
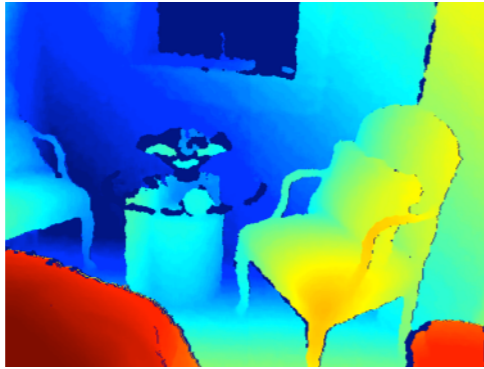
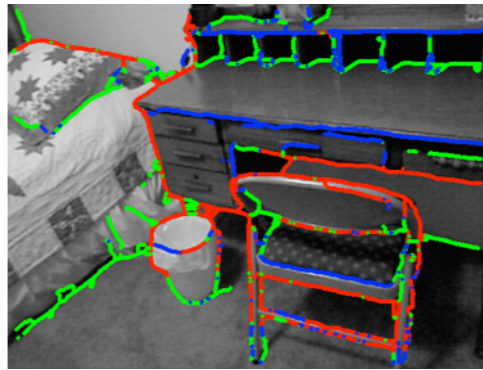
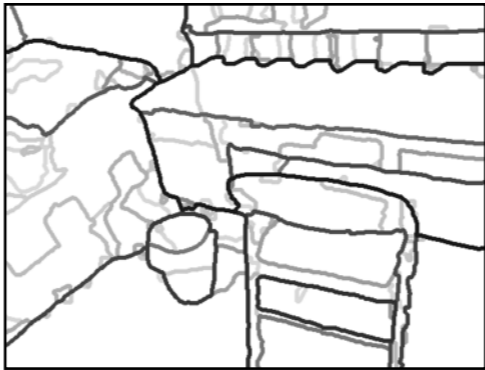
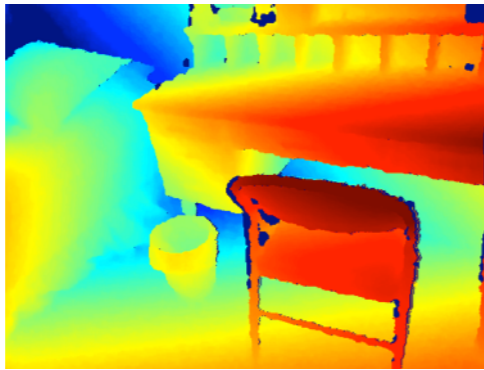
Contour Labels



Depth  
Discontinuities  
(Red)



Convex  
Normal  
Discontinuities  
(Blue)

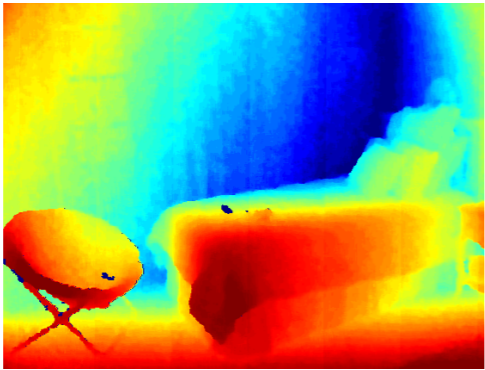


# Results

RGB



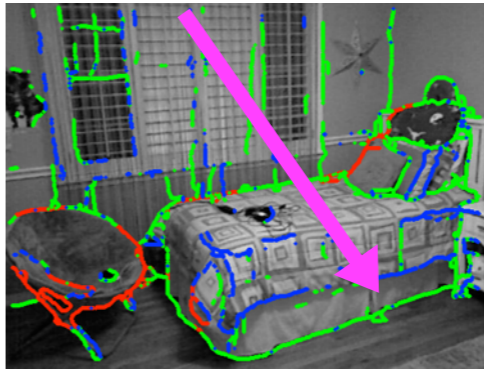
Depth



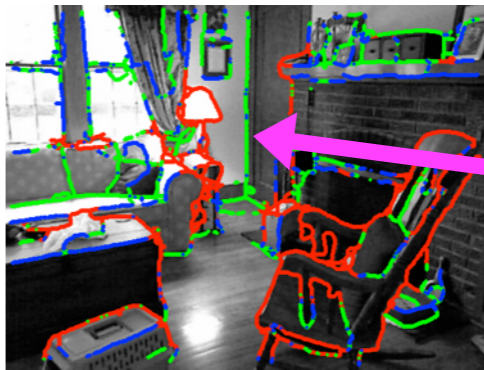
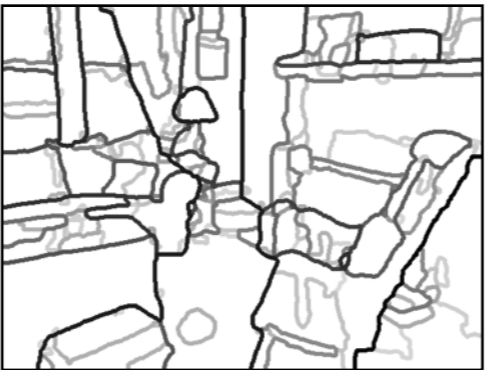
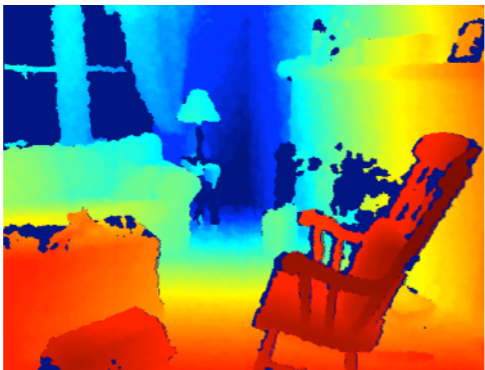
Contours



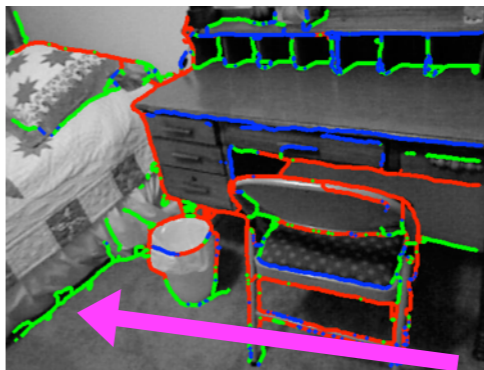
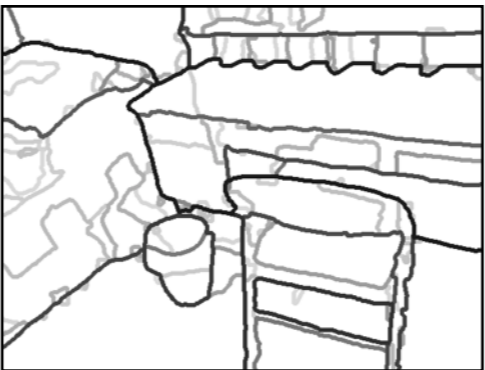
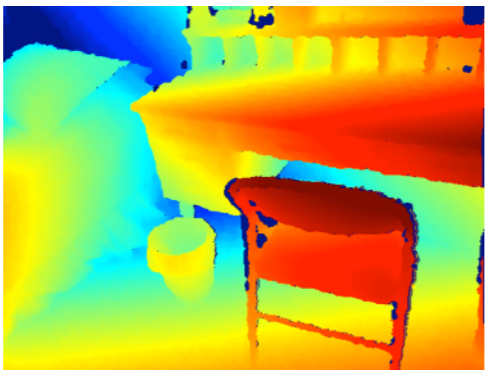
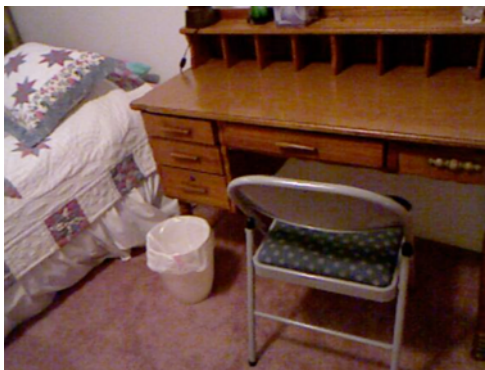
Contour Labels



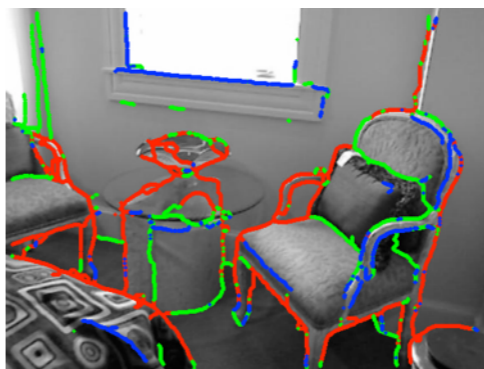
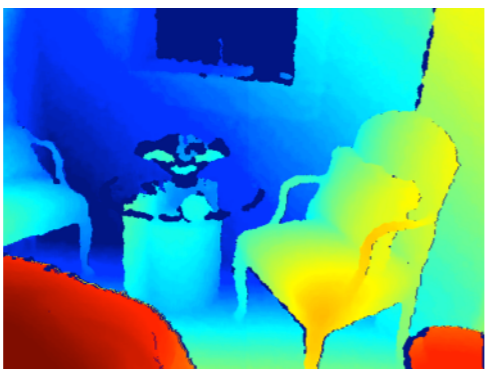
Depth Discontinuities (Red)



Convex Normal Discontinuities (Blue)



Concave Normal Discontinuities (Green)

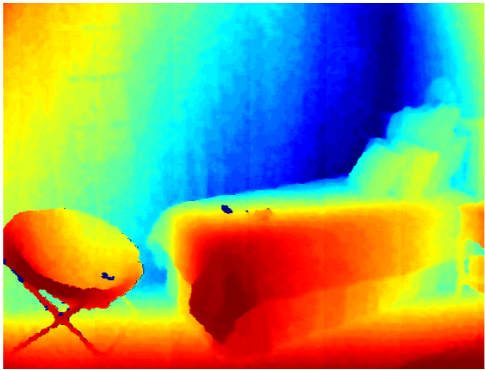


# Results

RGB



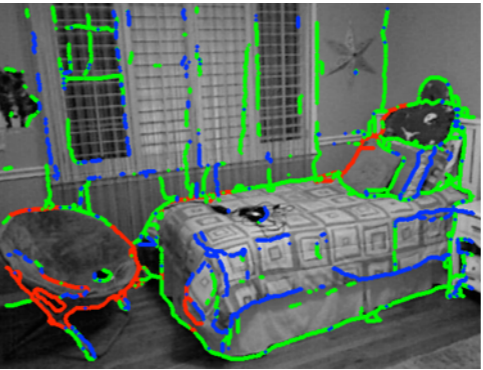
Depth



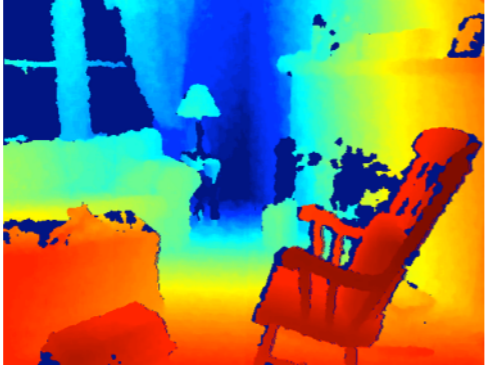
Contours



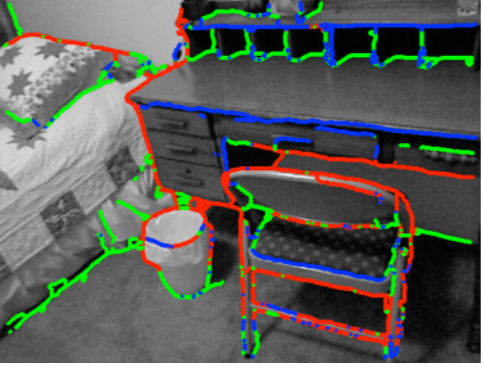
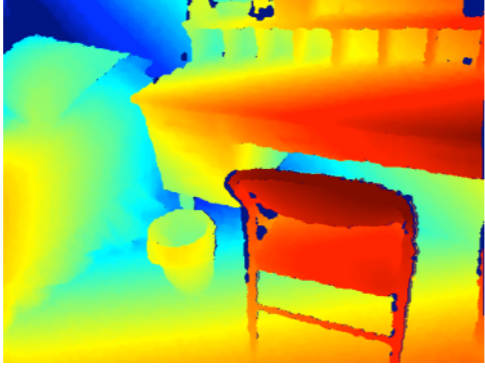
Contour Labels



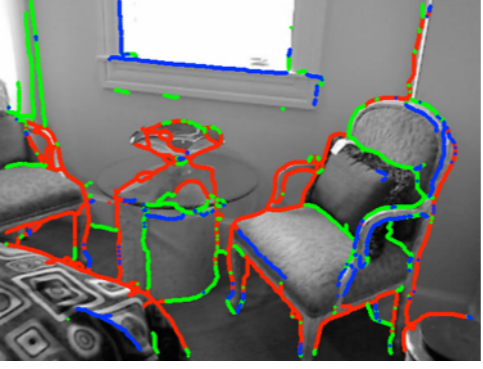
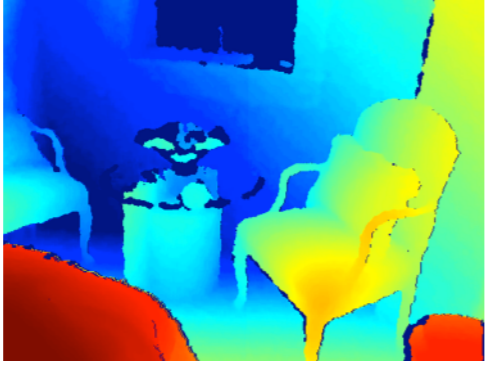
Depth  
Discontinuities  
(Red)



Convex  
Normal  
Discontinuities  
(Blue)

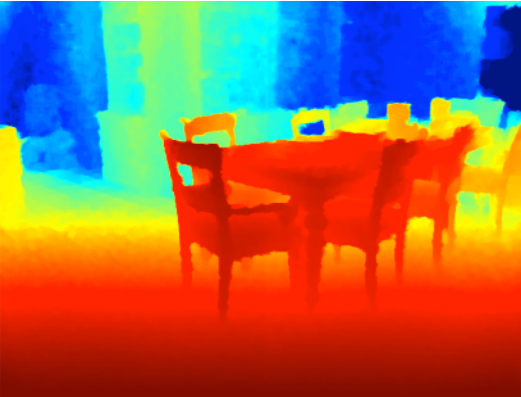


Concave  
Normal  
Discontinuities  
(Green)



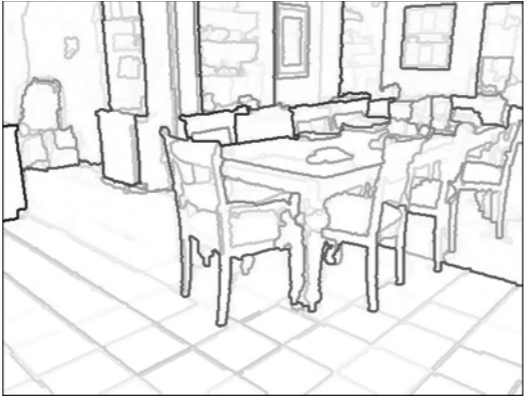
# Overview

## Input



Color and Depth Image Pair

## Re-organization

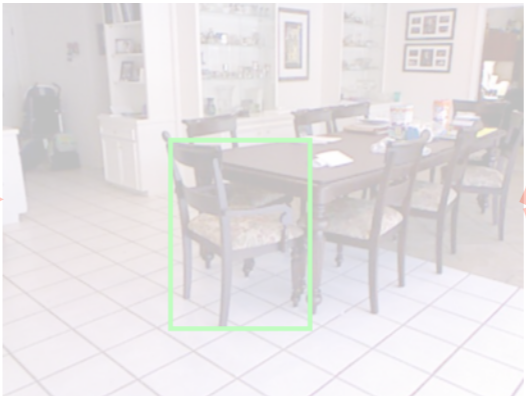


Contour Detection



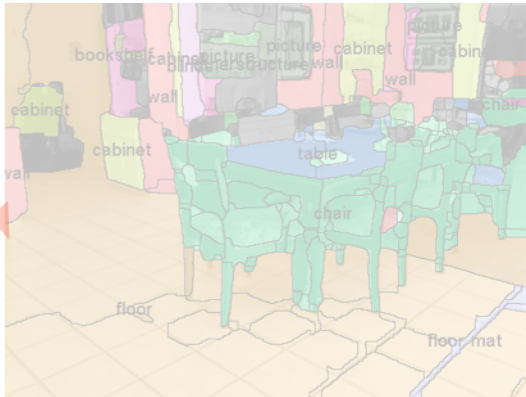
Region Proposal Generation

## Recognition

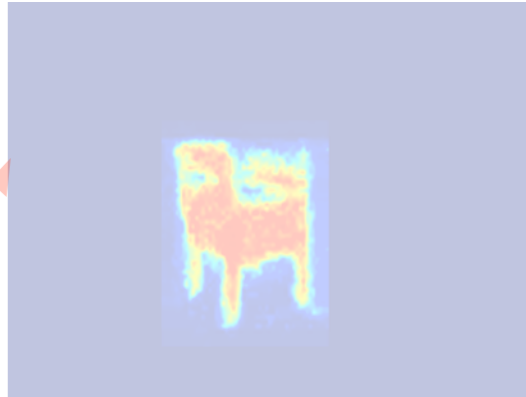


Object Detection

## Extensions



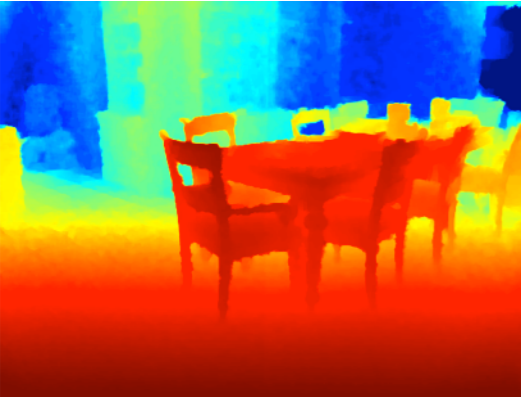
Semantic Segm.



Instance Segm.

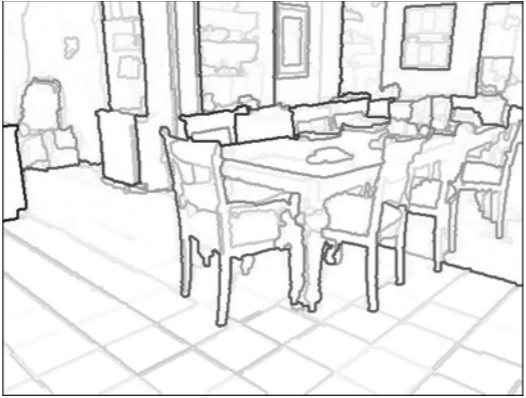
# Overview

## Input

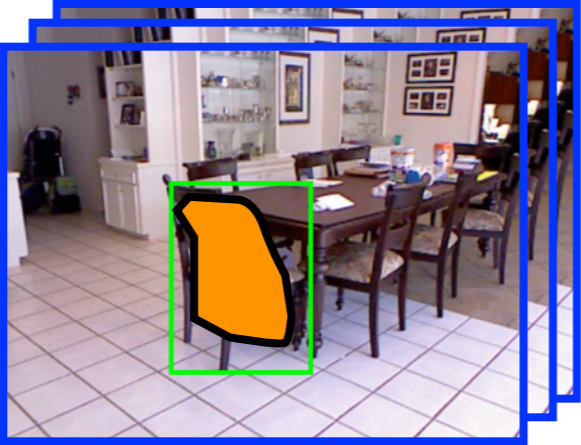


Color and Depth Image Pair

## Re-organization

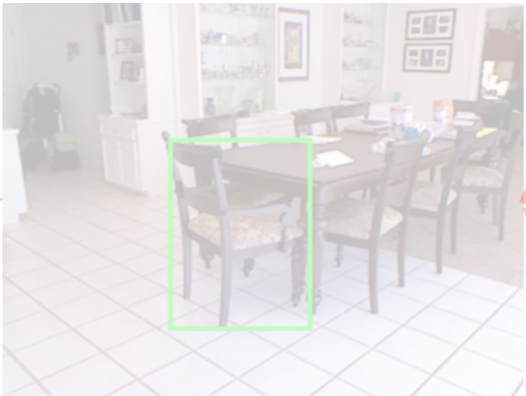


Contour Detection



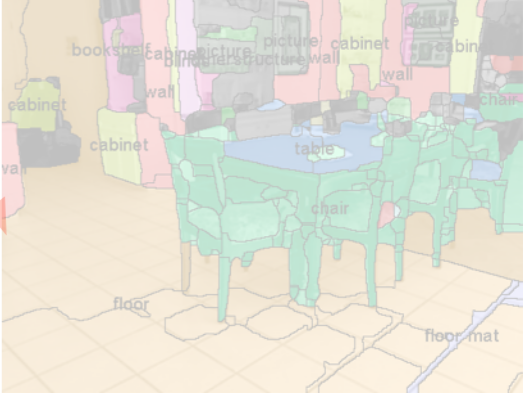
Region Proposal Generation

## Recognition

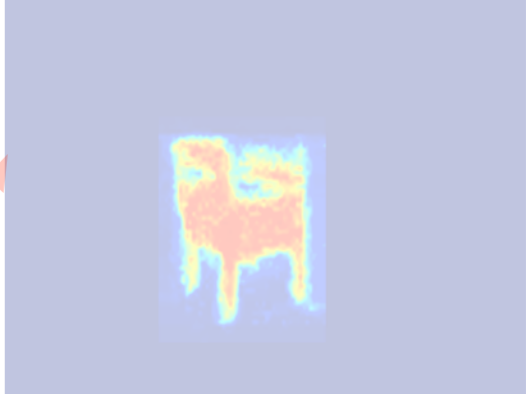


Object Detection

## Extensions



Semantic Segm.

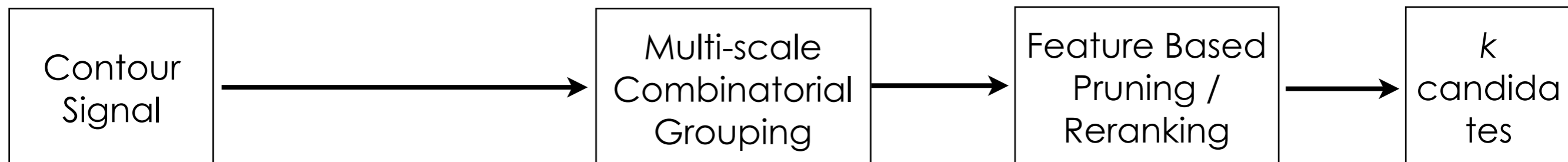
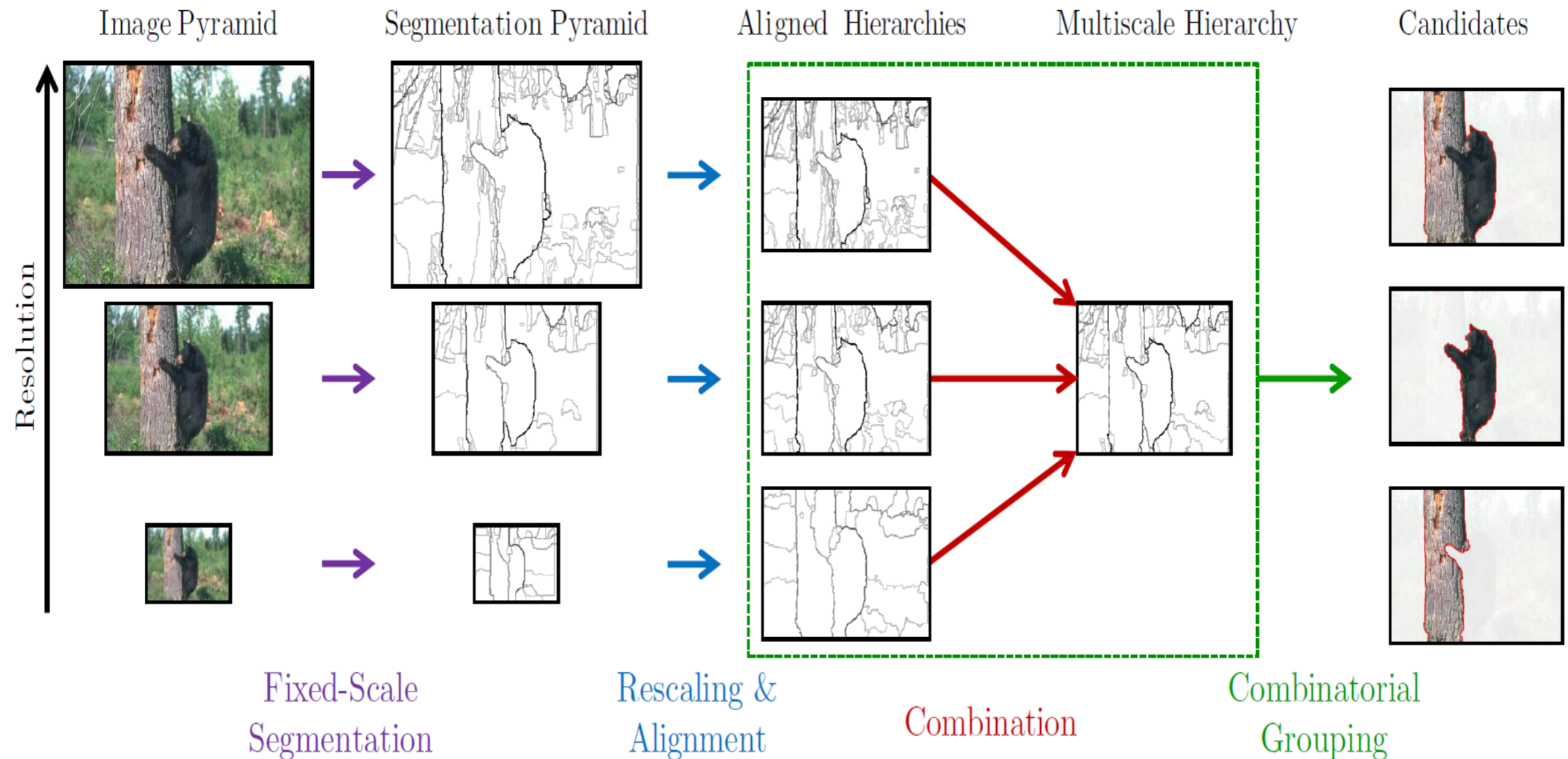


Instance Segm.



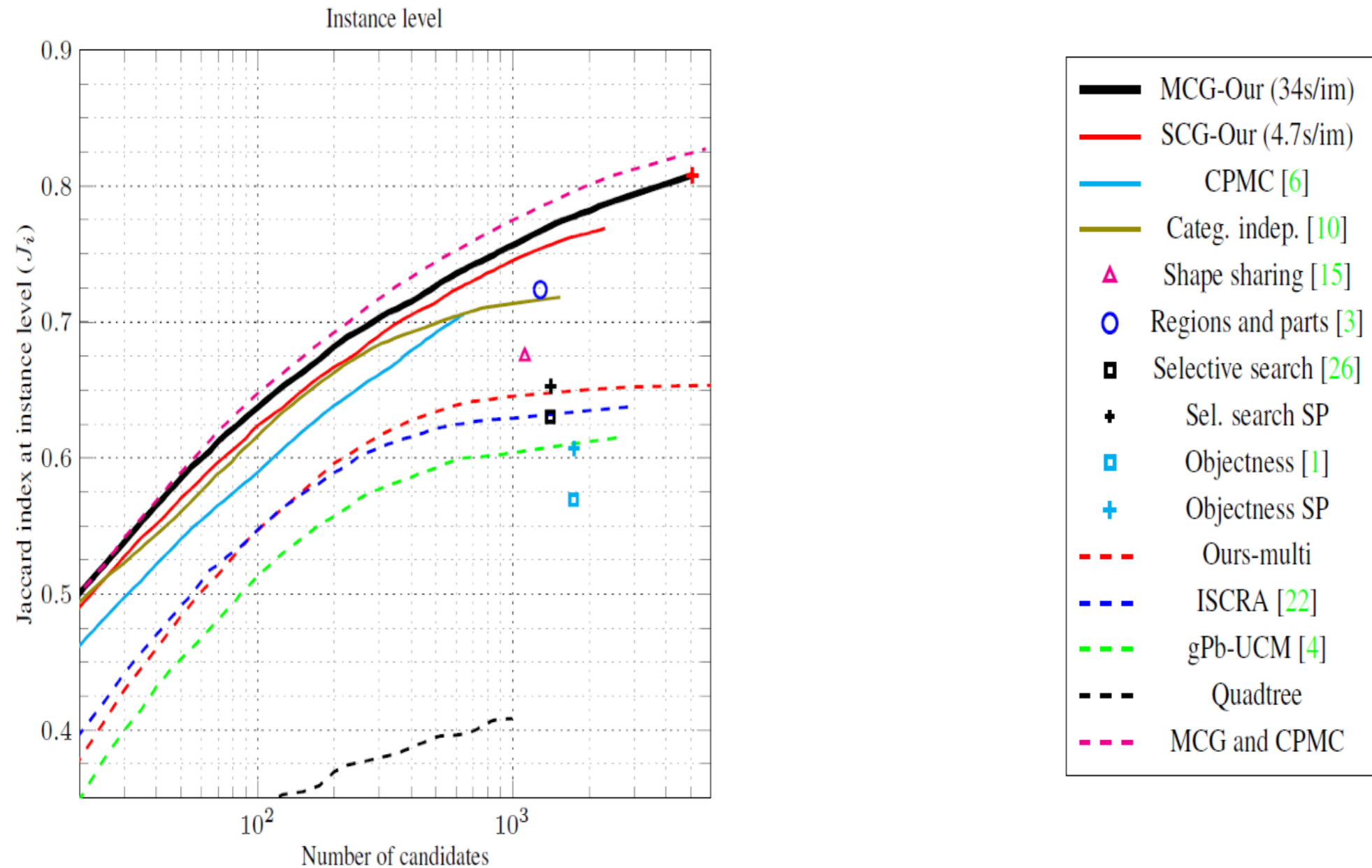
# Multiscale Combinatorial Regions

Arbelaez, Pont-Tuset, Barron, Marques & Malik, CVPR 2014



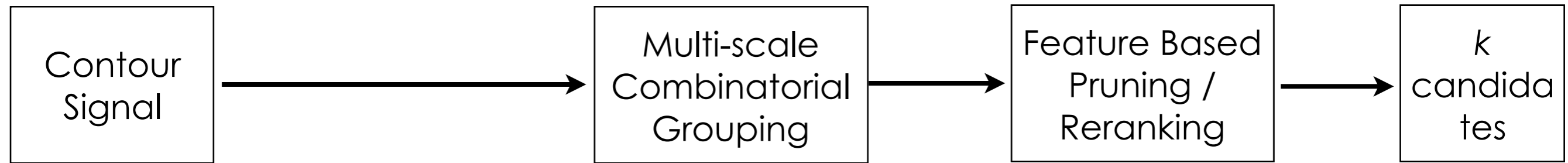
# Multiscale Combinatorial Regions

Arbelaez, Pont-Tuset, Barron, Marques & Malik, CVPR 2014



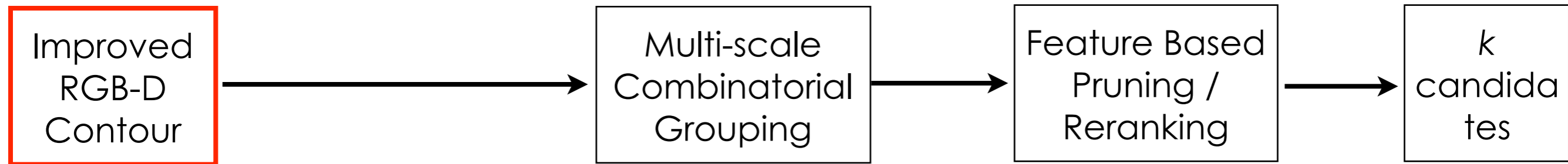
# Region Proposal Generation

Generalize MCG to RGB-D Images



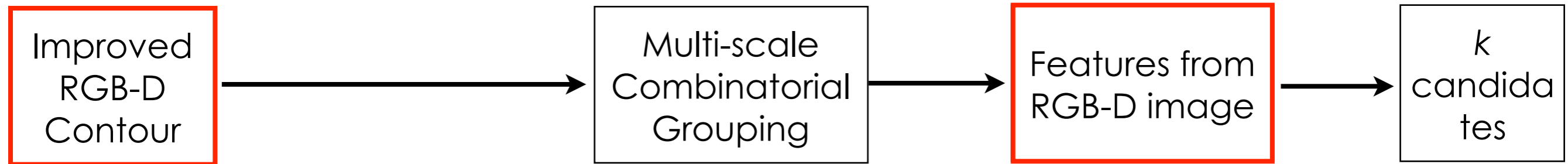
# Region Proposal Generation

Generalize MCG to RGB-D Images



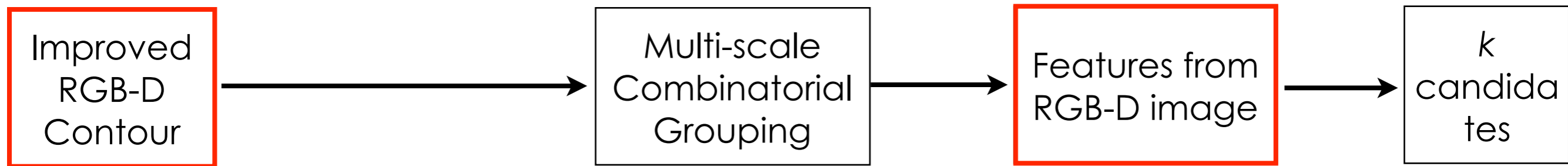
# Region Proposal Generation

Generalize MCG to RGB-D Images



# Region Proposal Generation

## Generalize MCG to RGB-D Images

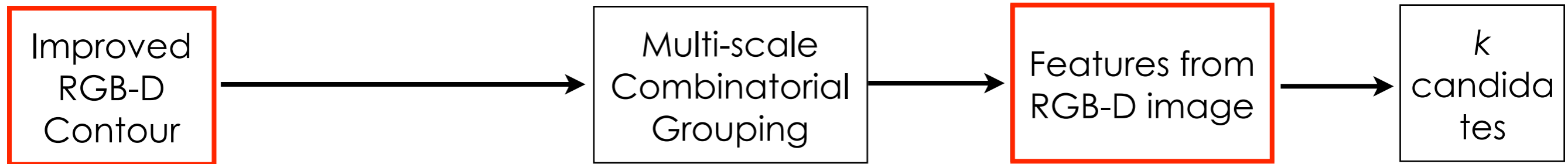


### Features for Pruning

- mean and std of
  - disparity
  - height above ground, angle with gravity
  - X, Y, Z coordinates
- 3D extent of the region
- min, max height above ground
- facing up area, vertical area ...
- tightest box in top-view

# Region Proposal Generation

## Generalize MCG to RGB-D Images



### Features for Pruning

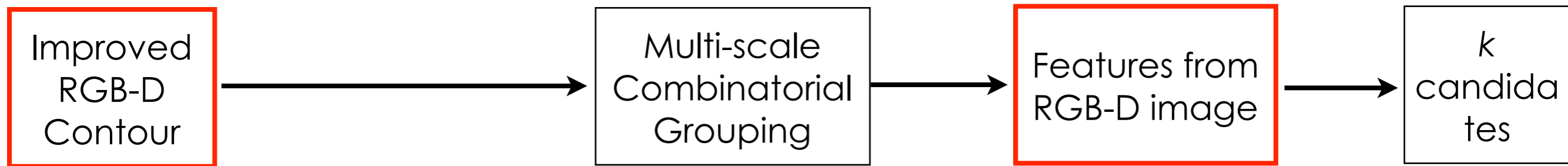
- mean and std of
  - disparity
  - height above ground, angle with gravity
  - X, Y, Z coordinates
- 3D extent of the region
- min, max height above ground
- facing up area, vertical area ...
- tightest box in top-view

### Can be efficiently computed

- decompose over superpixels
- compute over superpixels and accumulate over the regions

# Region Proposal Generation

## Generalize MCG to RGB-D Images



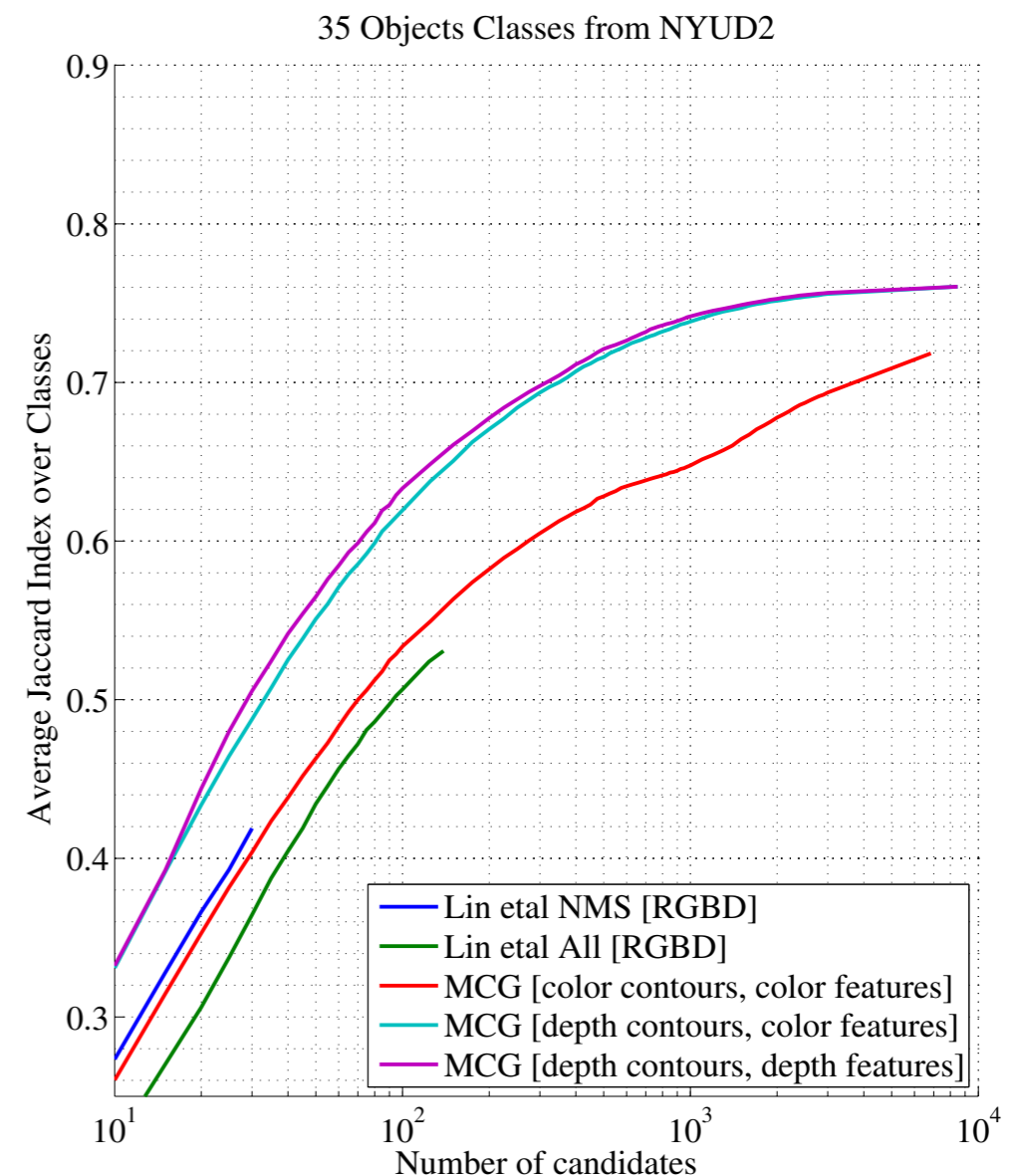
### Features for Pruning

- mean and std of
  - disparity
  - height above ground, angle with gravity
  - X, Y, Z coordinates
- 3D extent of the region
- min, max height above ground
- facing up area, vertical area ...
- tightest box in top-view

### Can be efficiently computed

- decompose over superpixels
- compute over superpixels and accumulate over the regions

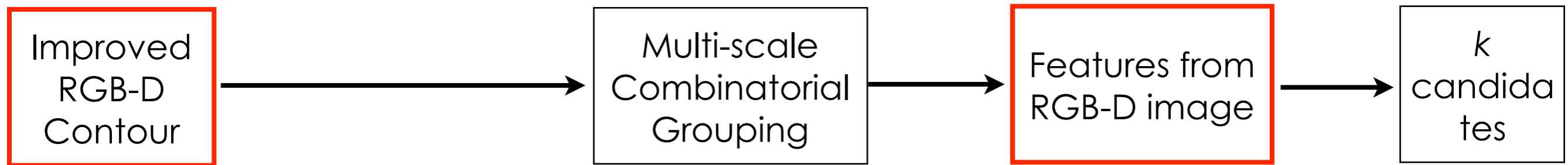
### Results





# Region Proposal Generation

## Generalize MCG to RGB-D Images



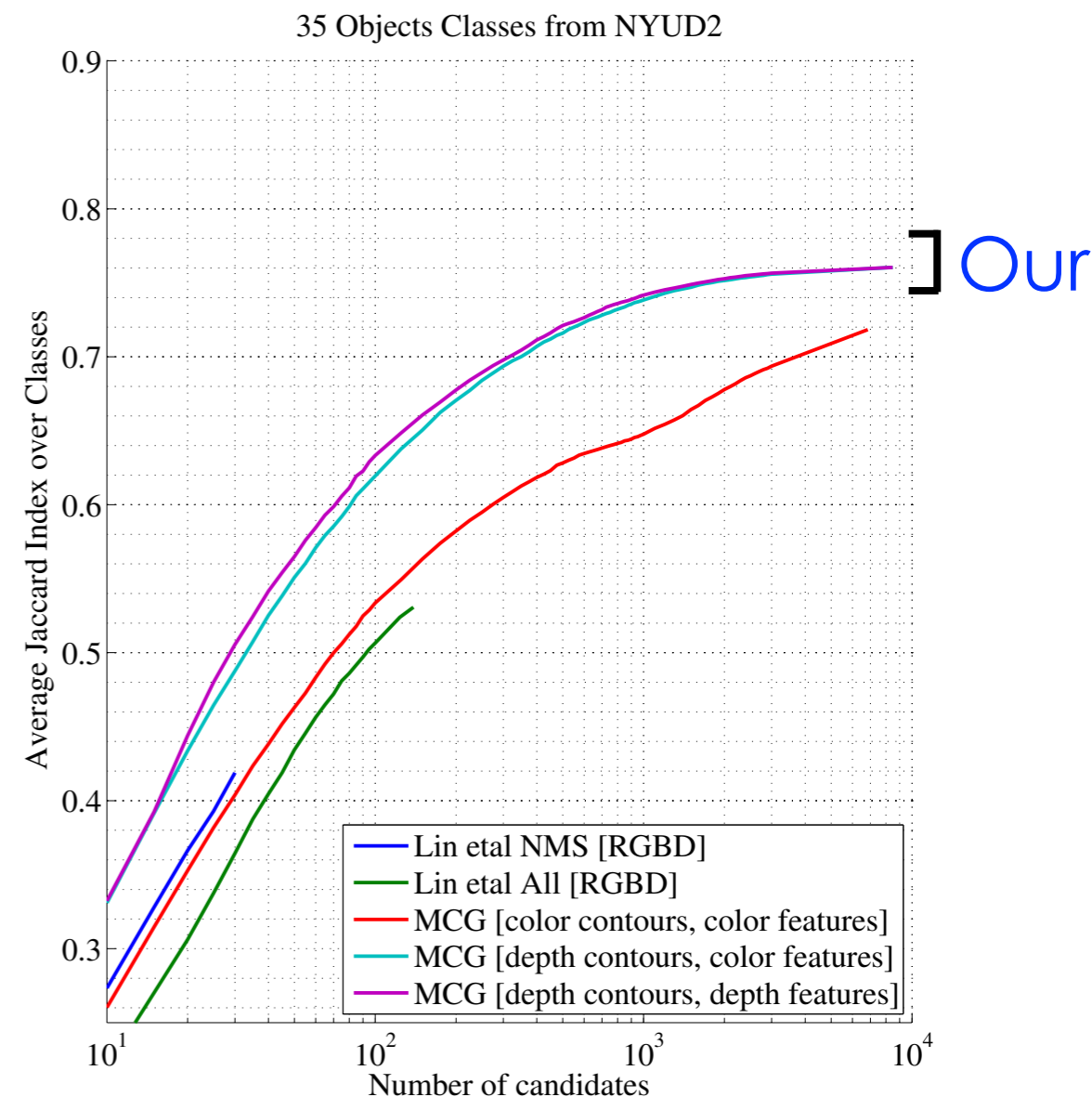
### Features for Pruning

- mean and std of
  - disparity
  - height above ground, angle with gravity
  - X, Y, Z coordinates
- 3D extent of the region
- min, max height above ground
- facing up area, vertical area ...
- tightest box in top-view

### Can be efficiently computed

- decompose over superpixels
- compute over superpixels and accumulate over the regions

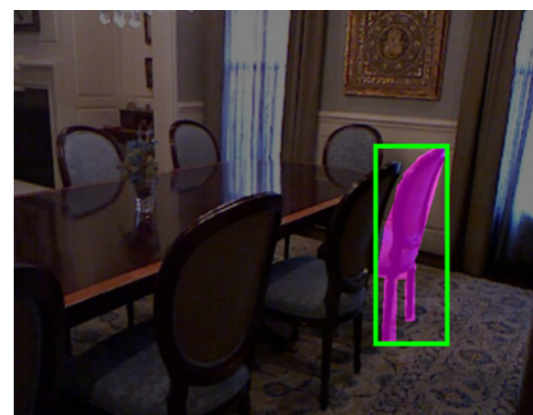
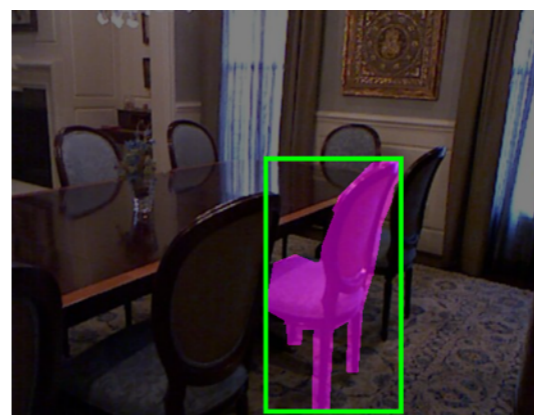
### Results



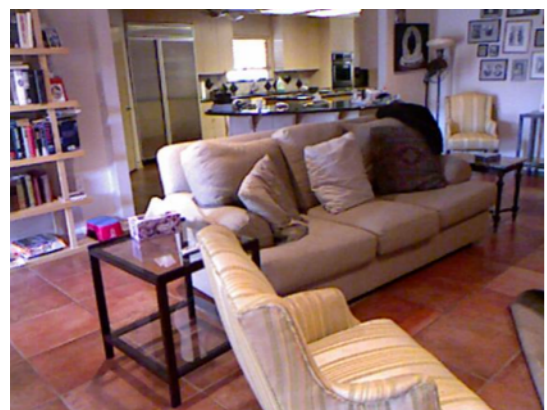
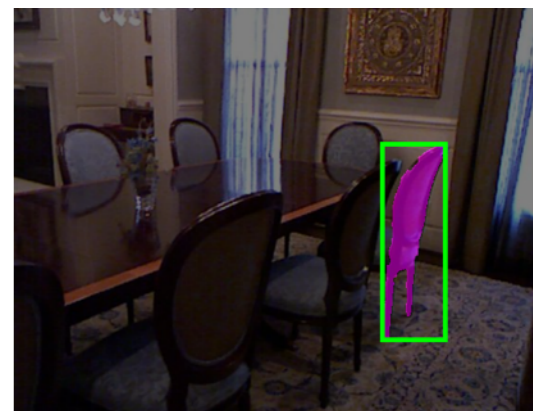
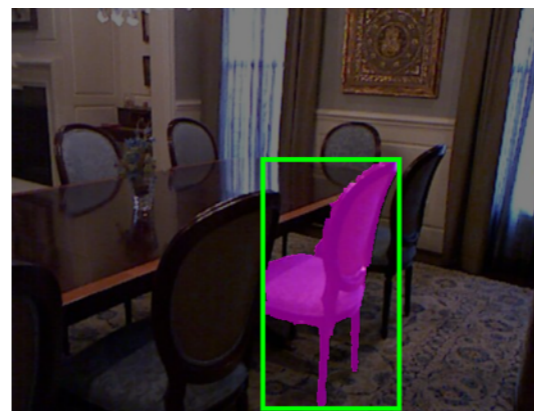
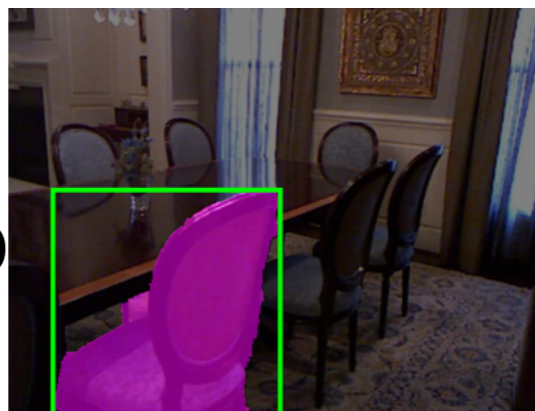
# Examples



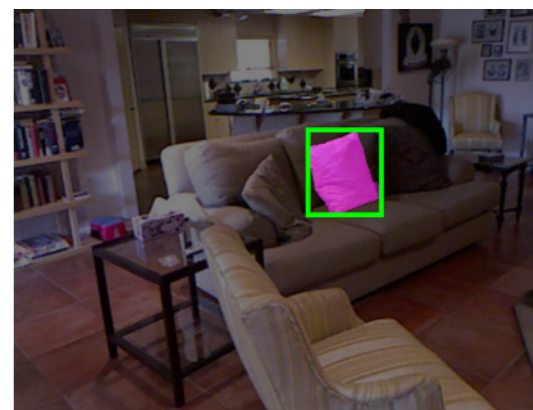
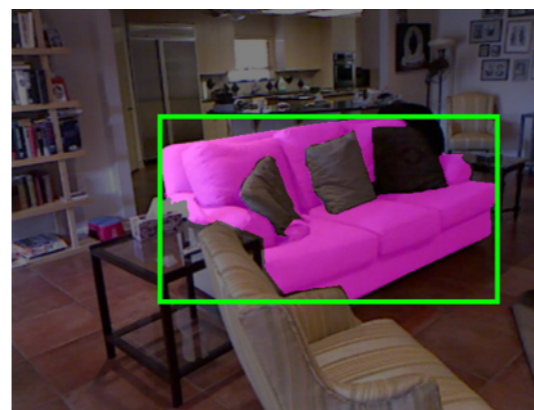
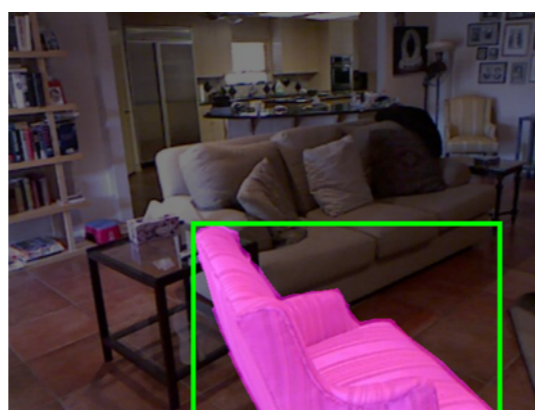
GT Mask



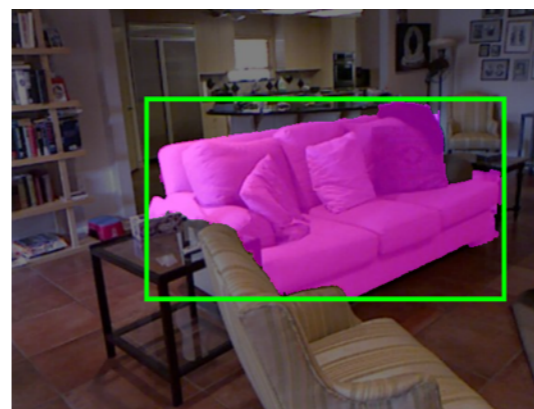
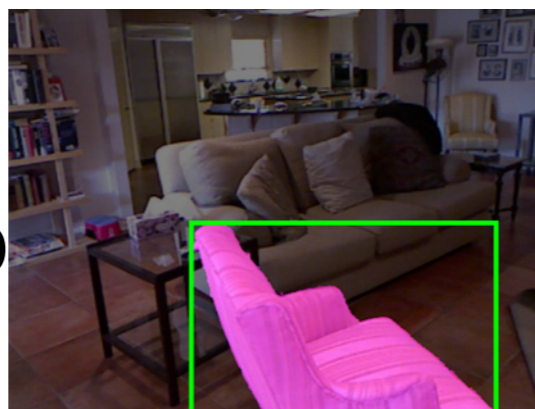
Best Proposal  
@500



GT Mask

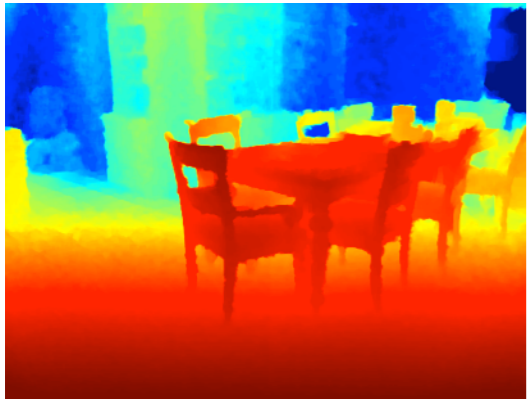


Best Proposal  
@500



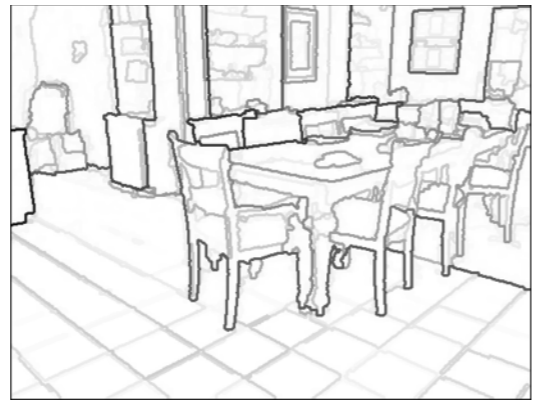
# Overview

## Input

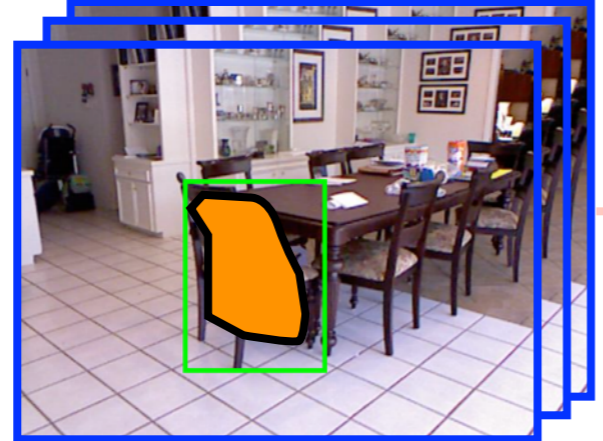


Color and Depth Image Pair

## Re-organization

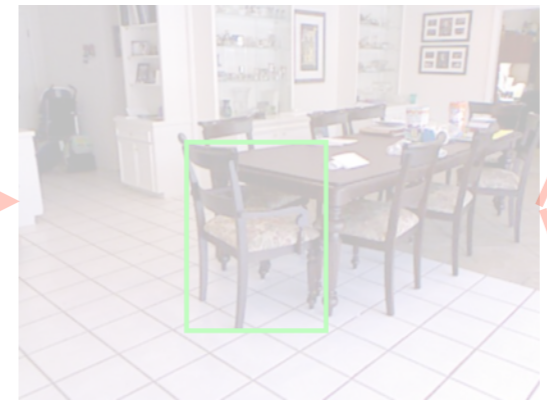


Contour Detection



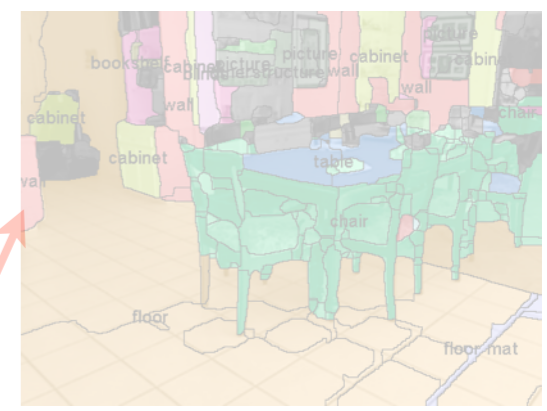
Region Proposal Generation

## Recognition

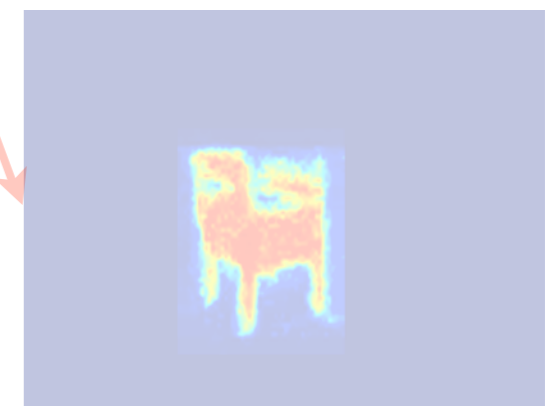


Object Detection

## Extensions



Semantic Segm.

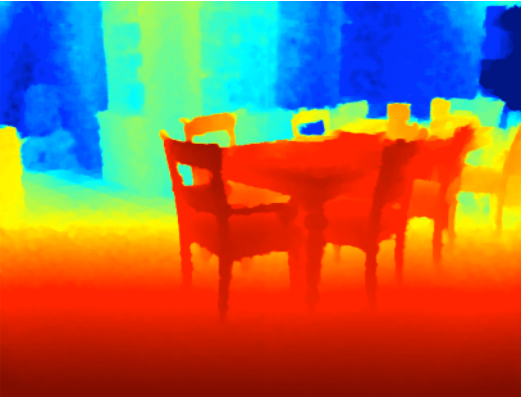


Instance Segm.



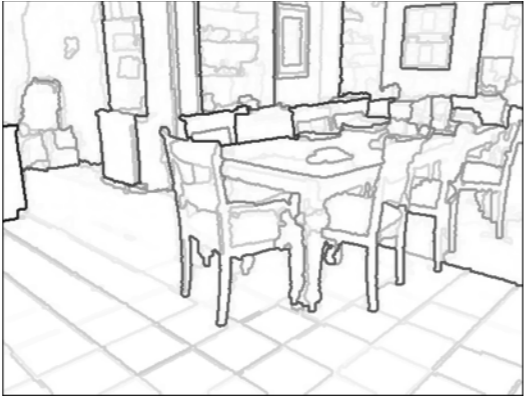
# Overview

## Input

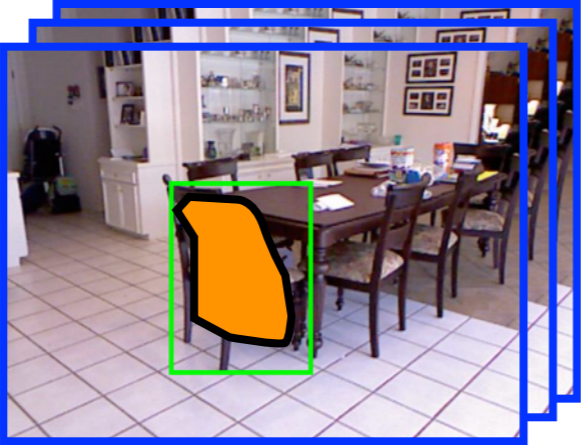


Color and Depth Image Pair

## Re-organization

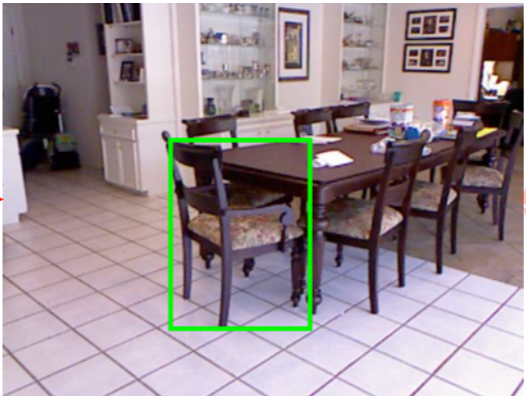


Contour Detection



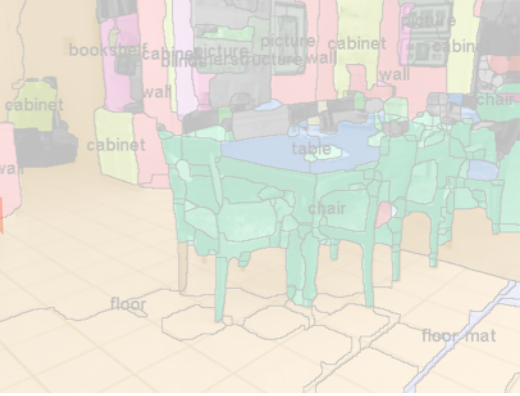
Region Proposal Generation

## Recognition

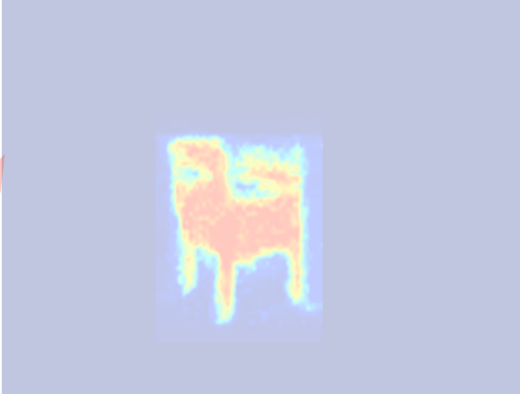


Object Detection

## Extensions



Semantic Segm.



Instance Segm.

# Object Detection

Related Work [RGB]

Sliding Window

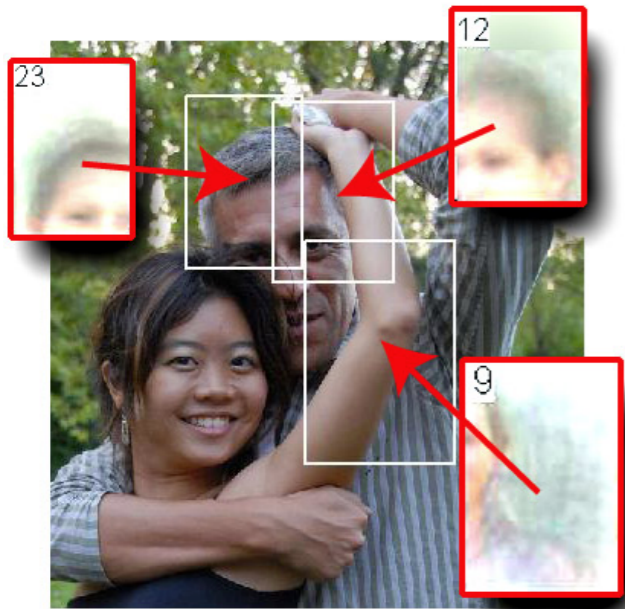
Region Classification

# Object Detection

## Related Work [RGB]

### Sliding Window

Bourdev et al., ECCV 10, Detecting People Using Mutually Consistent Poselet Activations



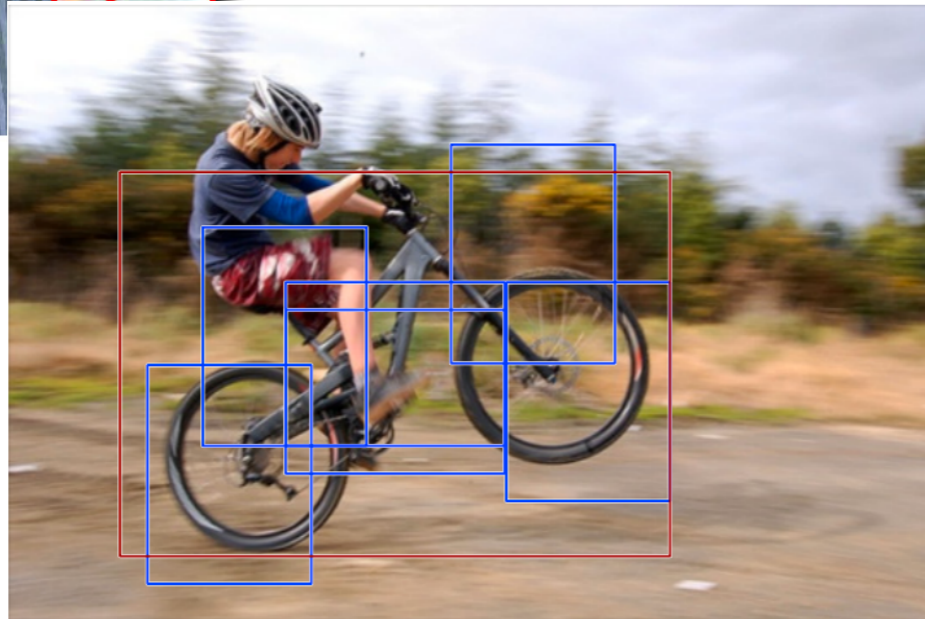
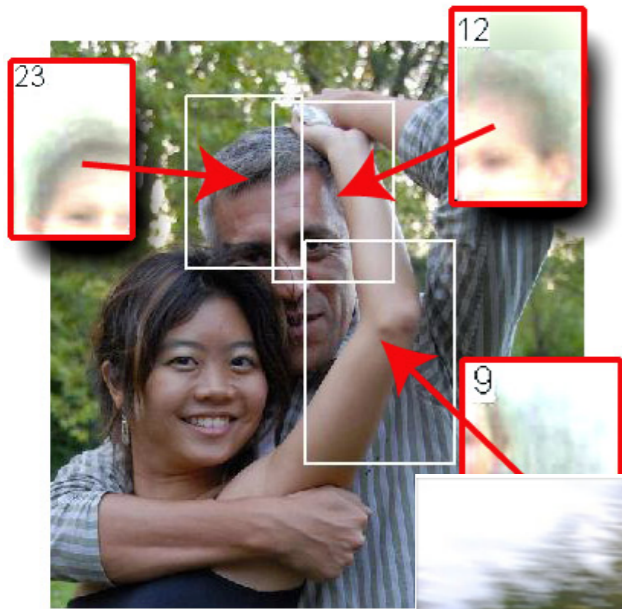
### Region Classification

# Object Detection

## Related Work [RGB]

### Sliding Window

Bourdev et al., ECCV 10, Detecting People Using Mutually Consistent Poselet Activations



Felzenszwalb et al., PAMI 10, Object Detection with Discriminatively Trained Part Based Models

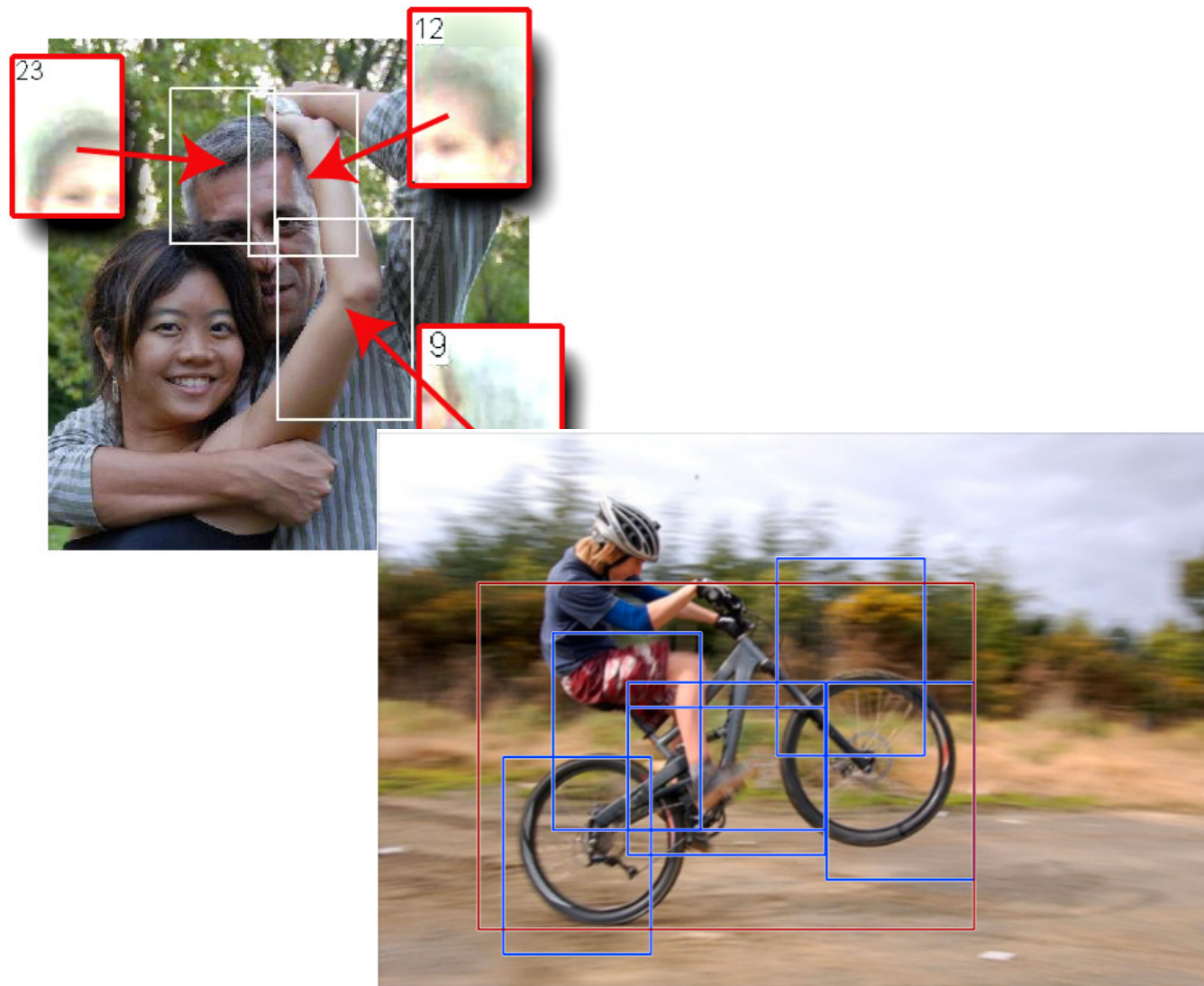
### Region Classification

# Object Detection

## Related Work [RGB]

### Sliding Window

Bourdev et al., ECCV 10, Detecting People Using Mutually Consistent Poselet Activations

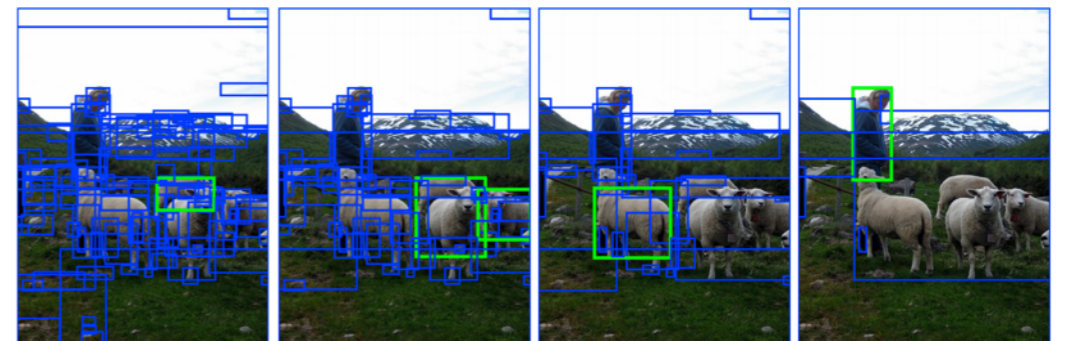


Felzenszwalb et al., PAMI 10, Object Detection with Discriminatively Trained Part Based Models

### Region Classification

Uijlings et al., IJCV 13, Selective Search for Object Recognition

Classifying bottom-up bounding box proposals using rich feature descriptors



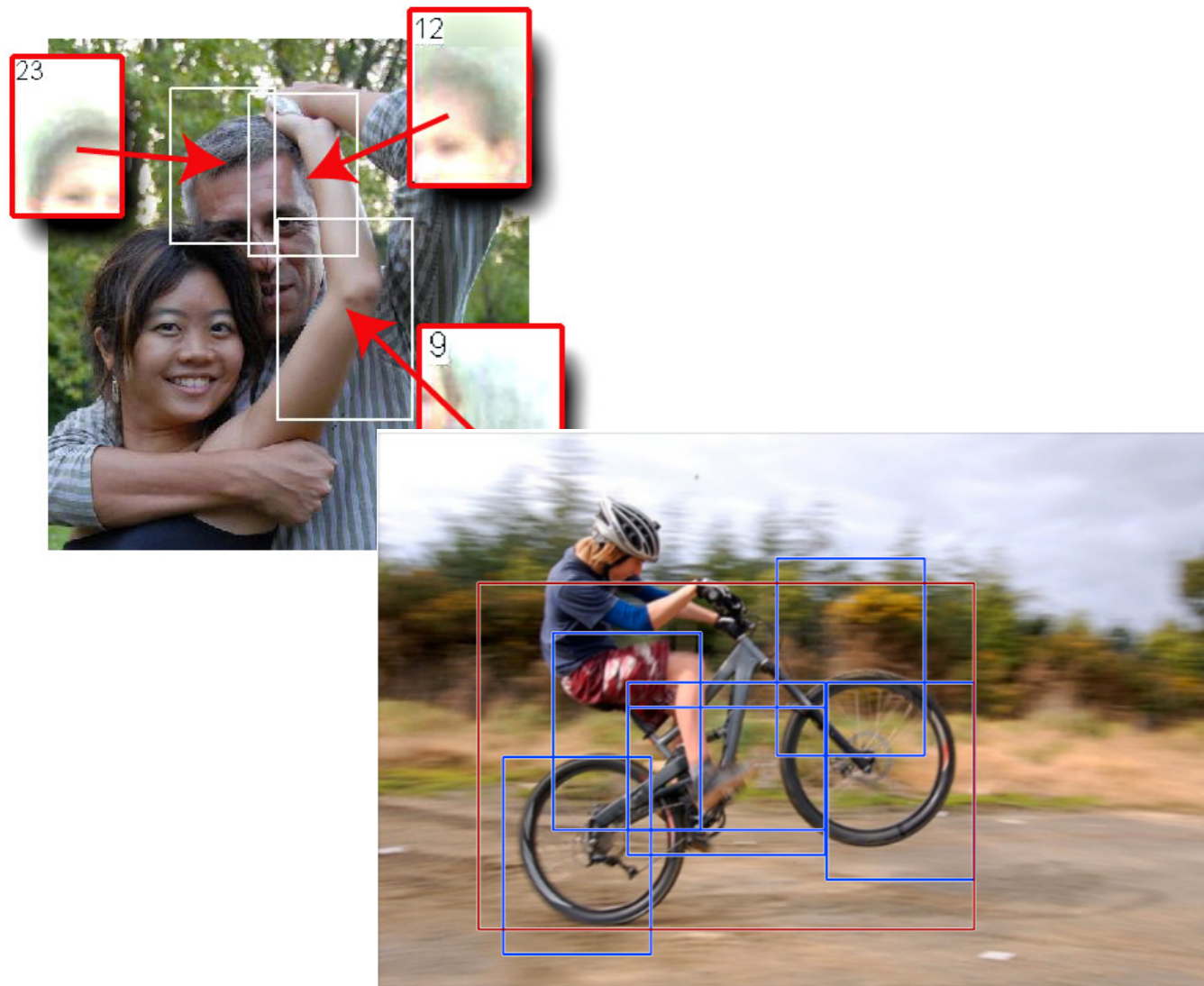


# Object Detection

## Related Work [RGB]

### Sliding Window

Bourdev et al., ECCV 10, Detecting People Using Mutually Consistent Poselet Activations

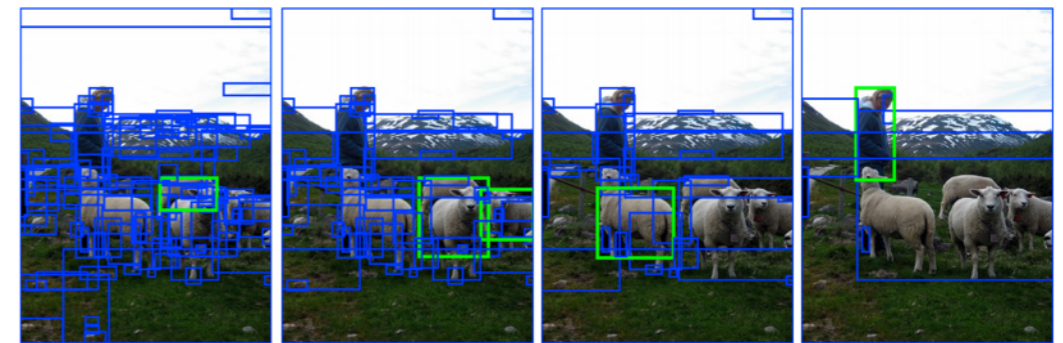


Felzenszwalb et al., PAMI 10, Object Detection with Discriminatively Trained Part Based Models

### Region Classification

Uijlings et al., IJCV 13, Selective Search for Object Recognition

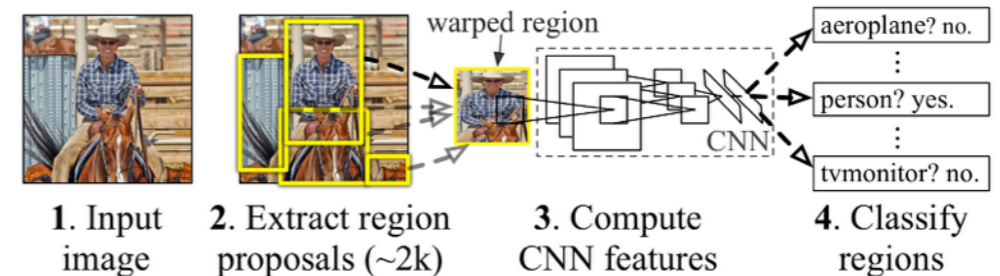
Classifying bottom-up bounding box proposals using rich feature descriptors



Girshick et al., CVPR 14, Rich feature hierarchies for accurate object detection and semantic segmentation

Recently, Convolutional Neural Network (CNN) based features for region proposals have resulted in large gain in performance.

### R-CNN: Regions with CNN features



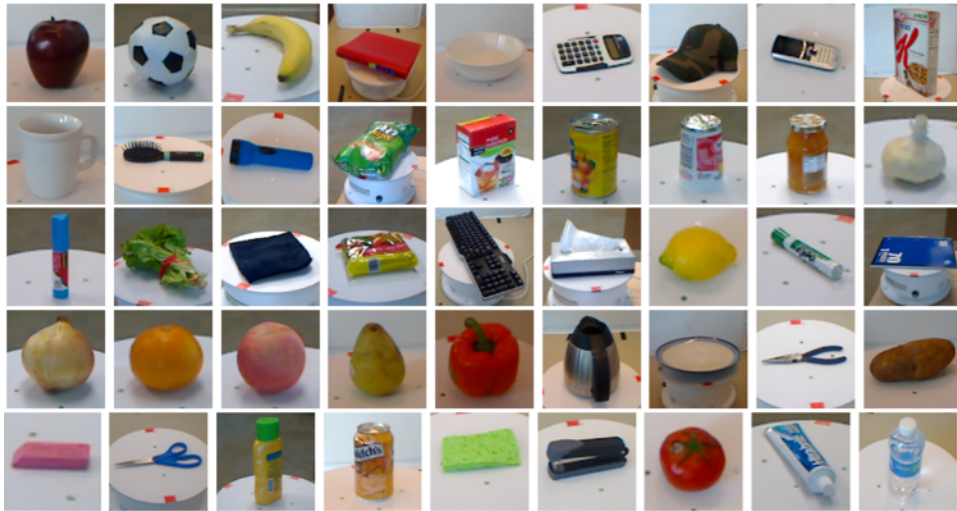
# Object Detection

Related Work [RGB-D, Robotics]

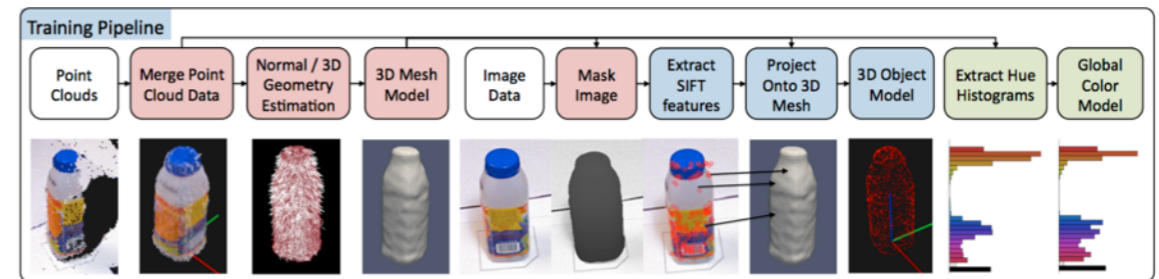
# Object Detection

## Related Work [RGB-D, Robotics]

Lai et al. ICRA 2011, A Large-Scale Hierarchical Multi-View RGB-D Object Dataset: RGB-D DPM, but instances and small table-top objects



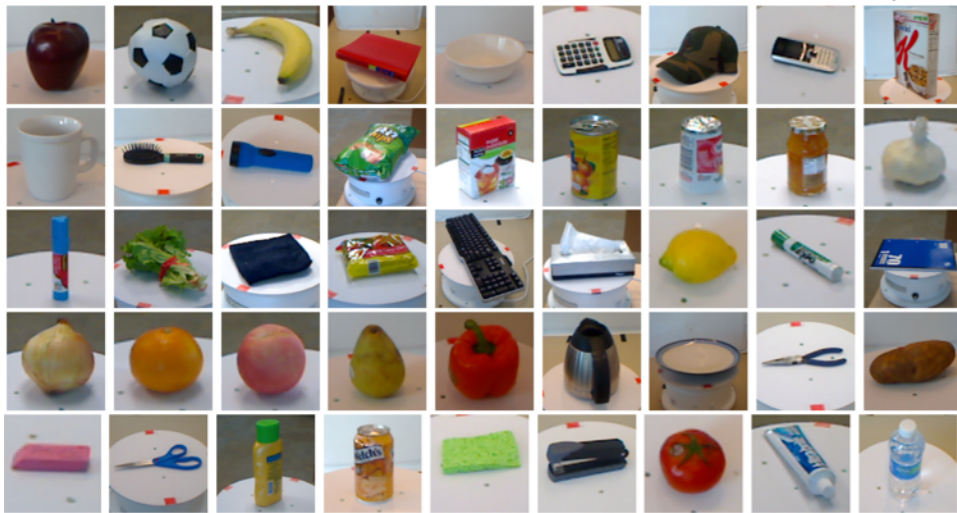
Tang et al. ICRA 2012, A Textured Object Recognition Pipeline for Color and Depth  
**Image Data:** Appearance matching, geometric verification



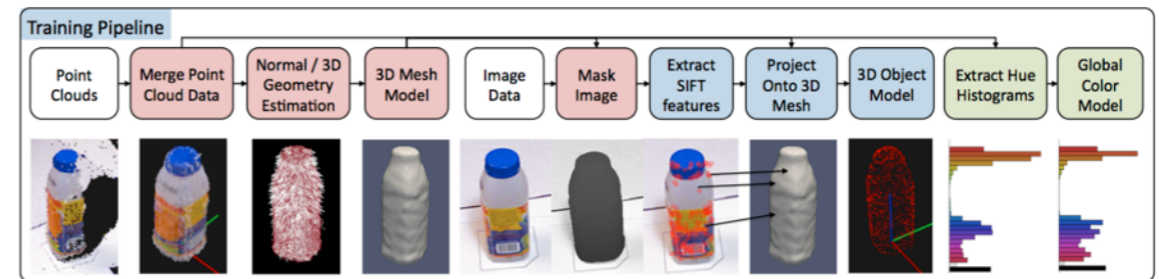
# Object Detection

## Related Work [RGB-D, Robotics]

Lai et al. ICRA 2011, A Large-Scale Hierarchical Multi-View RGB-D Object Dataset: RGB-D DPM, but instances and small table-top objects



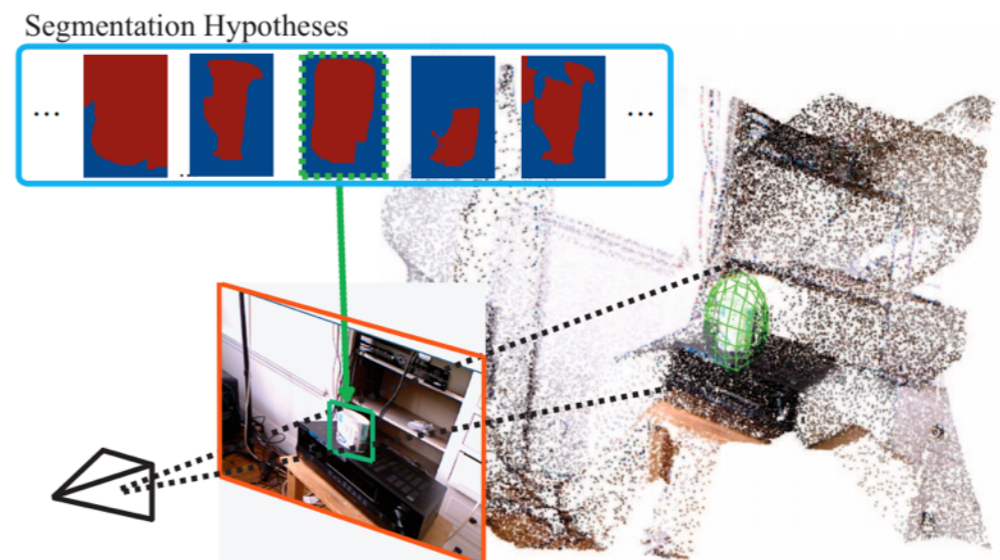
Tang et al. ICRA 2012, A Textured Object Recognition Pipeline for Color and Depth  
**Image Data:** Appearance matching, geometric verification



Janoch et al. ICCV-W 2011, A Category-Level 3-D Object Dataset: Putting the Kinect to Work, Absolute size based pruning and re-scoring with DPMs



Kim et al. CVPR 2013, Accurate Localization of 3D Objects from RGB-D Data using Segmentation Hypotheses, Extension to DPMs to model deformations in 3D



# Object Detection

## Related Work - RCNN

### R-CNN: *Regions with CNN features*



1. Input image

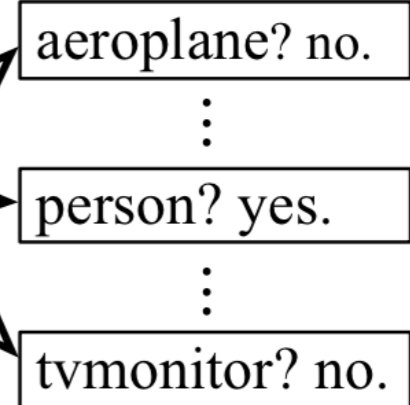
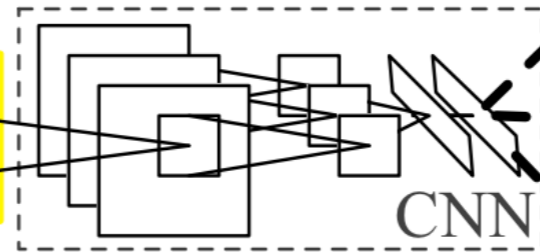


2. Extract region proposals (~2k)

warped region



3. Compute CNN features



4. Classify regions

# Object Detection

## Related Work - RCNN

### R-CNN: *Regions with CNN features*

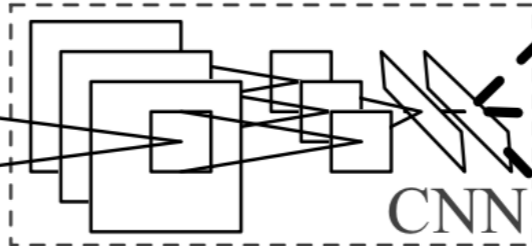


1. Input image

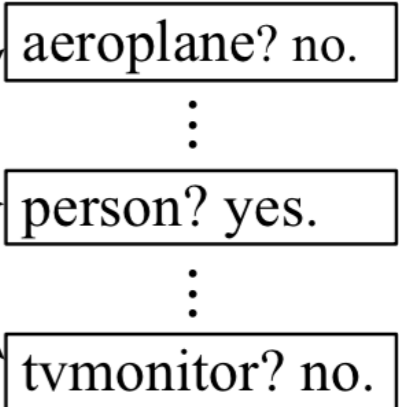


2. Extract region proposals (~2k)

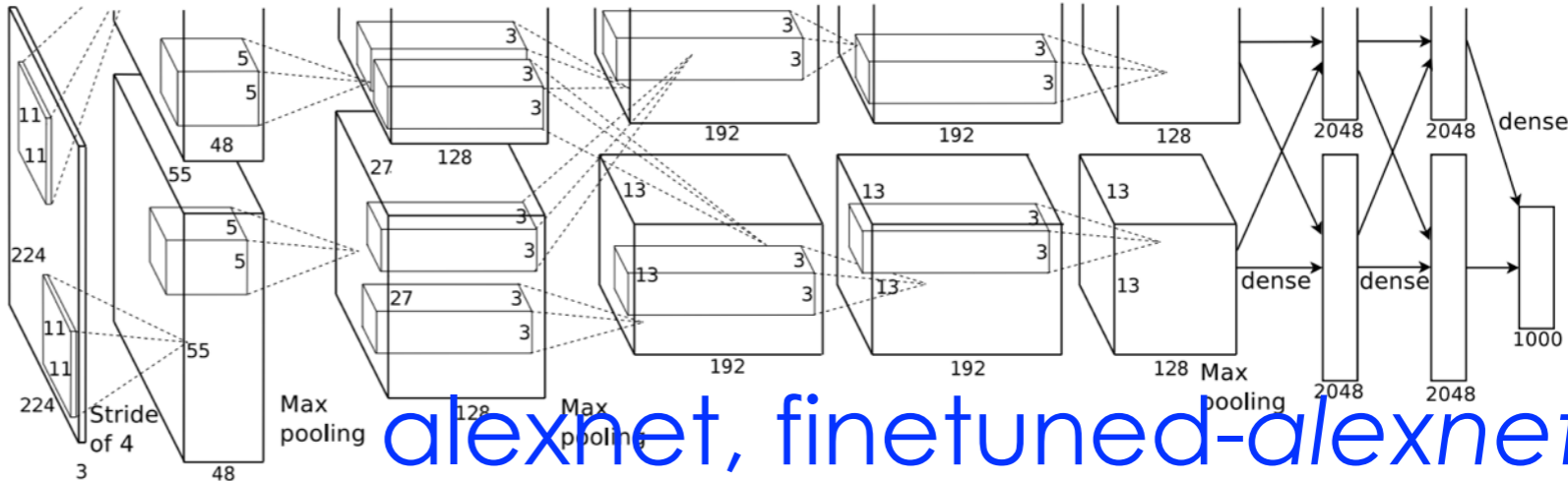
warped region



3. Compute CNN features



4. Classify regions



# Object Detection

## Related Work - RCNN

### R-CNN: *Regions with CNN features*

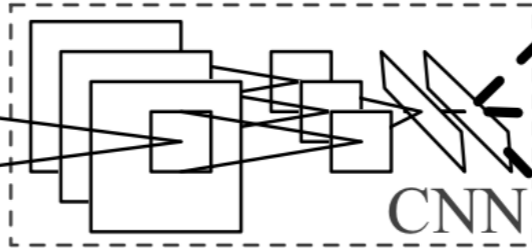


1. Input image

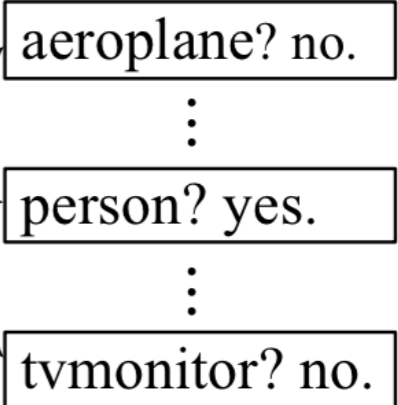


2. Extract region proposals (~2k)

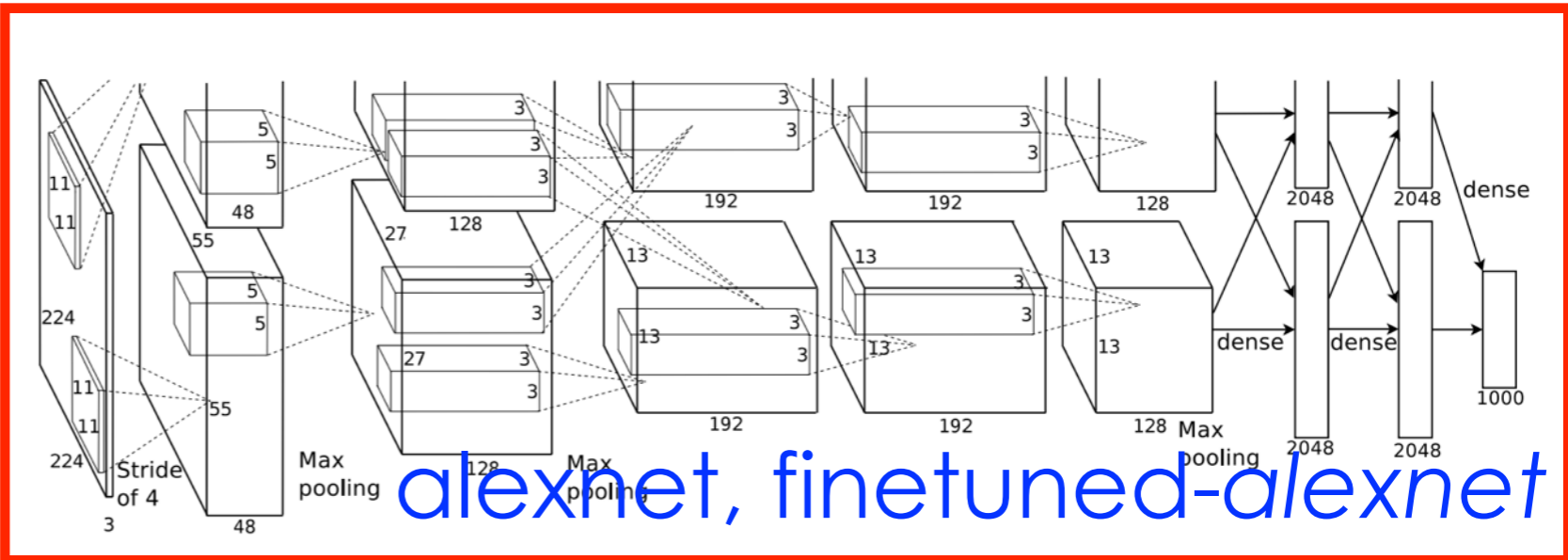
warped region



3. Compute CNN features



4. Classify regions



# Feature Learning for RGB-D Images

# Object Detection

## Key Insights



# Object Detection

## Key Insights

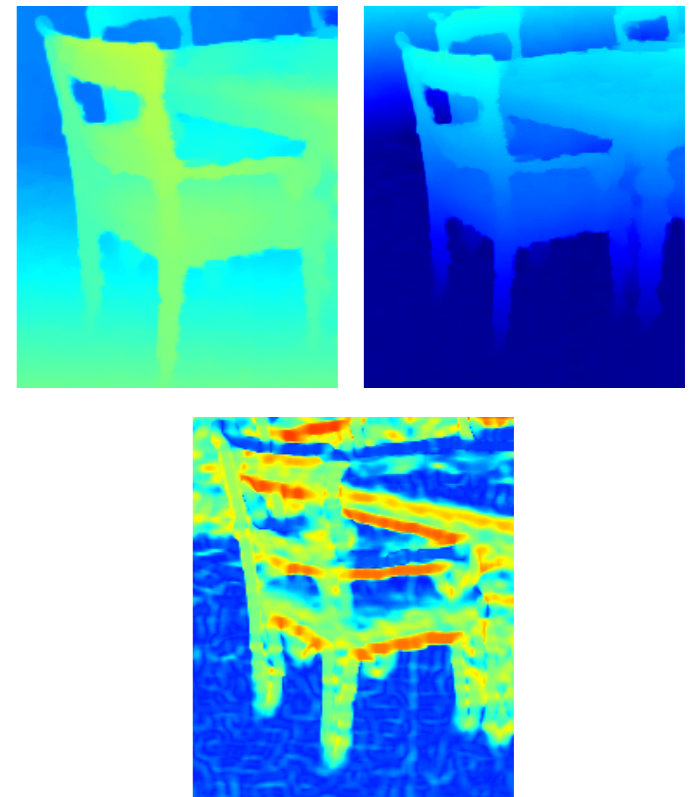
Depth Images are **image-like enough** to use Convolutional Neural Network models

# Object Detection

## Key Insights

Depth Images are **image-like enough** to use Convolutional Neural Network models

**Geocentric embedding** into *Horizontal Disparity, Height Above Ground, and Angle with Gravity (HHA)* works better than just raw disparity



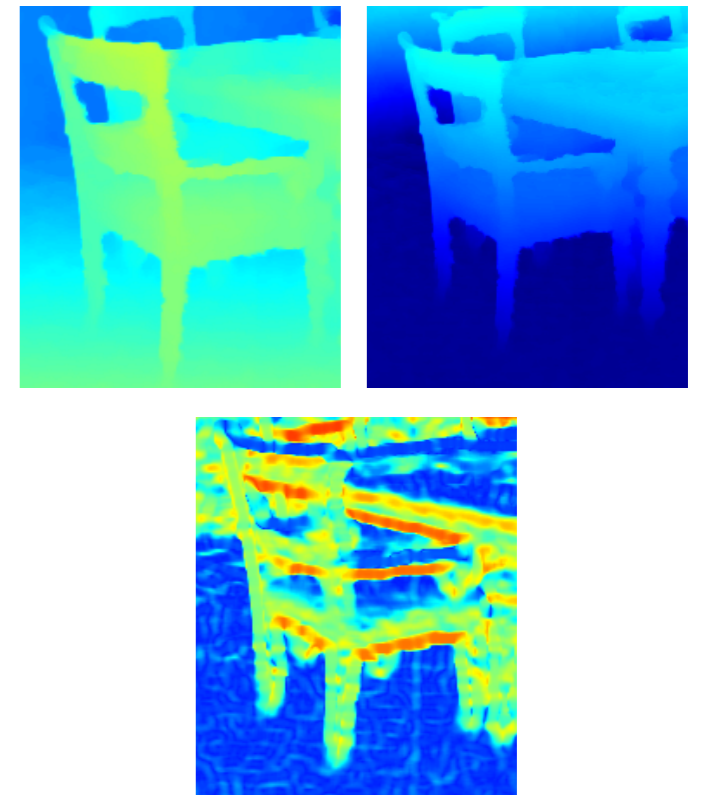
# Object Detection

## Key Insights

Depth Images are **image-like enough** to use Convolutional Neural Network models

**Geocentric embedding** into *Horizontal Disparity, Height Above Ground, and Angle with Gravity (HHA)* works better than just raw disparity

**Synthetic depth data** can help



# Object Detection

# Object Detection

## Experiments

# Object Detection

## Experiments

<b>finetuned?</b>
<b>layer</b>
<b>synthetic?</b>
<b>mAP</b>

# Object Detection

## Experiments

	<b>A</b>
<b>finetuned?</b>	
	DPM
<b>layer</b>	
	RGB
<b>synthetic?</b>	
<b>mAP</b>	8.4

# Object Detection

## Experiments

	<b>A</b>	<b>B</b>
<b>finetuned?</b>		
	DPM	DPM
<b>layer</b>		
	RGB	RGBD
<b>synthetic?</b>		
<b>mAP</b>	8.4	21.7



# Object Detection

## Experiments

	<b>A</b>	<b>B</b>	<b>C</b>
<b>finetuned?</b>			
	DPM	DPM	CNN
<b>layer</b>			fc6
	RGB	RGBD	RGB
<b>synthetic?</b>			
<b>mAP</b>	8.4	21.7	16.4

# Object Detection

## Experiments

	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>
<b>finetuned?</b>				yes
	DPM	DPM	CNN	CNN
<b>layer</b>			fc6	fc6
	RGB	RGBD	RGB	RGB
<b>synthetic?</b>				
<b>mAP</b>	8.4	21.7	16.4	19.7

# Object Detection

## Experiments

	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>
<b>finetuned?</b>				yes	
	DPM	DPM	CNN	CNN	CNN
<b>layer</b>			fc6	fc6	fc6
	RGB	RGBD	RGB	RGB	disparity
<b>synthetic?</b>					
<b>mAP</b>	8.4	21.7	16.4	19.7	11.3

# Object Detection

## Experiments

	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>	<b>F</b>
<b>finetuned?</b>				yes		yes
	DPM	DPM	CNN	CNN	CNN	CNN
<b>layer</b>			fc6	fc6	fc6	fc6
	RGB	RGBD	RGB	RGB	disparity	disparity
<b>synthetic?</b>						
<b>mAP</b>	8.4	21.7	16.4	19.7	11.3	20.1

# Object Detection

## Experiments

	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>	<b>F</b>	<b>G</b>
<b>finetuned?</b>				yes		yes	yes
	DPM	DPM	CNN	CNN	CNN	CNN	CNN
<b>layer</b>			fc6	fc6	fc6	fc6	fc6
	RGB	RGBD	RGB	RGB	disparity	disparity	HHA
<b>synthetic?</b>							
<b>mAP</b>	8.4	21.7	16.4	19.7	11.3	20.1	25.2

# Object Detection

## Experiments

	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>	<b>F</b>	<b>G</b>	<b>H</b>
<b>finetuned?</b>				yes		yes	yes	yes
	DPM	DPM	CNN	CNN	CNN	CNN	CNN	CNN
<b>layer</b>			fc6	fc6	fc6	fc6	fc6	fc6
	RGB	RGBD	RGB	RGB	disparity	disparity	HHA	HHA
<b>synthetic?</b>								2x
<b>mAP</b>	8.4	21.7	16.4	19.7	11.3	20.1	25.2	26.1

# Object Detection

## Experiments

	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>	<b>F</b>	<b>G</b>	<b>H</b>	<b>I</b>
<b>finetuned?</b>				yes		yes	yes	yes	yes
	DPM	DPM	CNN	CNN	CNN	CNN	CNN	CNN	CNN
<b>layer</b>			fc6	fc6	fc6	fc6	fc6	fc6	pool5
	RGB	RGBD	RGB	RGB	disparity	disparity	HHA	HHA	HHA
<b>synthetic?</b>								2x	2x
<b>mAP</b>	8.4	21.7	16.4	19.7	11.3	20.1	25.2	26.1	21.9

# Object Detection

## Experiments

	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>	<b>F</b>	<b>G</b>	<b>H</b>	<b>I</b>	<b>J</b>
<b>finetuned?</b>				yes		yes	yes	yes	yes	yes
	DPM	DPM	CNN	CNN	CNN	CNN	CNN	CNN	CNN	CNN
<b>layer</b>			fc6	fc6	fc6	fc6	fc6	fc6	pool5	fc7
	RGB	RGBD	RGB	RGB	disparity	disparity	HHA	HHA	HHA	HHA
<b>synthetic?</b>								2x	2x	2x
<b>mAP</b>	8.4	21.7	16.4	19.7	11.3	20.1	25.2	26.1	21.9	25.3



# Object Detection

## Experiments

	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>	<b>F</b>	<b>G</b>	<b>H</b>	<b>I</b>	<b>J</b>	<b>K</b>
<b>finetuned?</b>				yes		yes	yes	yes	yes	yes	yes
	DPM	DPM	CNN	CNN	CNN	CNN	CNN	CNN	CNN	CNN	CNN
<b>layer</b>			fc6	fc6	fc6	fc6	fc6	fc6	pool5	fc7	fc6
	RGB	RGBD	RGB	RGB	disparity	disparity	HHA	HHA	HHA	HHA	RGB +HHA
<b>synthetic?</b>								2x	2x	2x	2x
<b>mAP</b>	8.4	21.7	16.4	19.7	11.3	20.1	25.2	26.1	21.9	25.3	<b>32.5</b>

# Object Detection

## Experiments

	A	B	C	D	E	F	G	H	I	J	K
<b>finetuned?</b>				yes		yes	yes	yes	yes	yes	yes
<b>layer</b>	DPM	DPM	CNN	CNN	CNN	CNN	CNN	CNN	CNN	CNN	CNN
<b>layer</b>			fc6	fc6	fc6	fc6	fc6	fc6	pool5	fc7	fc6
<b>synthetic?</b>	RGB	RGBD	RGB	RGB	disparity	disparity	HHA	HHA	HHA	HHA	RGB+HHA
<b>mAP</b>	8.4	21.7	16.4	19.7	11.3	20.1	25.2	26.1	21.9	25.3	<b>32.5</b>

## Test Set

	mean	bath tub	bed	book shelf	box	chair	counter	desk	door	dresser	garbage bin	lamp	monitor	night stand	pillow	sink	sofa	table	television	toilet
<b>RGB DPM</b>	<b>9</b>	1	28	9	0	8	7	1	3	1	7	22	10	9	4	6	9	6	6	34
<b>RGBD DPM</b>	<b>24</b>	19	56	18	<b>1</b>	24	24	6	10	16	27	27	35	33	21	23	34	17	20	45
<b>RGB RCNN</b>	<b>22</b>	17	45	28	<b>1</b>	26	30	10	16	19	16	28	32	17	11	17	29	13	27	44
<b>Our</b>	<b>37</b>	<b>44</b>	<b>71</b>	<b>33</b>	<b>1</b>	<b>43</b>	<b>44</b>	<b>15</b>	<b>24</b>	<b>30</b>	<b>39</b>	<b>37</b>	<b>53</b>	<b>40</b>	<b>35</b>	<b>36</b>	<b>54</b>	<b>24</b>	<b>38</b>	<b>47</b>

# Object Detection

## Experiments

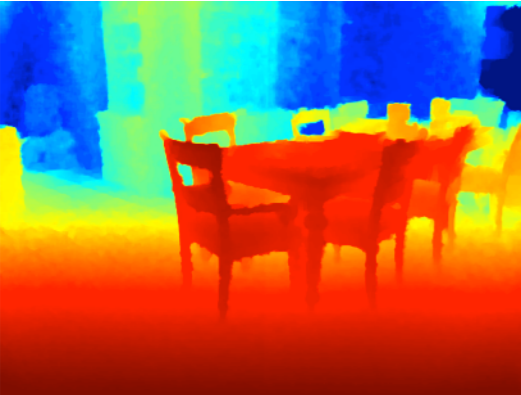
	A	B	C	D	E	F	G	H	I	J	K
<b>finetuned?</b>				yes		yes	yes	yes	yes	yes	yes
<b>layer</b>	DPM	DPM	CNN	CNN	CNN	CNN	CNN	CNN	CNN	CNN	CNN
<b>layer</b>			fc6	fc6	fc6	fc6	fc6	fc6	pool5	fc7	fc6
<b>synthetic?</b>	RGB	RGBD	RGB	RGB	disparity	disparity	HHA	HHA	HHA	HHA	RGB+HHA
<b>mAP</b>	8.4	21.7	16.4	19.7	11.3	20.1	25.2	26.1	21.9	25.3	<b>32.5</b>

## Test Set

	mean	bath tub	bed	book shelf	box	chair	counter	desk	door	dresser	garbage bin	lamp	monitor	night stand	pillow	sink	sofa	table	television	toilet
<b>RGB DPM</b>	<b>9</b>	1	28	9	0	8	7	1	3	1	7	22	10	9	4	6	9	6	6	34
<b>RGBD DPM</b>	<b>24</b>	19	56	18	1	24	24	6	10	16	27	27	35	33	21	23	34	17	20	45
<b>RGB RCNN</b>	<b>22</b>	17	45	28	1	26	30	10	16	19	16	28	32	17	11	17	29	13	27	44
<b>Our</b>	<b>37</b>	<b>44</b>	<b>71</b>	<b>33</b>	<b>1</b>	<b>43</b>	<b>44</b>	<b>15</b>	<b>24</b>	<b>30</b>	<b>39</b>	<b>37</b>	<b>53</b>	<b>40</b>	<b>35</b>	<b>36</b>	<b>54</b>	<b>24</b>	<b>38</b>	<b>47</b>

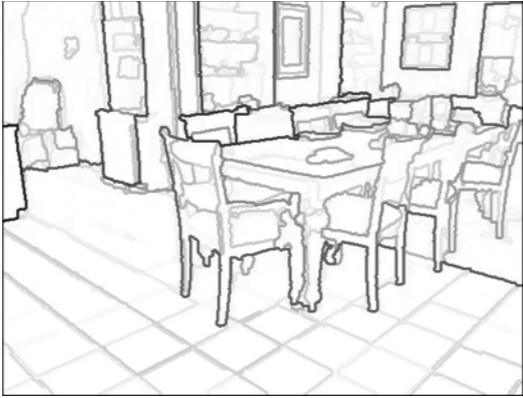
# Overview

## Input

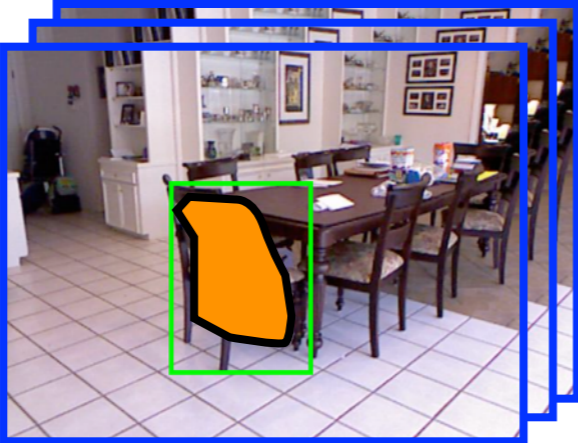


Color and Depth Image Pair

## Re-organization

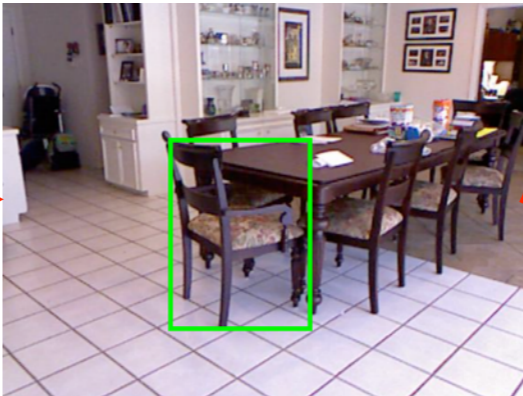


Contour Detection



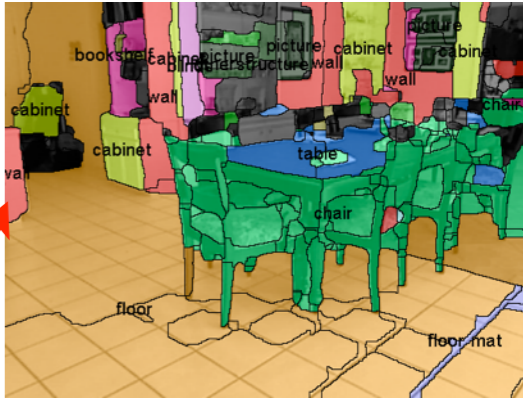
Region Proposal Generation

## Recognition

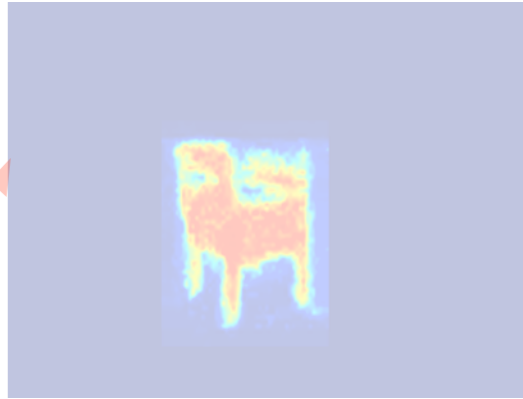


Object Detection

## Extensions



Semantic Segm.



Instance Segm.

# Object Detection

## For Semantic Segmentation

# Object Detection

## For Semantic Segmentation

Use output from object detectors to compute **additional features** for superpixels

# Object Detection

## For Semantic Segmentation

Use output from object detectors to compute **additional features** for superpixels

## Feature Computation

# Object Detection

## For Semantic Segmentation

Use output from object detectors to compute **additional features** for superpixels

## Feature Computation



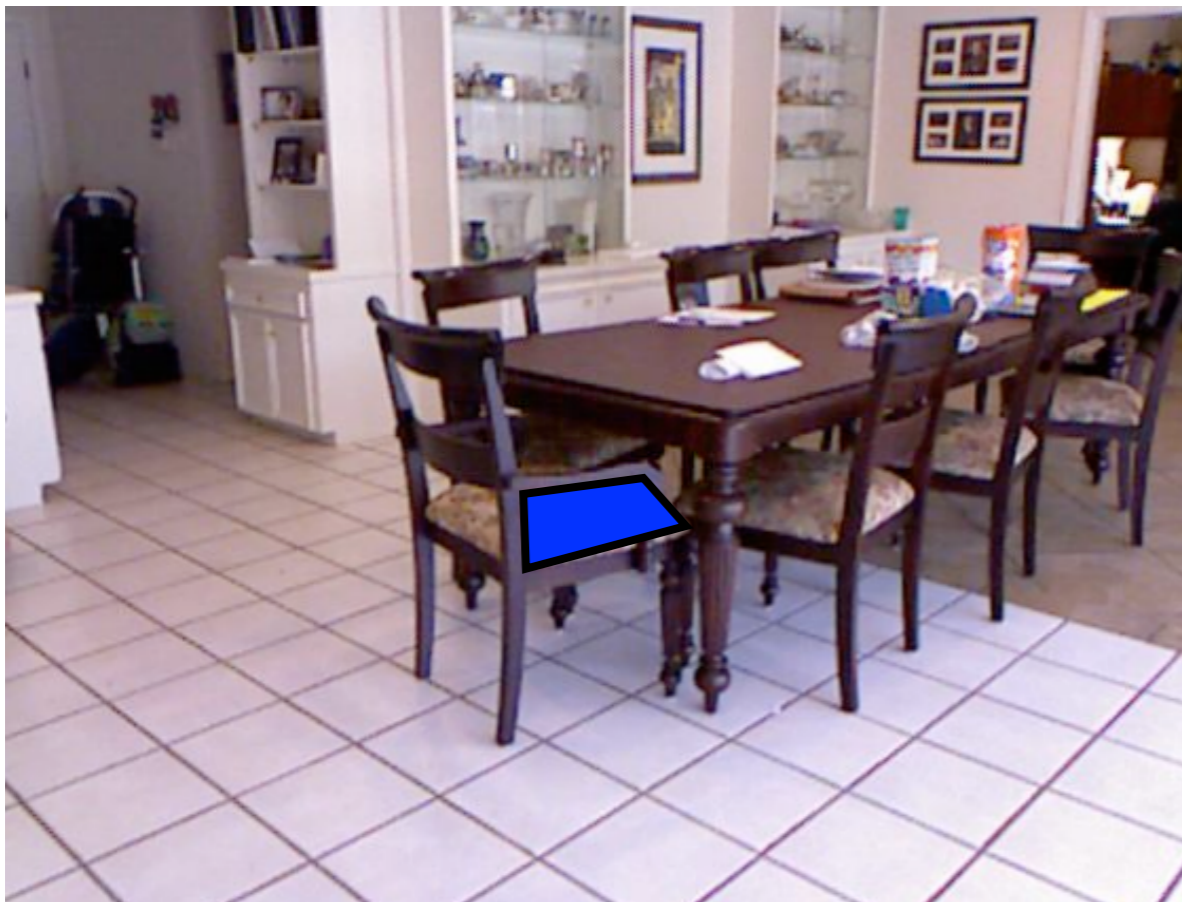


# Object Detection

## For Semantic Segmentation

Use output from object detectors to compute **additional features** for superpixels

## Feature Computation

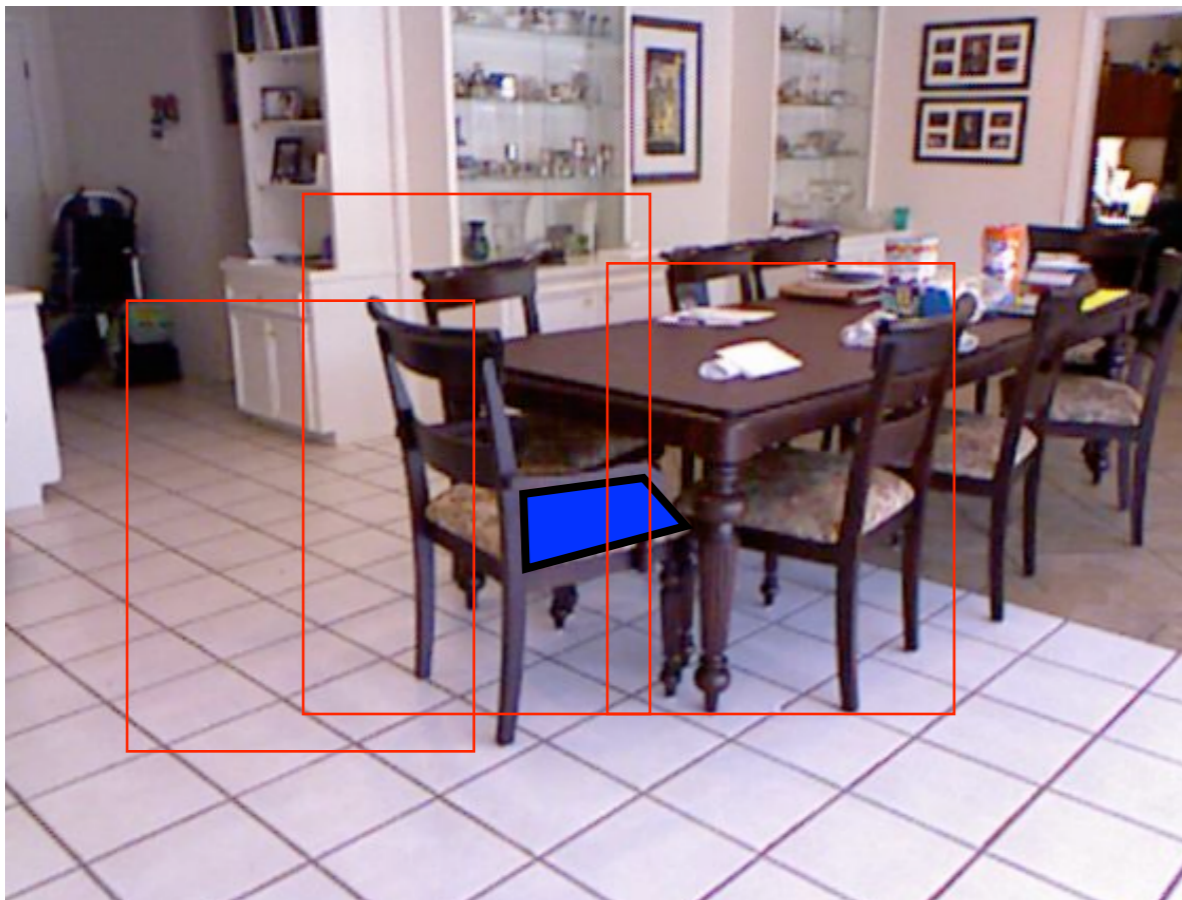


# Object Detection

## For Semantic Segmentation

Use output from object detectors to compute **additional features** for superpixels

## Feature Computation

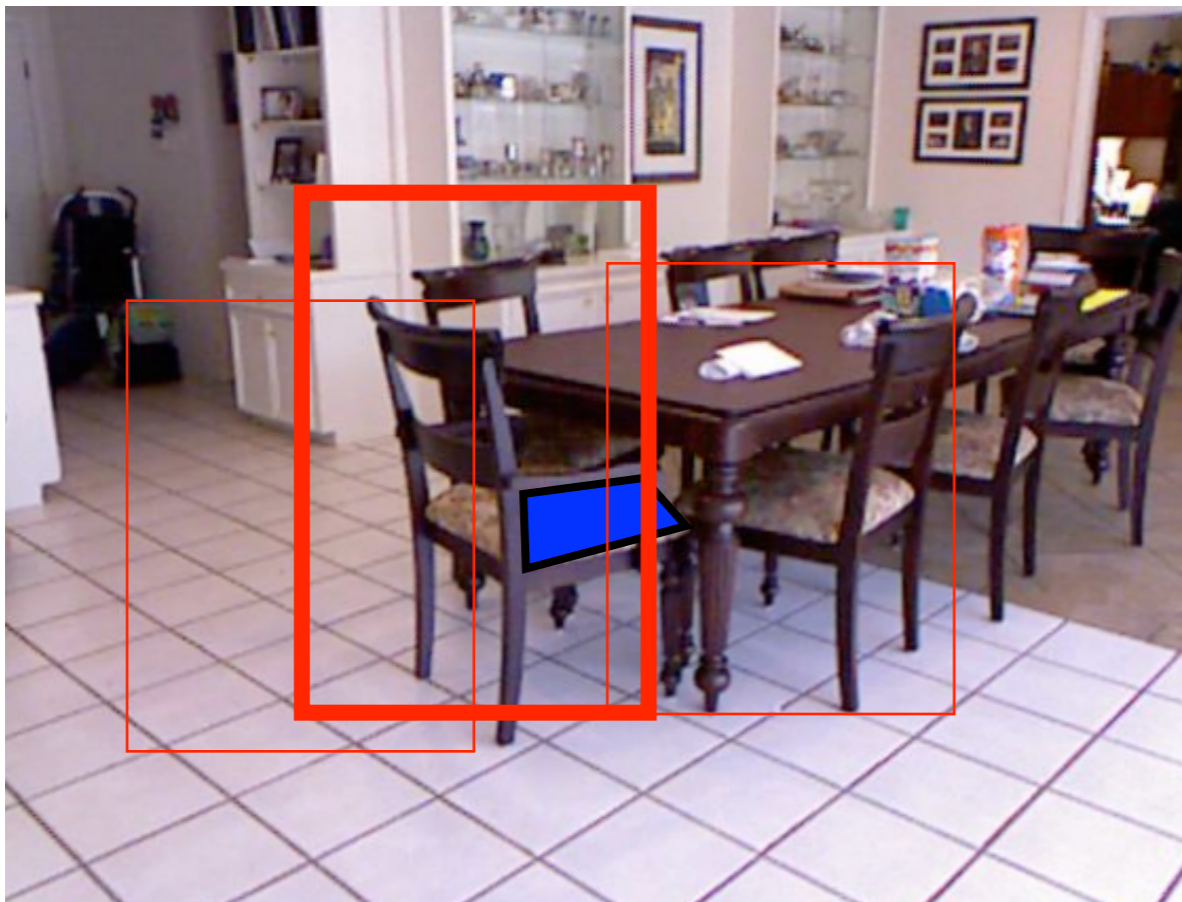


# Object Detection

## For Semantic Segmentation

Use output from object detectors to compute **additional features** for superpixels

## Feature Computation



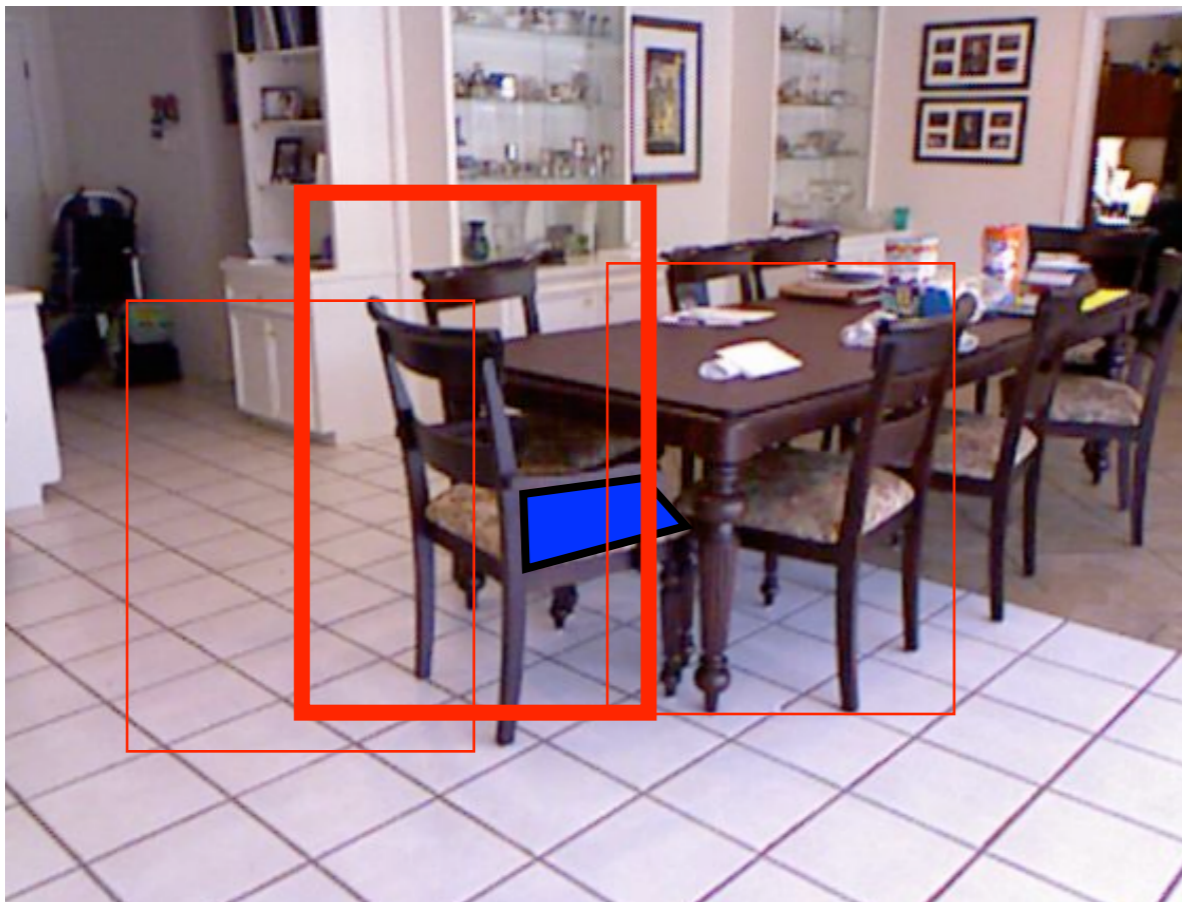
1. Highest scoring detection

# Object Detection

## For Semantic Segmentation

Use output from object detectors to compute **additional features** for superpixels

## Feature Computation



1. Highest scoring detection
2. Use as features for the superpixel
  - detection score
  - overlap
  - difference in mean depth of superpixel and detection
  - non-linear combinations

# Object Detection

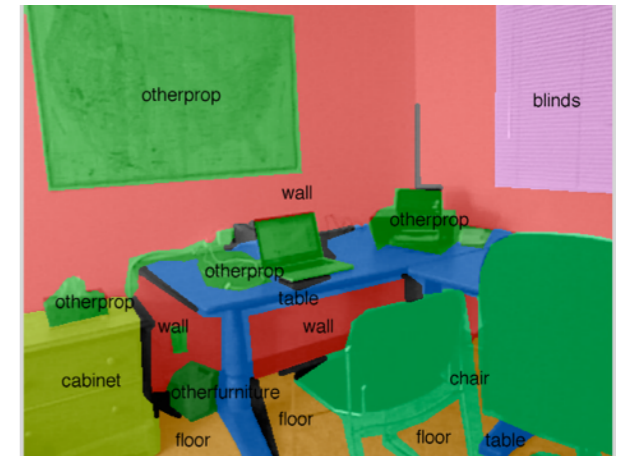
## For Semantic Segmentation (Performance)

### 40 Class Task

**Scene Surfaces** - Floors, walls, ceiling, windows, doors, ...

**Furniture** - Beds, chairs, sofa, table, desks, ...

**Objects** - Pillow, books, bottles, ...



Ground Truth 40

	Silberman et al. ECCV 12	Ren et al. CVPR 12	Gupta et al. CVPR 13	Gupta et al. (13) + RGB-D DPM	Gupta et al. (13) + Our Obj Det.
<b>fwavacc</b>	38.2	37.6	43.4	45.2	<b>47</b>
<b>avacc</b>	19	20.5	24.3	27.3	<b>28.6</b>
<b>mean (maxIU)</b>	-	21.4	27.9	29.6	<b>31.3</b>
<b>pixacc</b>	54.6	49.3	57.9	59	<b>60.3</b>
<b>obj avg</b>	18.4	21.1	26.4	31.1	<b>35.1</b>

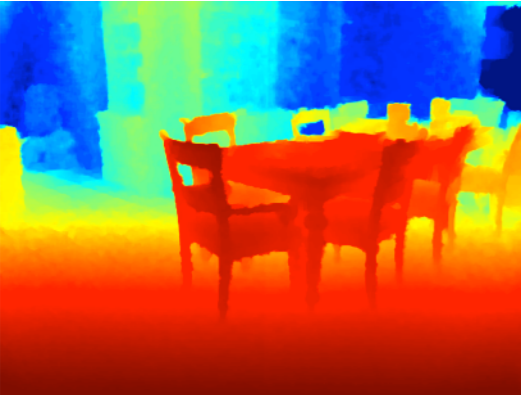
Silberman et al., ECCV12, Indoor segmentation and support inference from RGBD images.

Ren et al., CVPR12, RGB-(D) scene labeling: Features and algorithms

Gupta et al., CVPR13, Perceptual Organization and Recognition of Indoor Scenes from RGB-D Images.

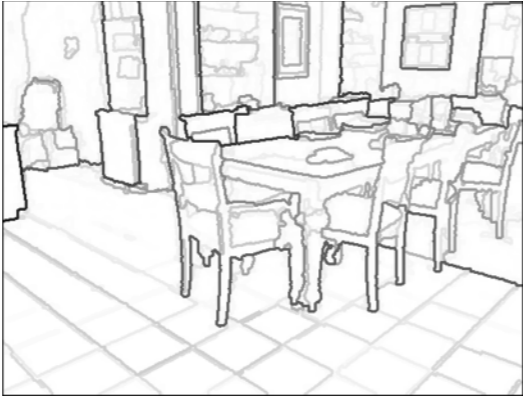
# Overview

## Input

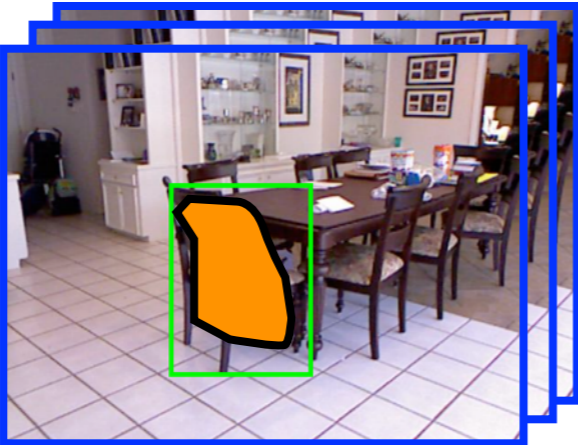


Color and Depth Image Pair

## Re-organization

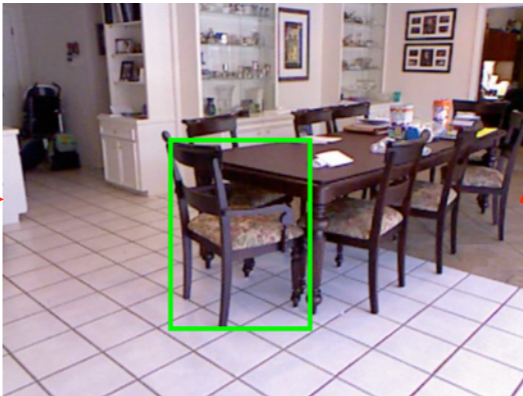


Contour Detection



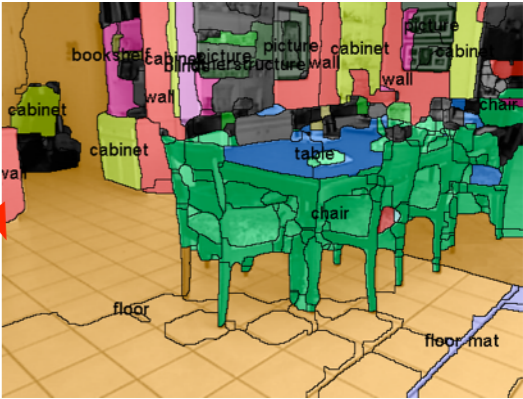
Region Proposal Generation

## Recognition

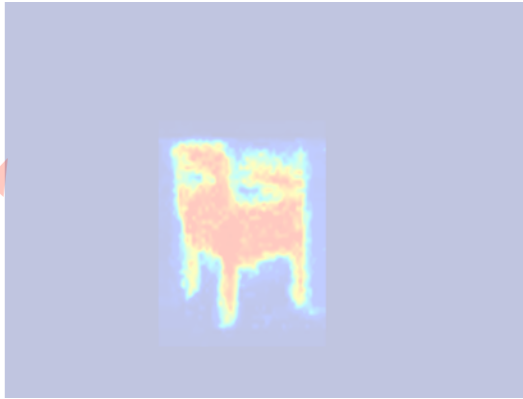


Object Detection

## Extensions



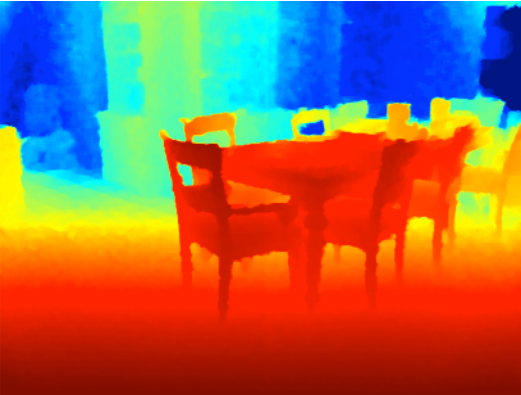
Semantic Segm.



Instance Segm.

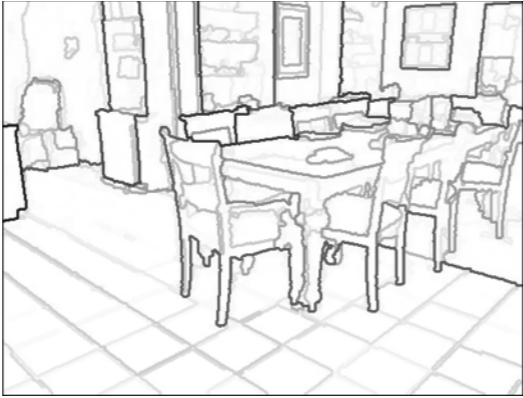
# Overview

## Input

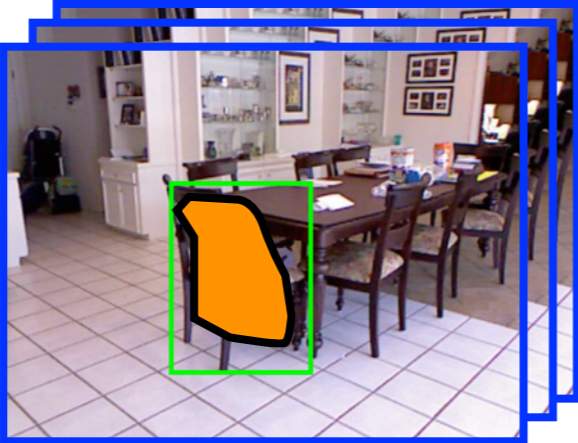


Color and Depth Image Pair

## Re-organization

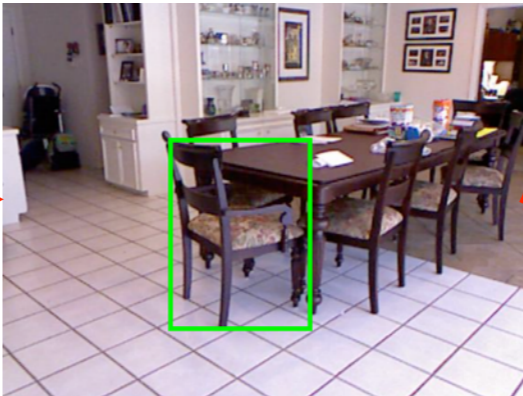


Contour Detection



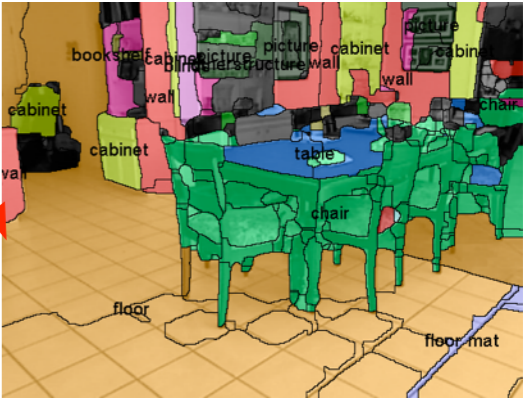
Region Proposal Generation

## Recognition

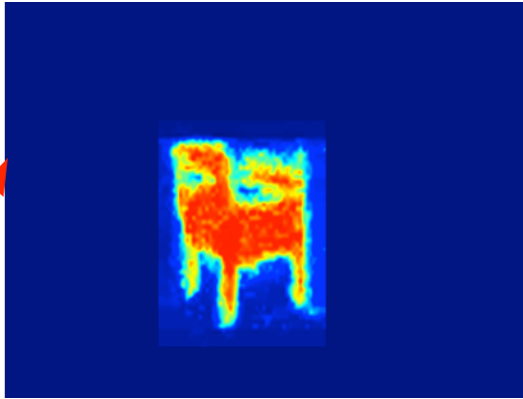


Object Detection

## Extensions



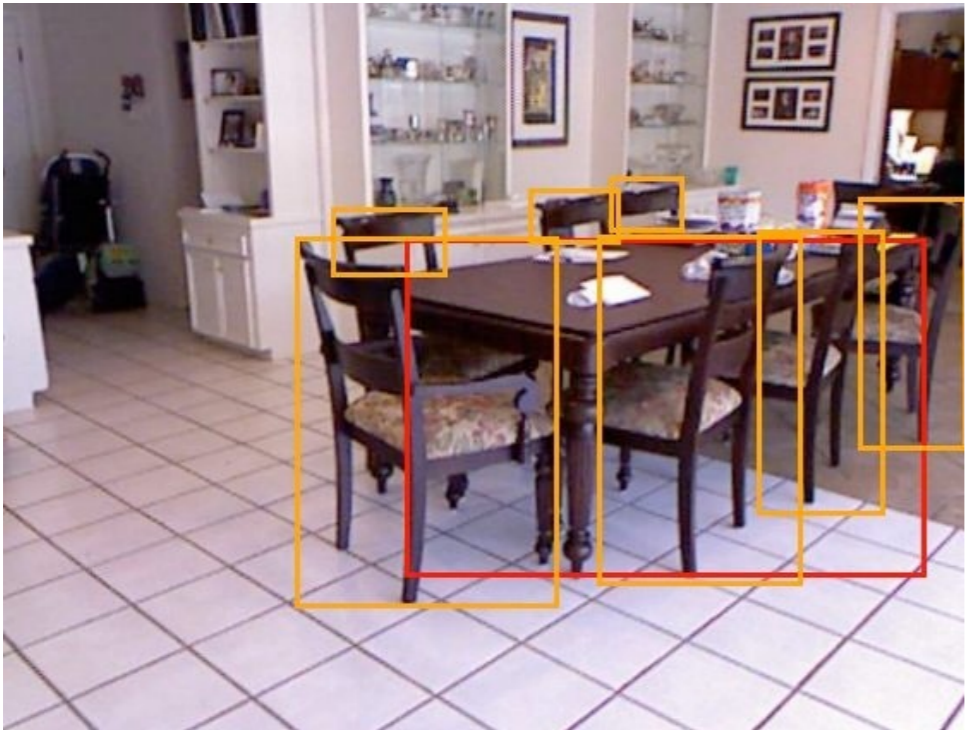
Semantic Segm.



Instance Segm.

# Object Detection

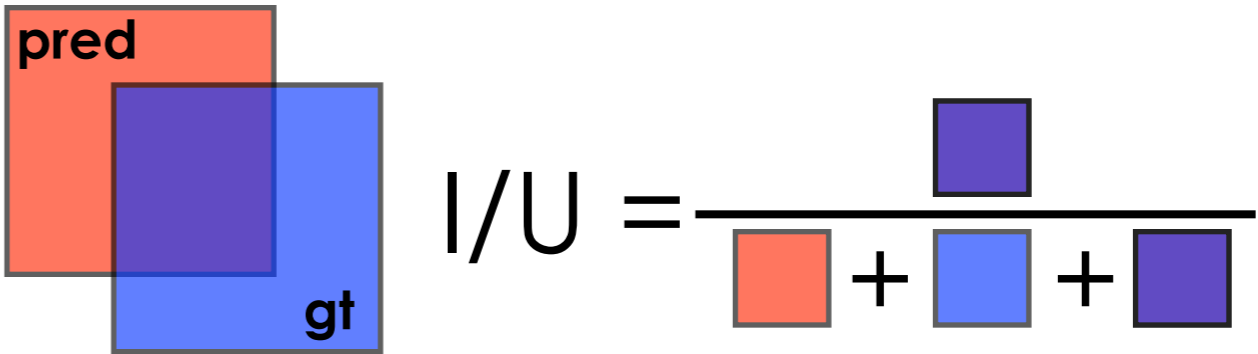
## Task



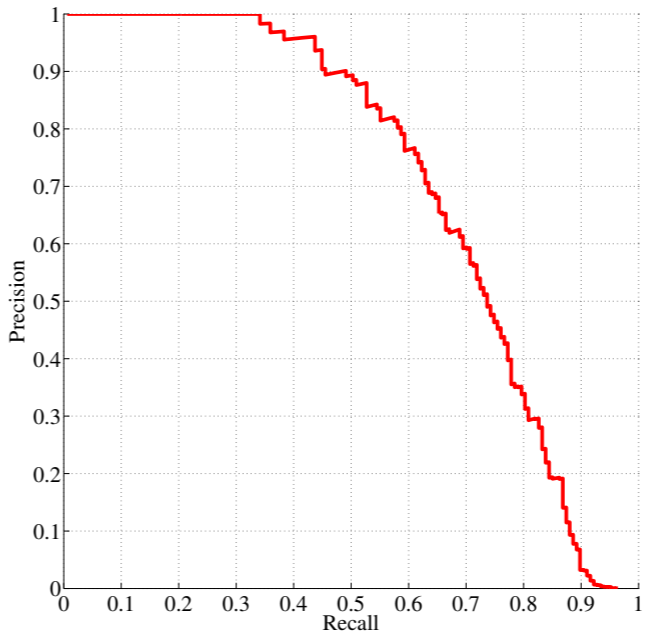
## Metric

# Average Precision

assign detection boxes to ground truth boxes based on I/U overlap



assign true positive, ...  
to compute precision and recall



**AP** = area under the Precision recall curve



# Instance Segmentation

## Task



# Instance Segmentation

## Task



# Instance Segmentation

## Task



# Instance Segmentation

## Task



# Instance Segmentation

## Task



# Instance Segmentation

## Task



# Instance Segmentation

## Task



# Instance Segmentation

## Task





# Instance Segmentation

## Task



# Instance Segmentation

Task

Metric



# Instance Segmentation

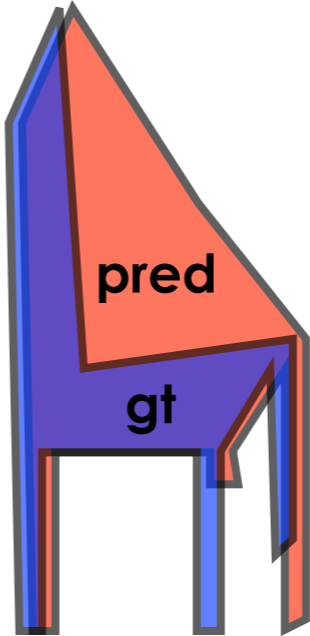
## Task



## Metric

$AP^r$

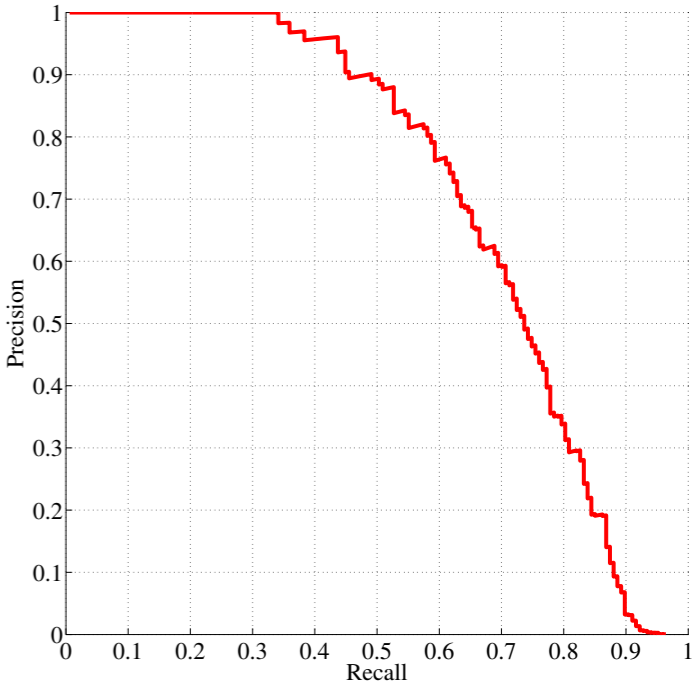
assign detection **regions** to ground truth **regions** based on **region I/U** overlap



$$I/U = \frac{\text{purple square}}{\text{red square} + \text{blue square} + \text{purple square}}$$

assign true positive, ...  
to compute Precision  
and Recall

$AP^r$  = area under the  
Precision-Recall curve



# Instance Segmentation

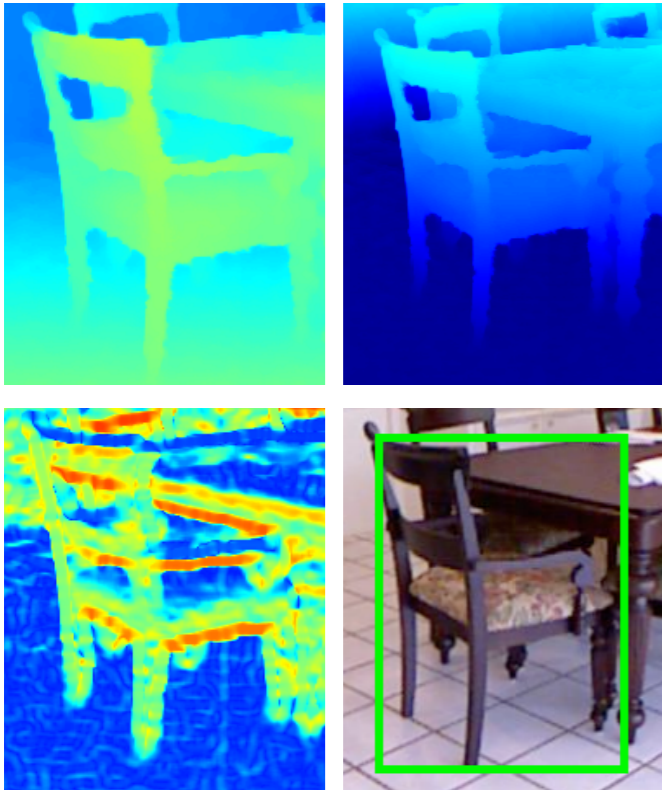
# Instance Segmentation

Predict each pixel in the detection window to be foreground or background

# Instance Segmentation

Predict each pixel in the detection window to be foreground or background

Compute Feature Channels

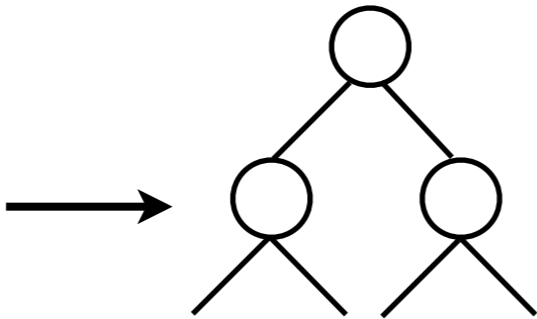
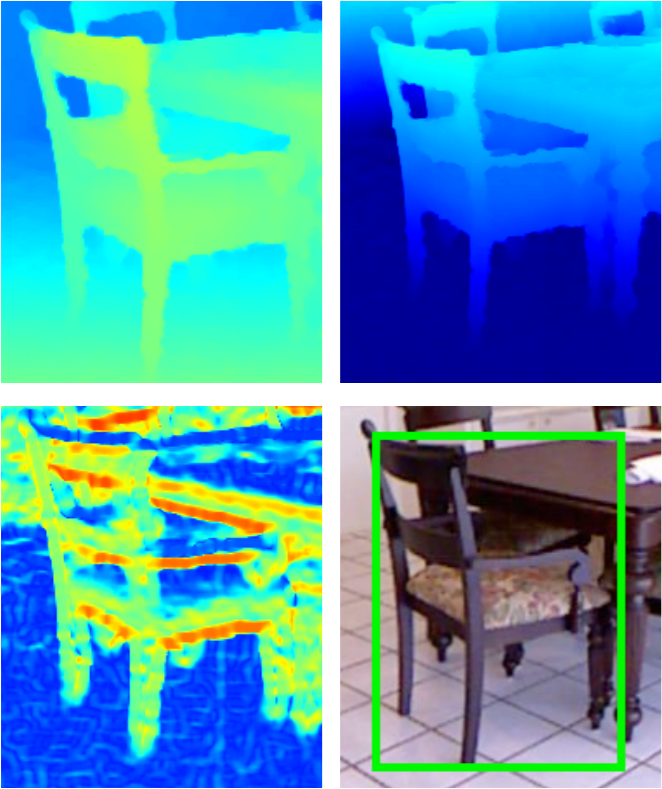


# Instance Segmentation

Predict each pixel in the detection window to be foreground or background

Compute Feature Channels

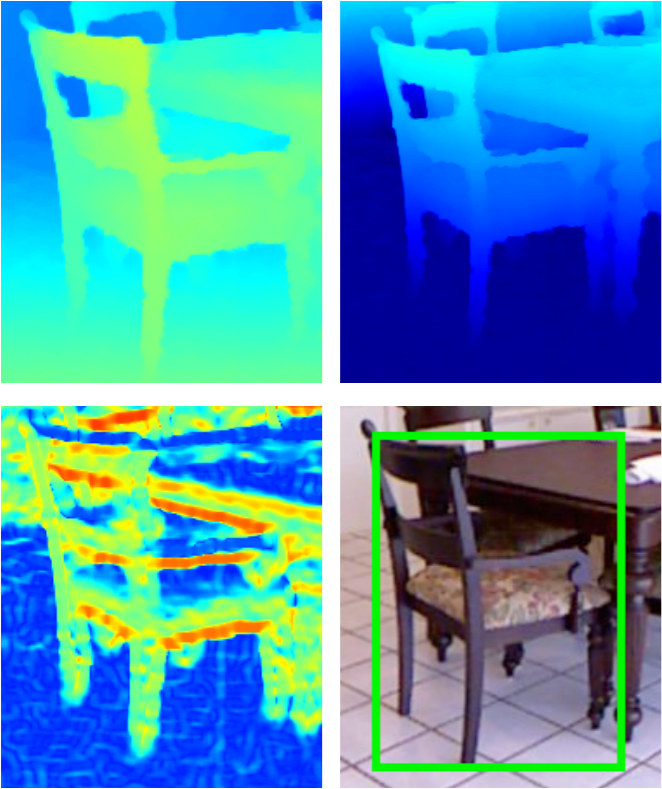
Decision Tree



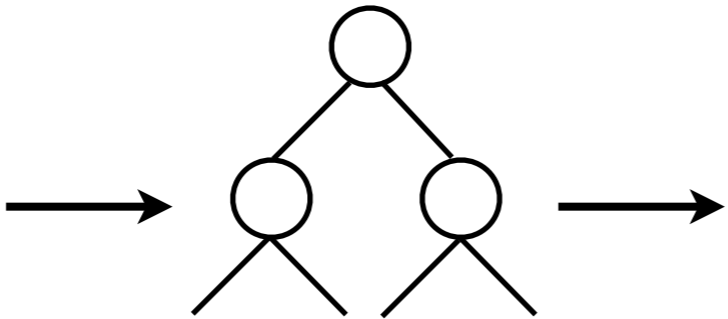
# Instance Segmentation

Predict each pixel in the detection window to be foreground or background

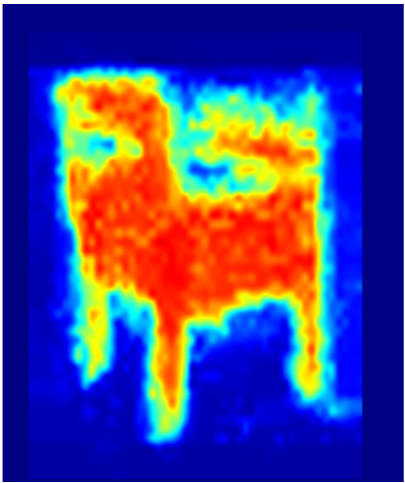
Compute Feature Channels



Decision Tree



Pixel Prediction

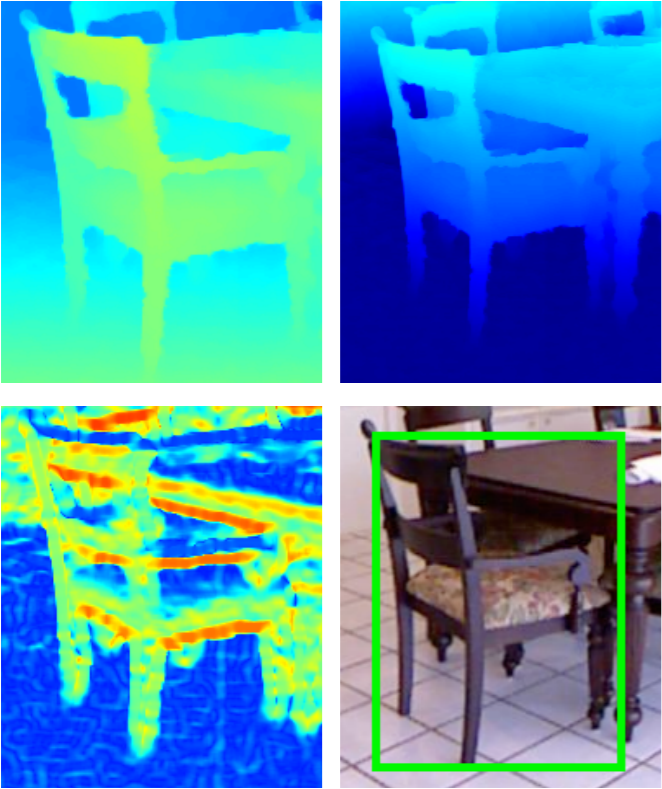




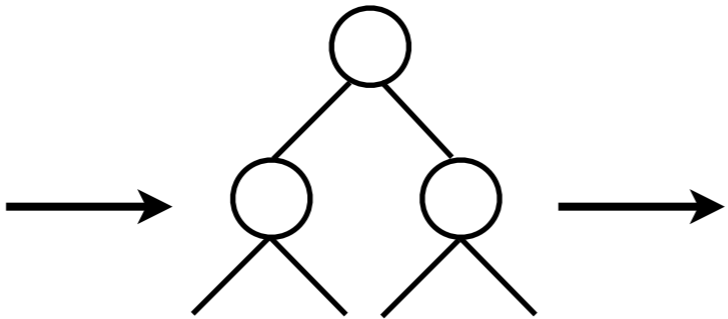
# Instance Segmentation

Predict each pixel in the detection window to be foreground or background

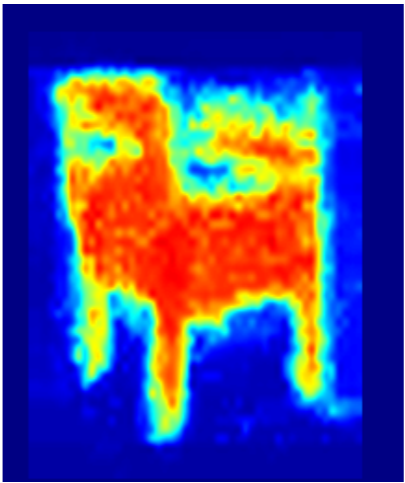
Compute Feature Channels



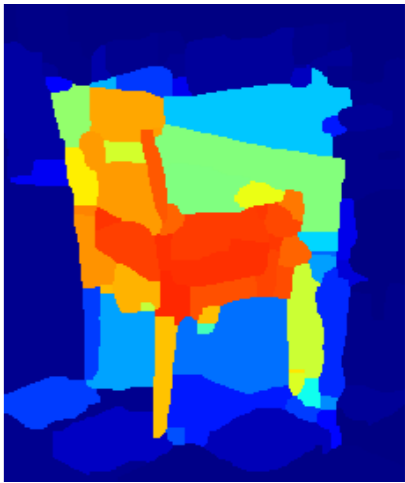
Decision Tree



Pixel Prediction



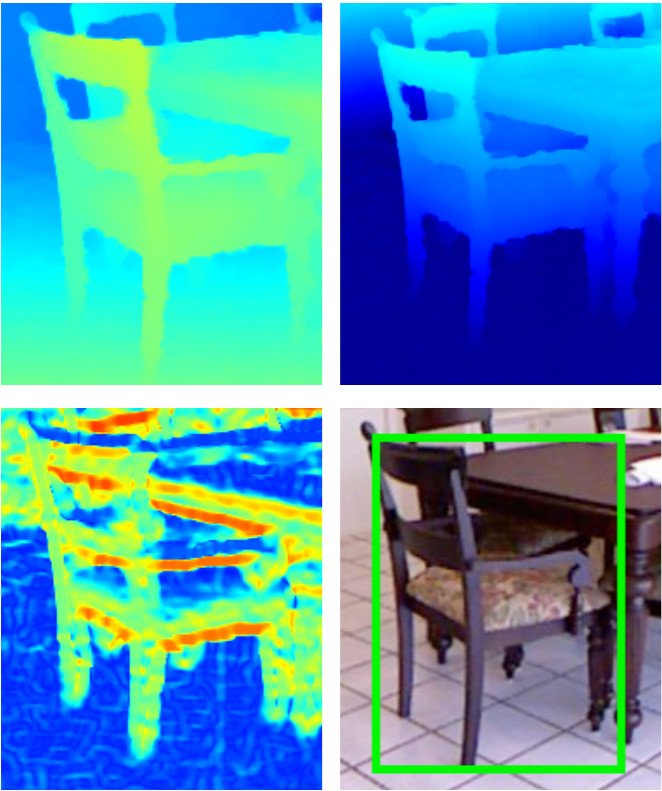
Super pixel Projection



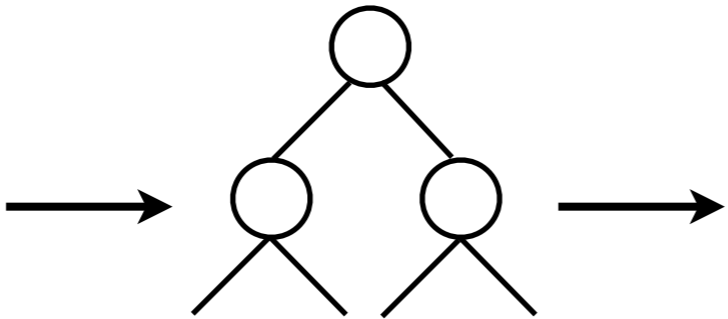
# Instance Segmentation

Predict each pixel in the detection window to be foreground or background

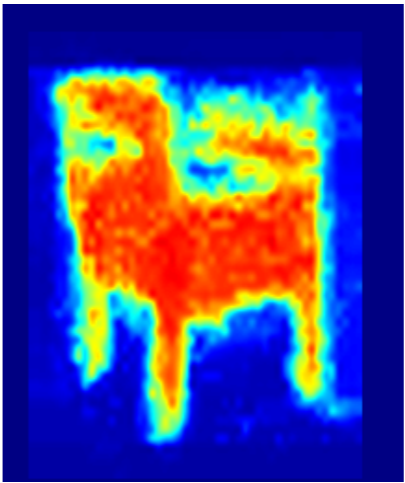
Compute Feature Channels



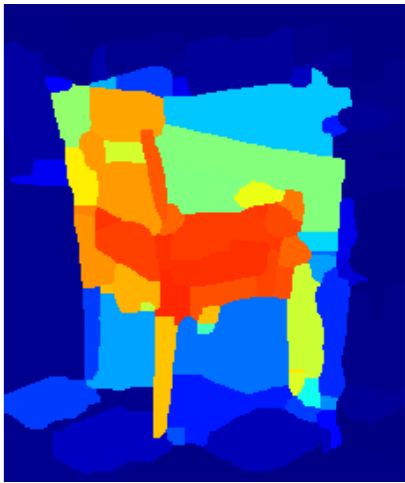
Decision Tree



Pixel Prediction



Super pixel Projection

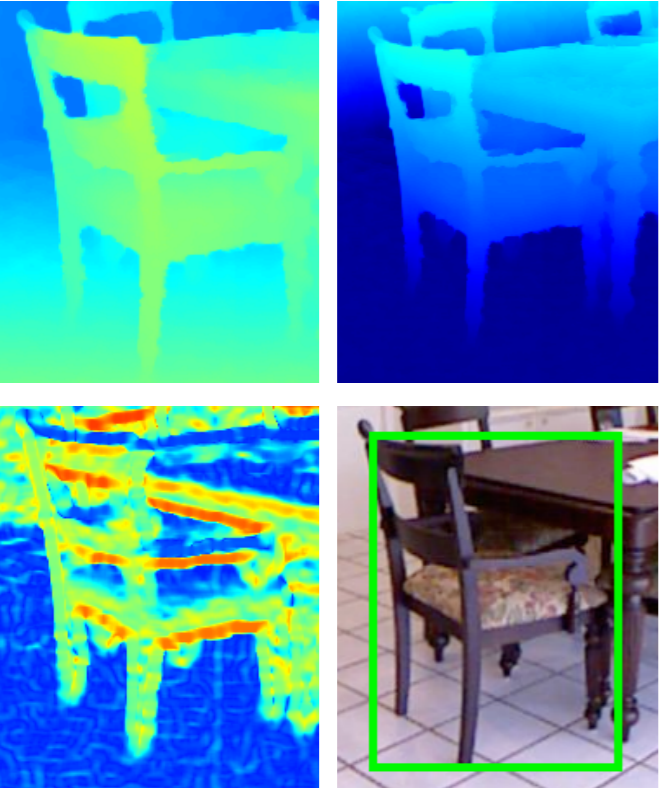


location of pixel in box  
depth, relative depth  
height above ground  
angle with gravity,  
azimuth, normal vector  
Luv color channels,  
is missing data

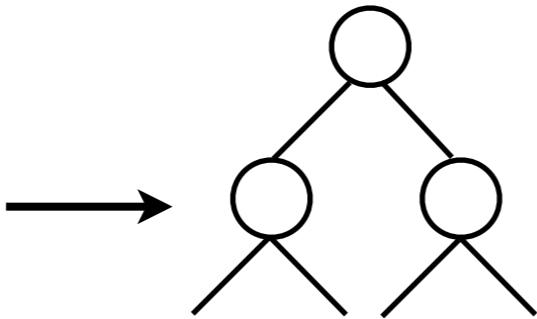
# Instance Segmentation

Predict each pixel in the detection window to be foreground or background

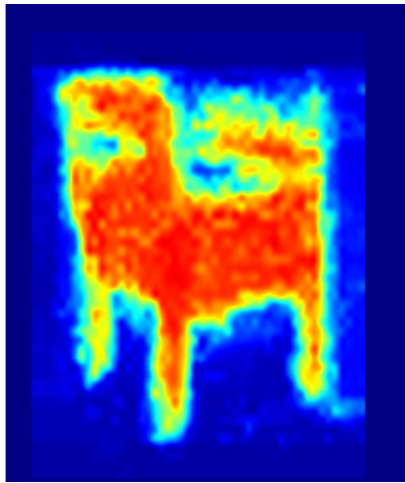
Compute Feature Channels



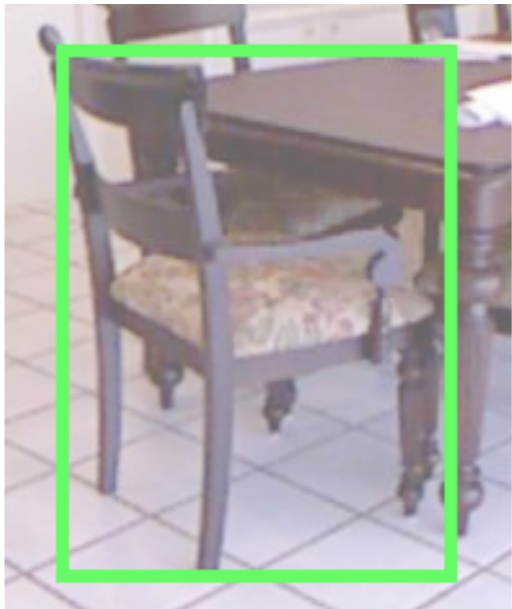
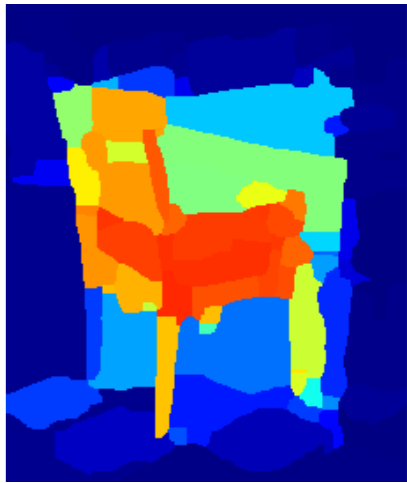
Decision Tree



Pixel Prediction



Super pixel Projection

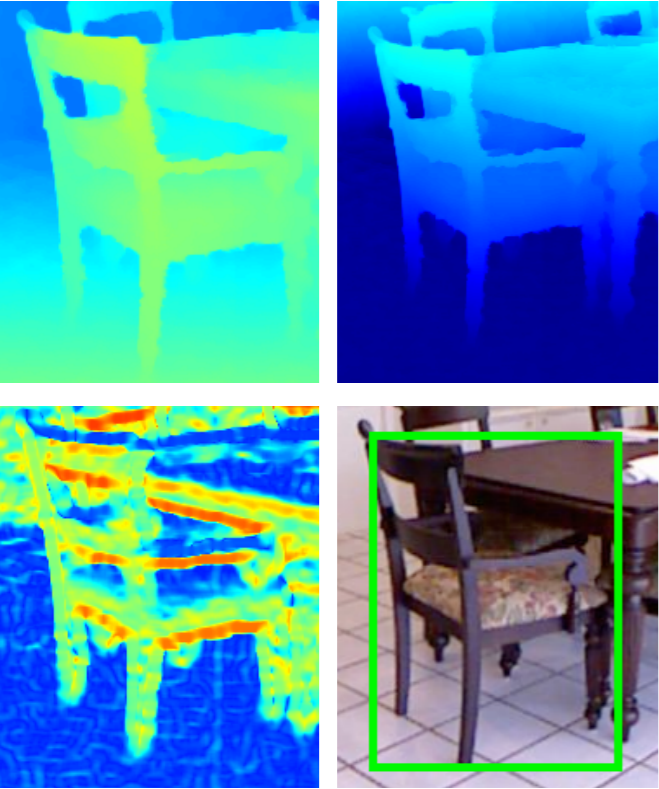


location of pixel in box  
depth, relative depth  
height above ground  
angle with gravity,  
azimuth, normal vector  
Luv color channels,  
is missing data

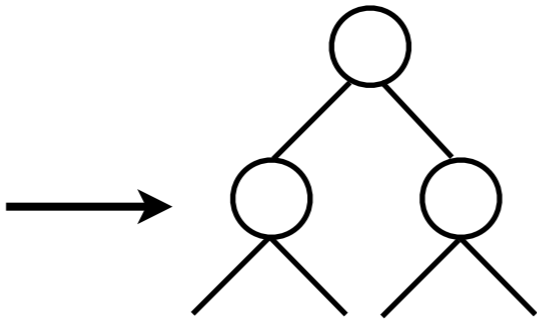
# Instance Segmentation

Predict each pixel in the detection window to be foreground or background

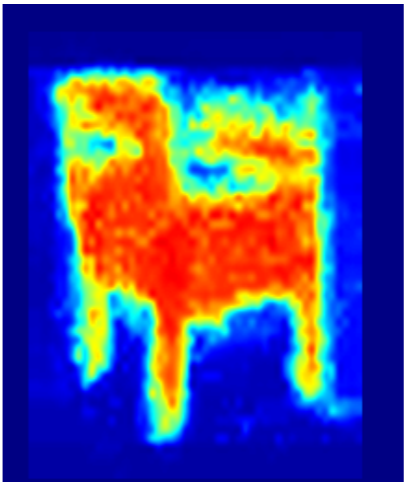
Compute Feature Channels



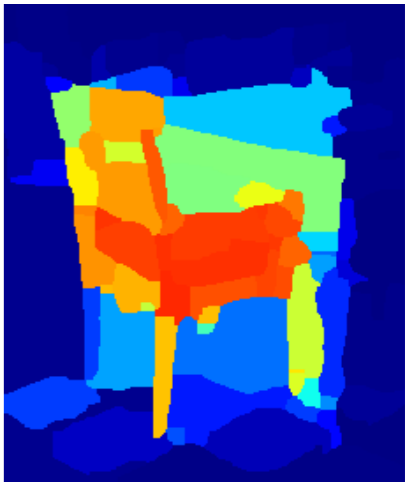
Decision Tree



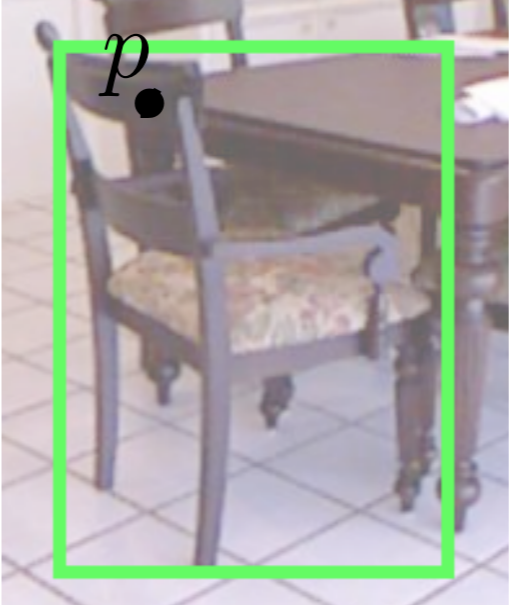
Pixel Prediction



Super pixel Projection



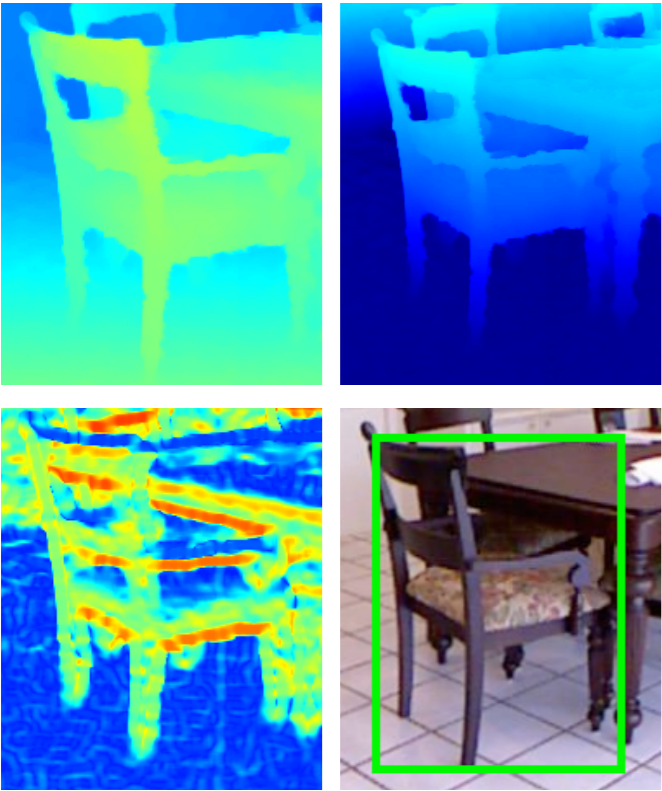
location of pixel in box  
depth, relative depth  
height above ground  
angle with gravity,  
azimuth, normal vector  
Luv color channels,  
is missing data



# Instance Segmentation

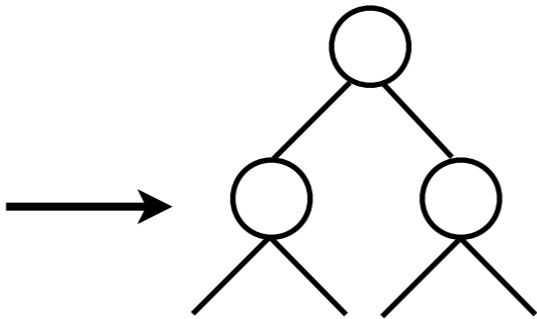
Predict each pixel in the detection window to be foreground or background

Compute Feature Channels

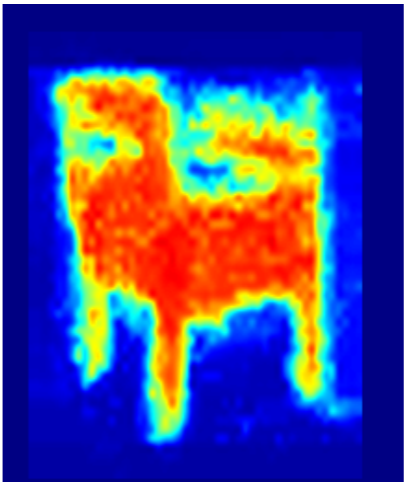


location of pixel in box  
depth, relative depth  
height above ground  
angle with gravity,  
azimuth, normal vector  
Luv color channels,  
is missing data

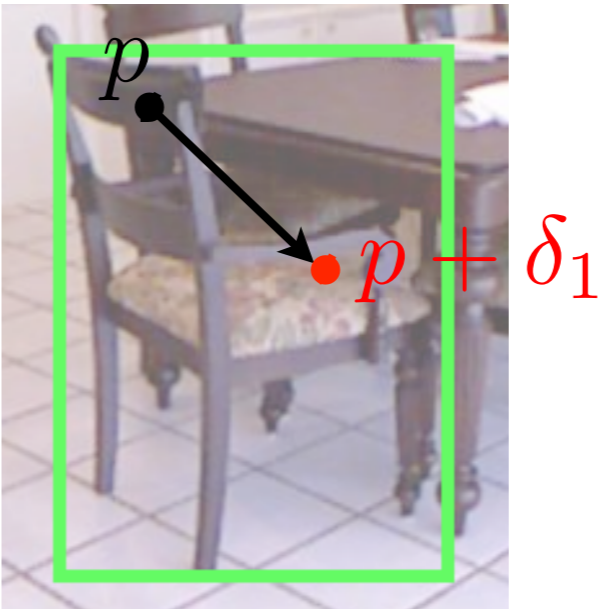
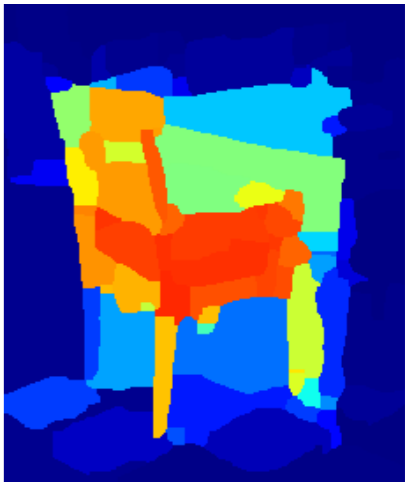
Decision Tree



Pixel Prediction



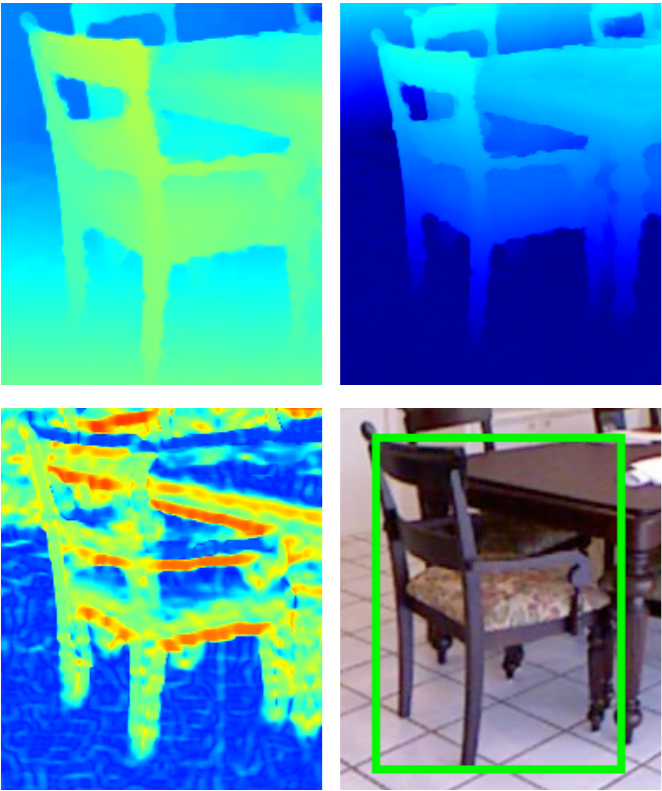
Super pixel Projection



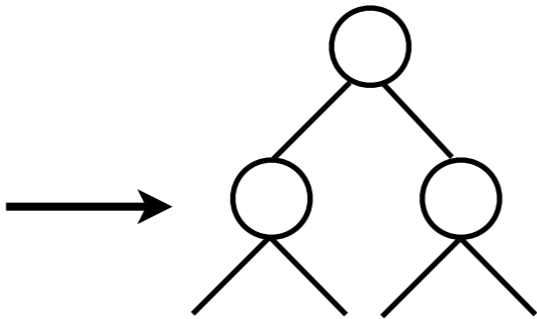
# Instance Segmentation

Predict each pixel in the detection window to be foreground or background

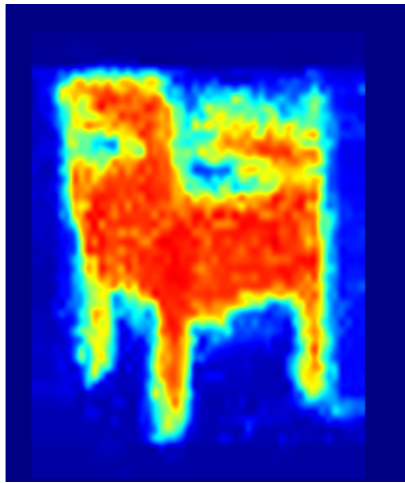
Compute Feature Channels



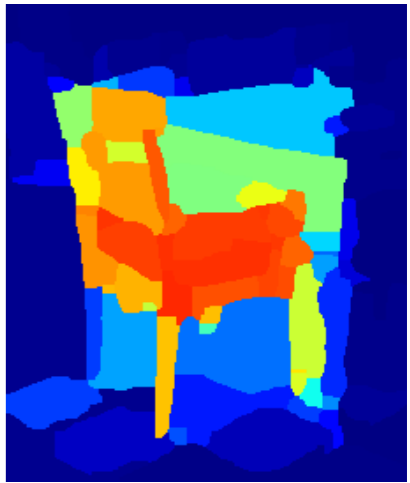
Decision Tree



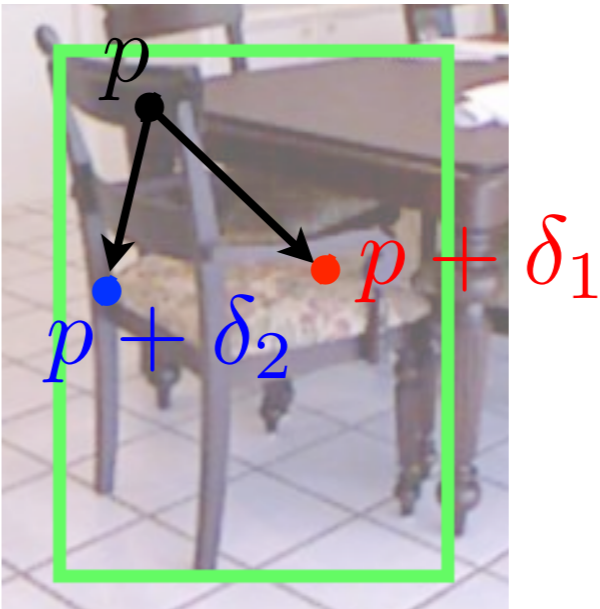
Pixel Prediction



Super pixel Projection



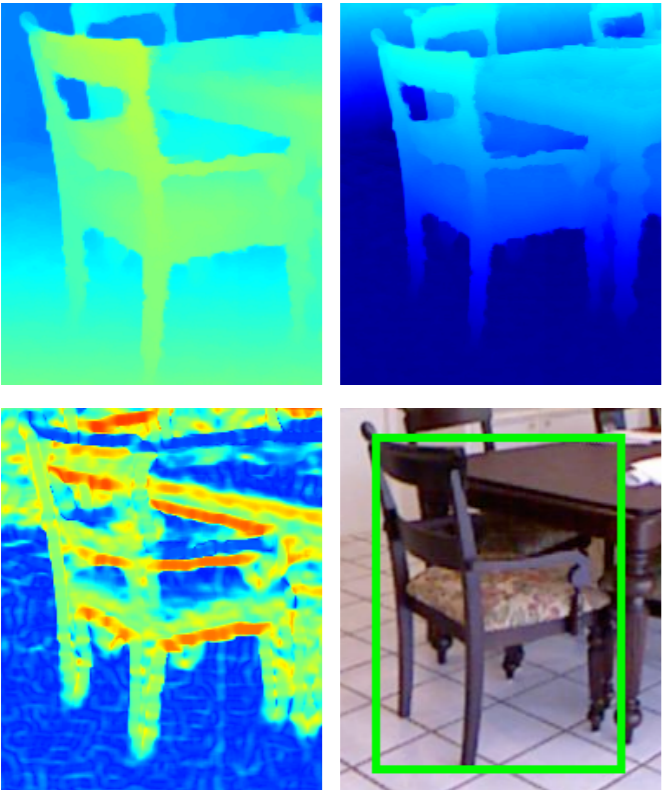
location of pixel in box  
depth, relative depth  
height above ground  
angle with gravity,  
azimuth, normal vector  
Luv color channels,  
is missing data



# Instance Segmentation

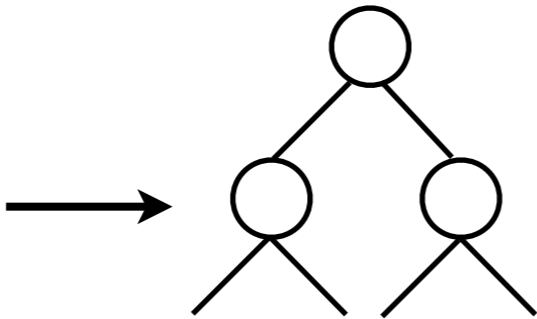
Predict each pixel in the detection window to be foreground or background

Compute Feature Channels

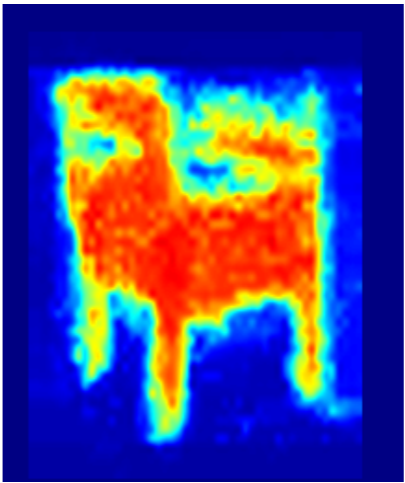


location of pixel in box  
 depth, relative depth  
 height above ground  
 angle with gravity,  
 azimuth, normal vector  
 Luv color channels,  
 is missing data

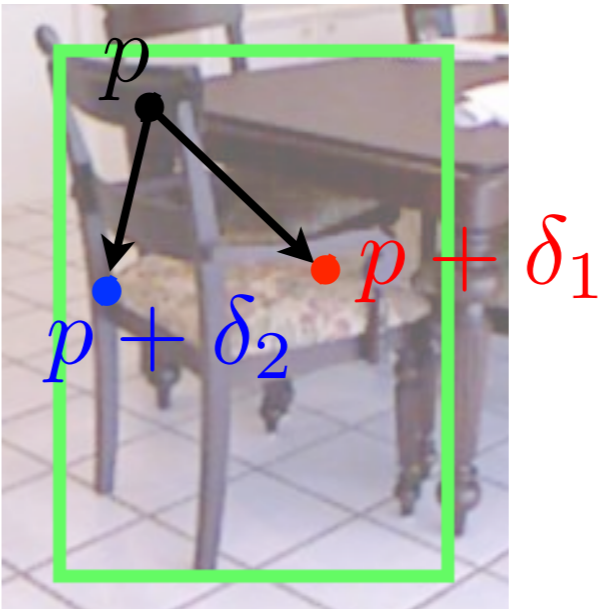
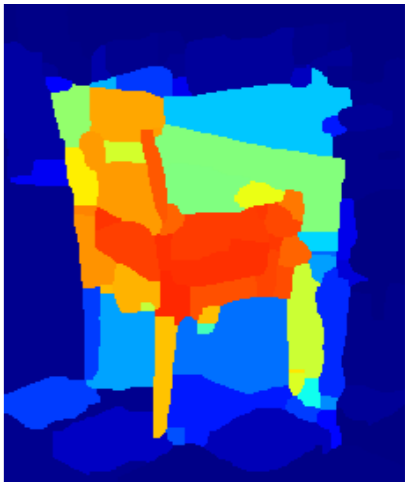
Decision Tree



Pixel Prediction



Super pixel Projection

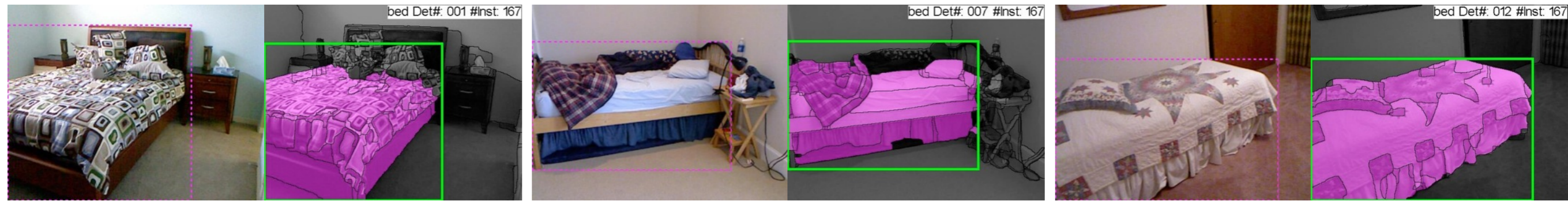


$$d_i (f_i(p + \delta_1), f_i(p + \delta_2)) \geq \tau$$

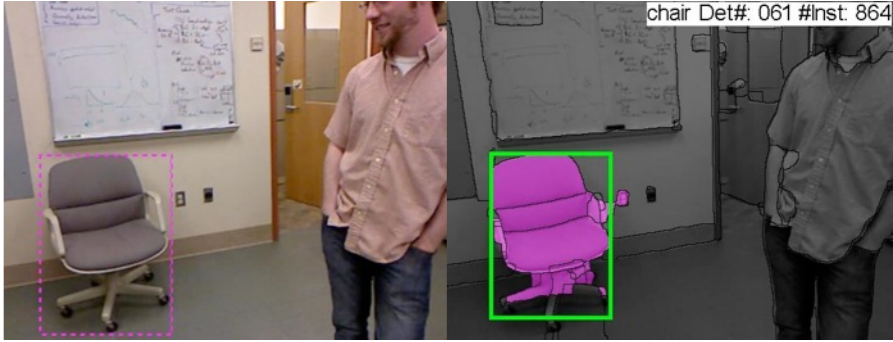
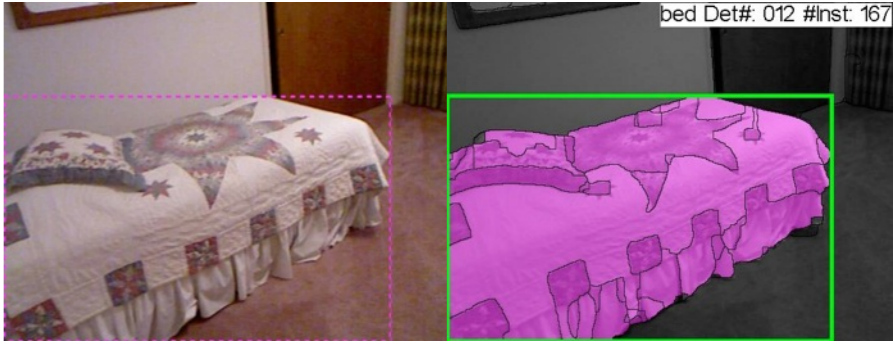
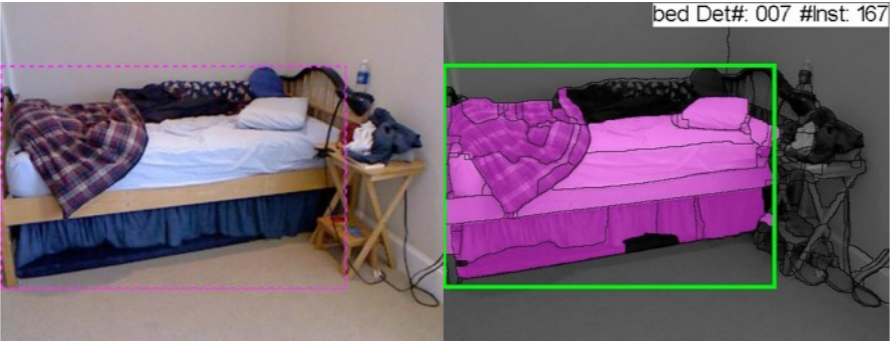
# Instance Segmentation



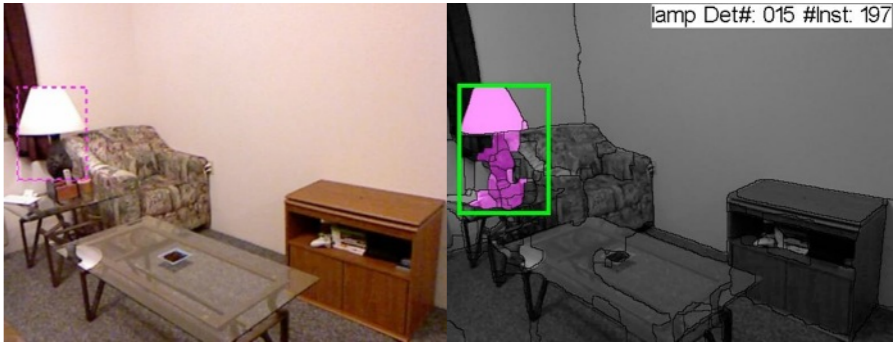
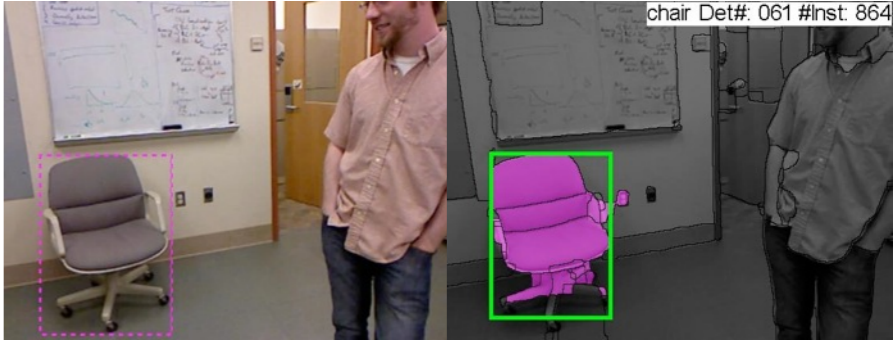
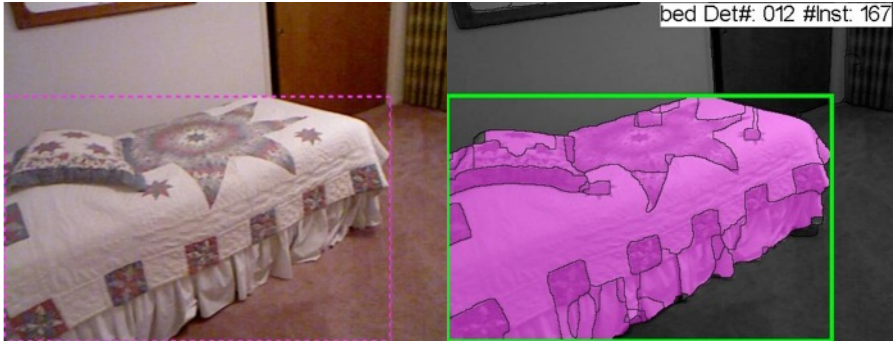
# Instance Segmentation



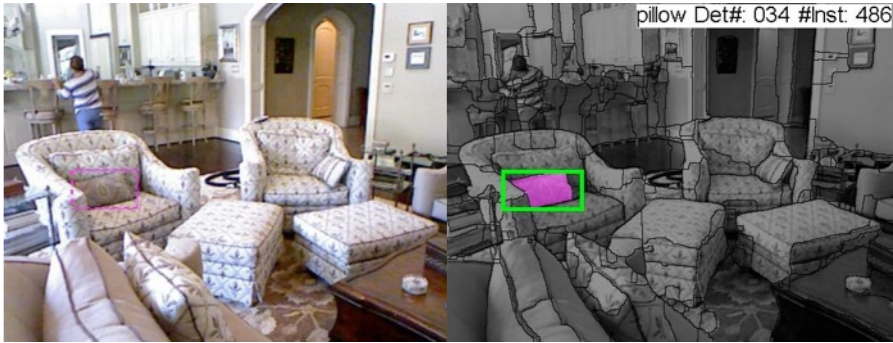
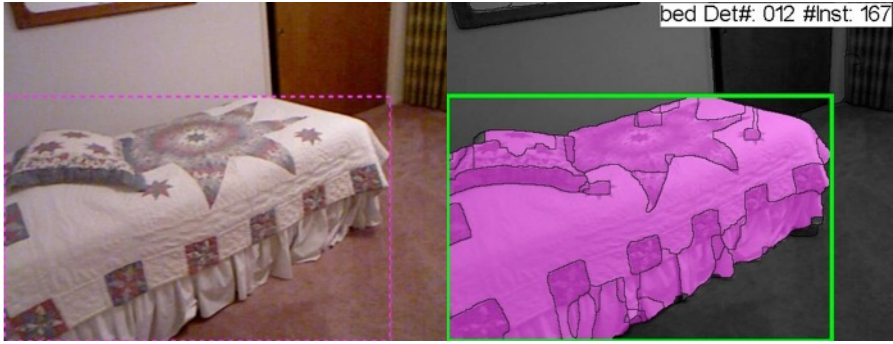
# Instance Segmentation



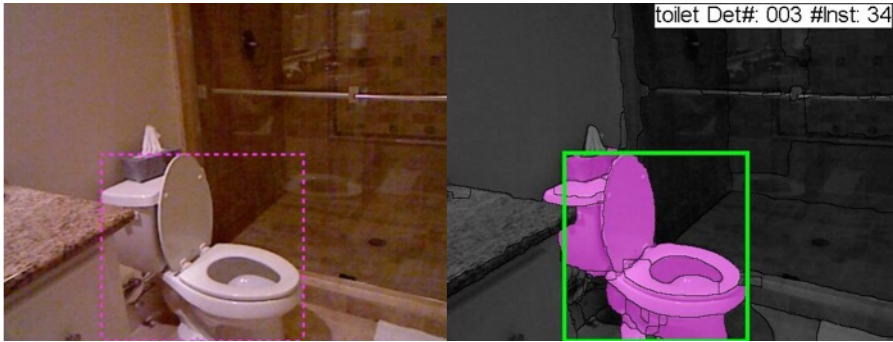
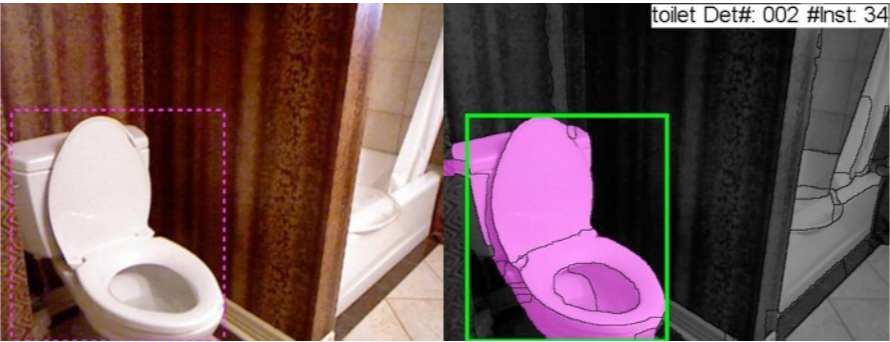
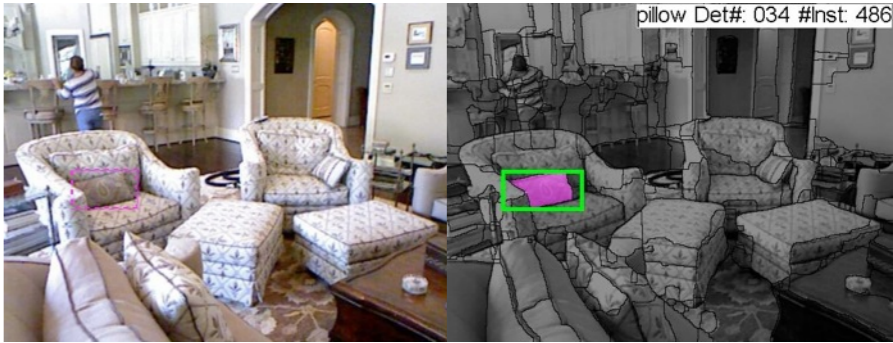
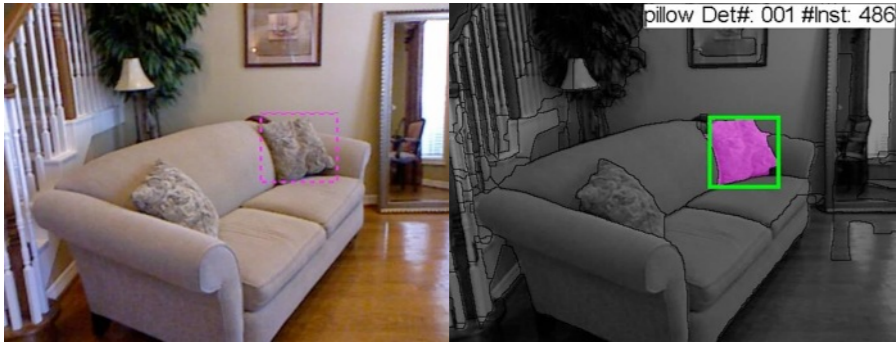
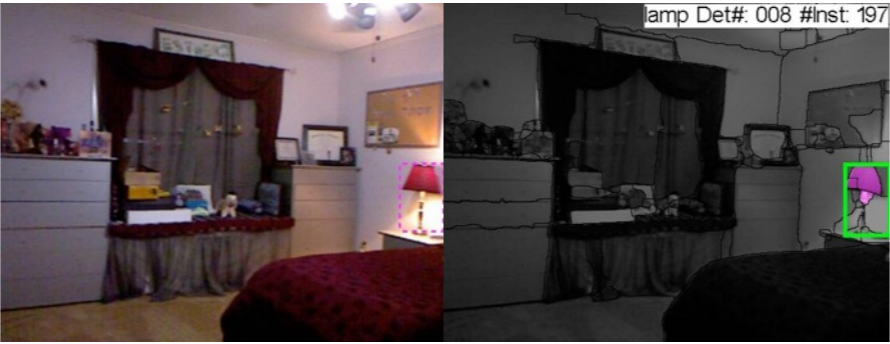
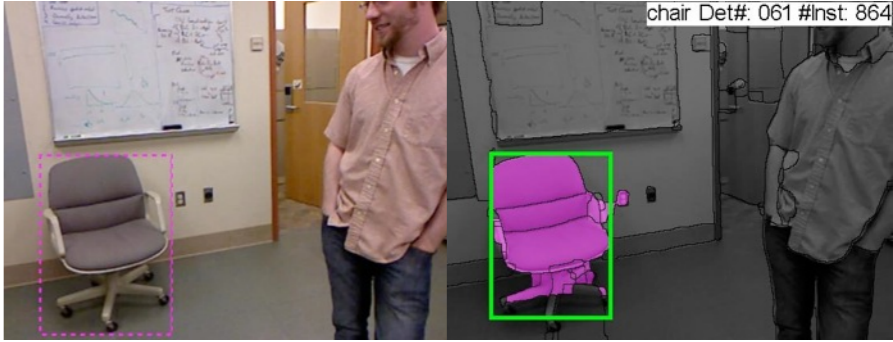
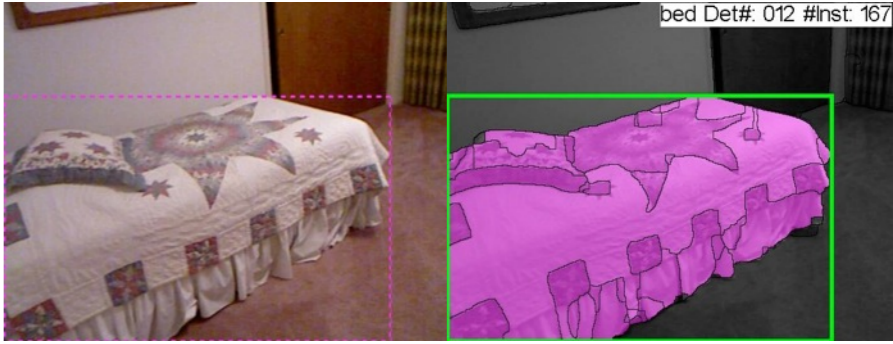
# Instance Segmentation



# Instance Segmentation



# Instance Segmentation



# Instance Segmentation

Performance (region average precision,  $AP^r$ )

# Instance Segmentation

Performance (region average precision, AP<sup>r</sup>)

## Test Set

	mean	bath tub	bed	book shelf	box	chair	counter	desk	door	dress er	garb age bin	lamp	moni tor	night stan d	pillo w	sink	sofa	table	televi sion	toilet
box	<b>14</b>	6	40	4	1	6	1	3	15	27	33	1	40	11	<b>6</b>	9	14	3	35	12
region	<b>28</b>	<b>32</b>	55	9	1	27	21	9	20	29	37	26	48	39	<b>33</b>	31	31	10	34	40
fg mask	<b>28</b>	15	60	9	1	29	5	7	<b>23</b>	33	<b>38</b>	31	<b>55</b>	39	32	32	36	11	37	38
our	<b>32</b>	19	<b>66</b>	<b>10</b>	<b>2</b>	<b>36</b>	<b>33</b>	<b>10</b>	<b>23</b>	<b>34</b>	<b>38</b>	<b>36</b>	53	<b>43</b>	32	<b>34</b>	<b>41</b>	<b>14</b>	<b>37</b>	<b>51</b>

# Instance Segmentation

Performance (region average precision, AP<sup>r</sup>)

## Test Set

	mean	bath tub	bed	book shelf	box	chair	counter	desk	door	dress er	garb age bin	lamp	moni tor	night stan d	pillo w	sink	sofa	table	televi sion	toilet
box	<b>14</b>	6	40	4	1	6	1	3	15	27	33	1	40	11	<b>6</b>	9	14	3	35	12
region	<b>28</b>	<b>32</b>	55	9	1	27	21	9	20	29	37	26	48	39	<b>33</b>	31	31	10	34	40
fg mask	<b>28</b>	15	60	9	1	29	5	7	<b>23</b>	33	<b>38</b>	31	<b>55</b>	39	32	32	36	11	37	38
our	<b>32</b>	19	<b>66</b>	<b>10</b>	<b>2</b>	<b>36</b>	<b>33</b>	<b>10</b>	<b>23</b>	<b>34</b>	<b>38</b>	<b>36</b>	53	<b>43</b>	32	<b>34</b>	<b>41</b>	<b>14</b>	<b>37</b>	<b>51</b>
box AP	<b>37</b>	44	71	33	1	43	44	15	24	30	39	37	53	40	35	36	54	24	38	47



# Instance Segmentation

Performance (region average precision,  $AP^r$ )

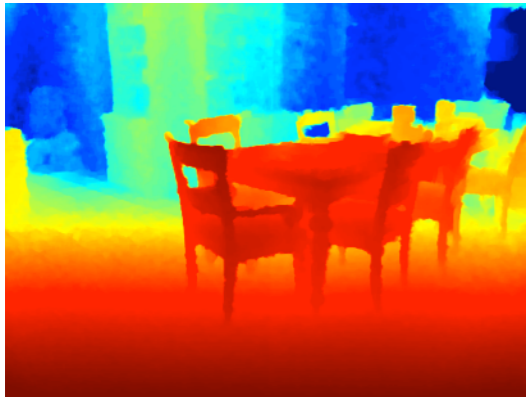
## Test Set

	mean	bath tub	bed	book shelf	box	chair	counter	desk	door	dress er	garb age bin	lamp	moni tor	night stan d	pillo w	sink	sofa	table	televi sion	toilet
box	<b>14</b>	6	40	4	1	6	1	3	15	27	33	1	40	11	<b>6</b>	9	14	3	35	12
region	<b>28</b>	<b>32</b>	55	9	1	27	21	9	20	29	37	26	48	39	<b>33</b>	31	31	10	34	40
fg mask	<b>28</b>	15	60	9	1	29	5	7	<b>23</b>	33	<b>38</b>	31	<b>55</b>	39	32	32	36	11	37	38
our	<b>32</b>	19	<b>66</b>	<b>10</b>	<b>2</b>	<b>36</b>	<b>33</b>	<b>10</b>	<b>23</b>	<b>34</b>	<b>38</b>	<b>36</b>	53	<b>43</b>	32	<b>34</b>	<b>41</b>	<b>14</b>	<b>37</b>	<b>51</b>
box AP	<b>37</b>	44	71	33	1	43	44	15	24	30	39	37	53	40	35	36	54	24	38	47

Better localization makes  $AP^r$  larger than  $AP^b$

# Overview

## Input

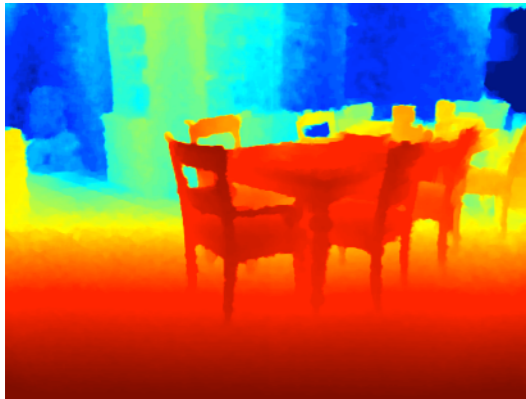


Color and Depth  
Image Pair

# Overview

Input

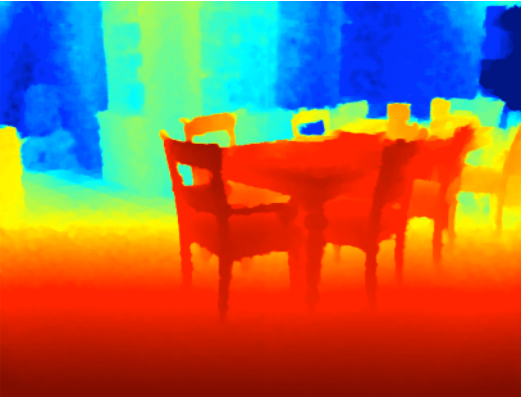
Re-organization



Color and Depth  
Image Pair

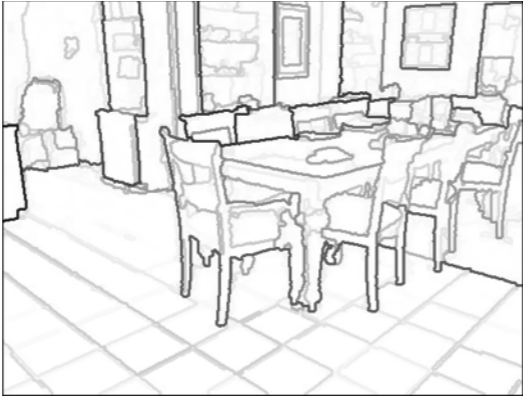
# Overview

## Input



Color and Depth  
Image Pair

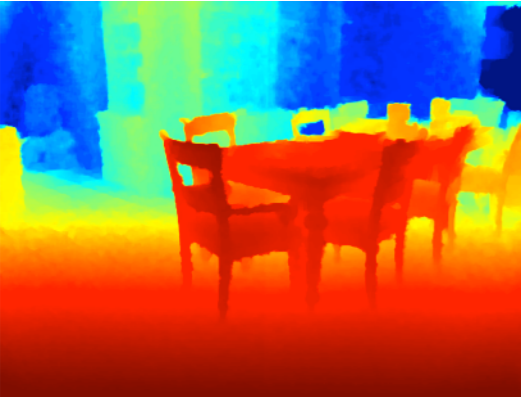
## Re-organization



Contour Detection

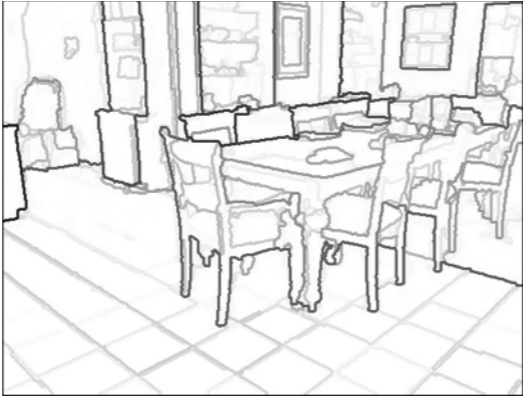
# Overview

## Input

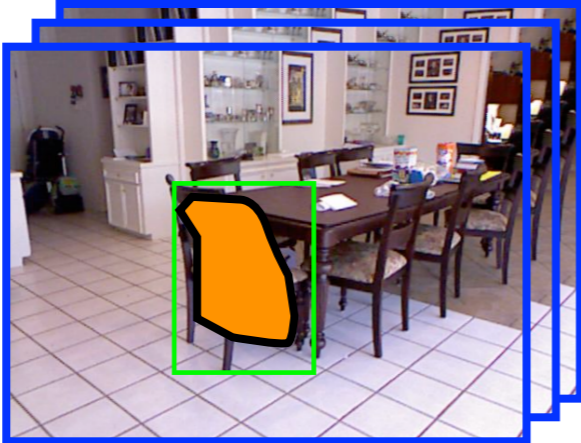


Color and Depth Image Pair

## Re-organization



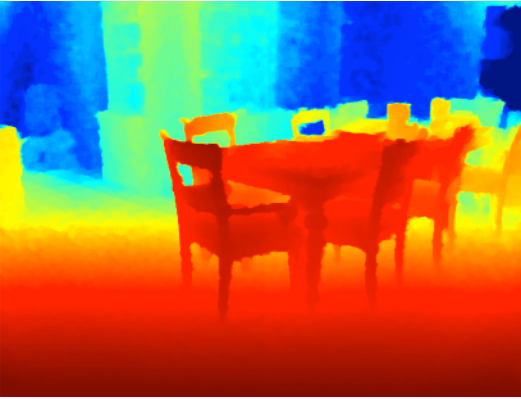
Contour Detection



Region Proposal Generation

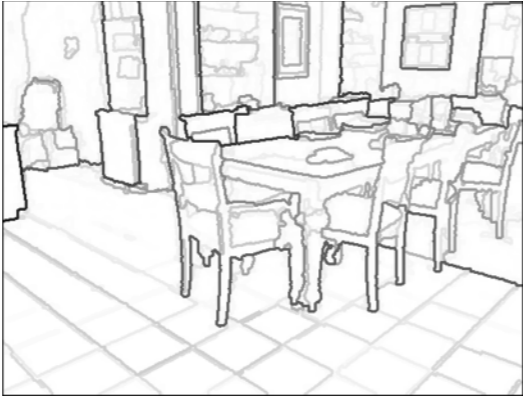
# Overview

## Input



Color and Depth Image Pair

## Re-organization

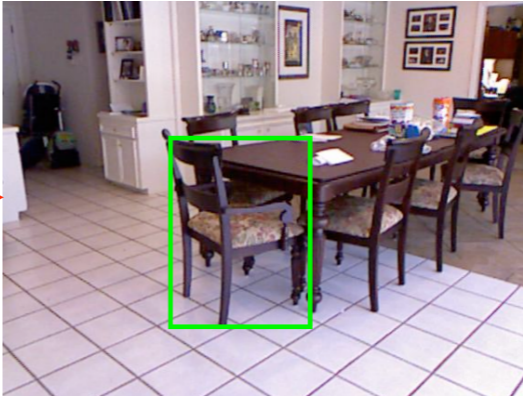


Contour Detection



Region Proposal Generation

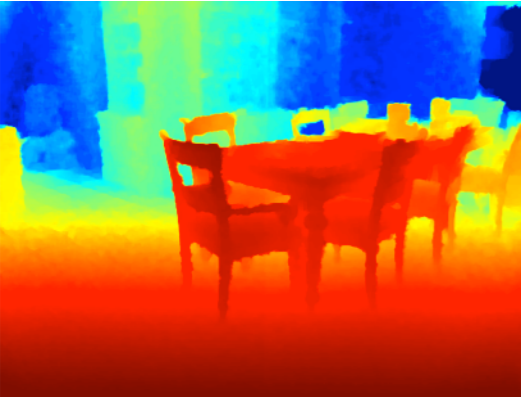
## Recognition



Object Detection

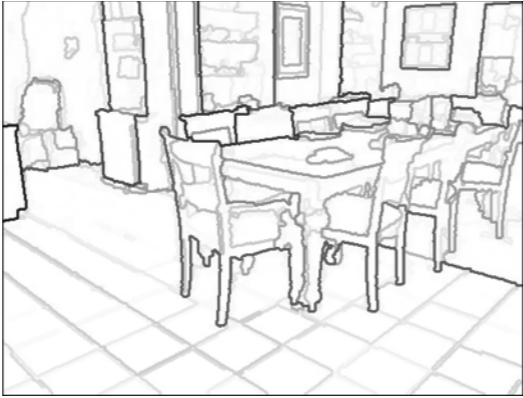
# Overview

## Input



Color and Depth  
Image Pair

## Re-organization

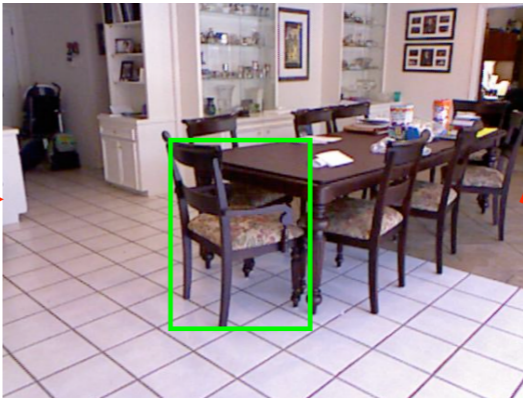


Contour Detection



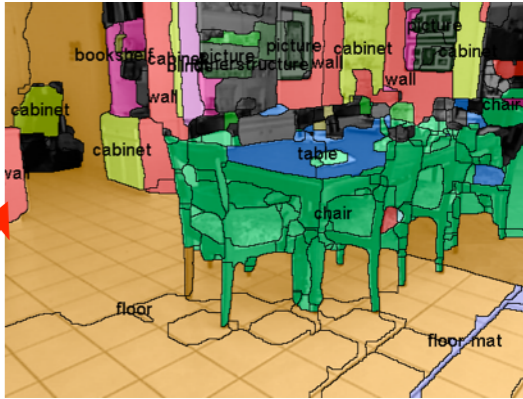
Region Proposal  
Generation

## Recognition



Object Detection

## Extensions

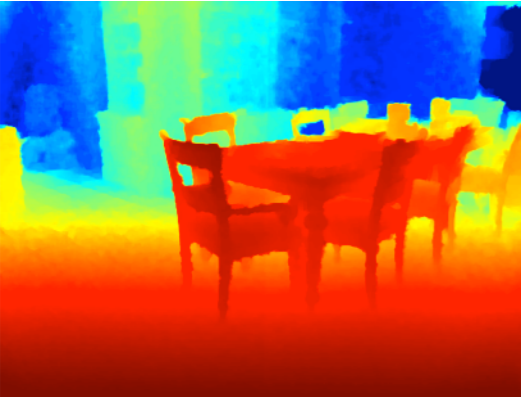


Semantic Segm.



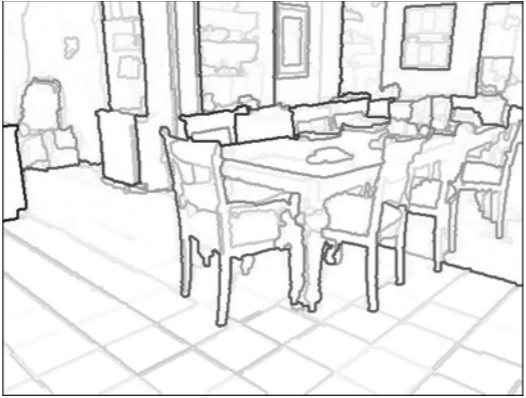
# Overview

## Input

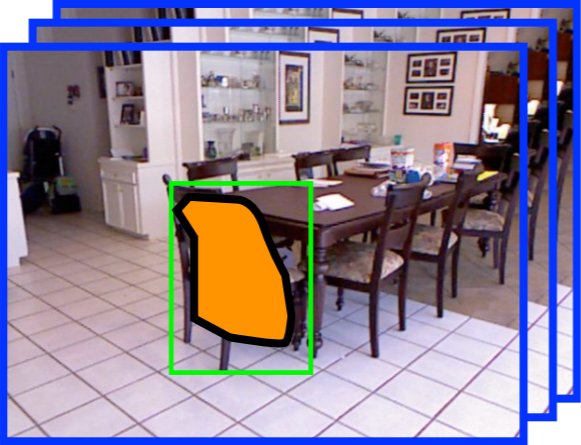


Color and Depth Image Pair

## Re-organization

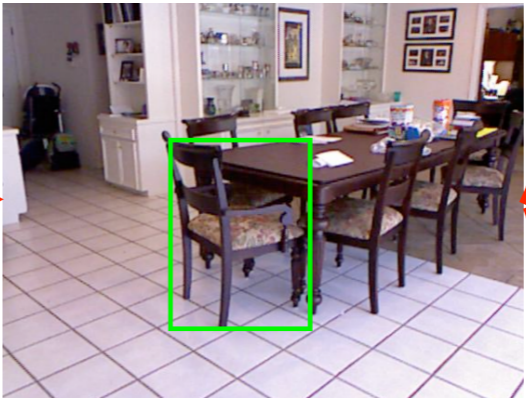


Contour Detection



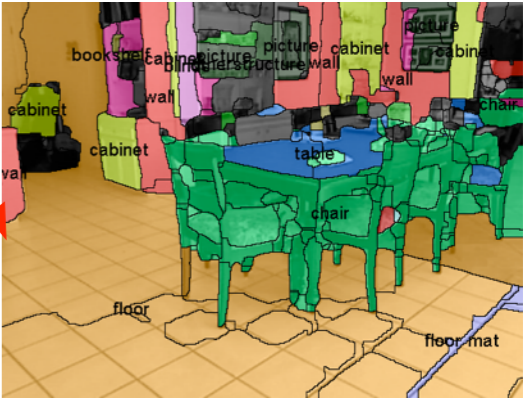
Region Proposal Generation

## Recognition

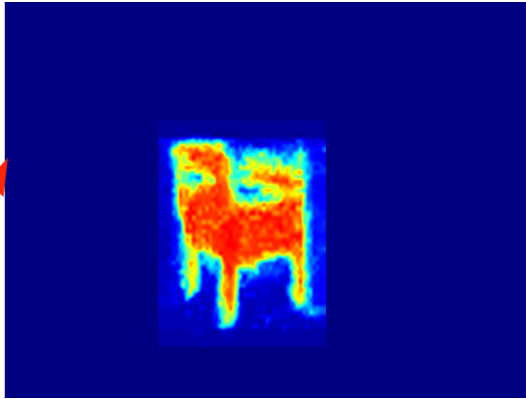


Object Detection

## Extensions



Semantic Segm.

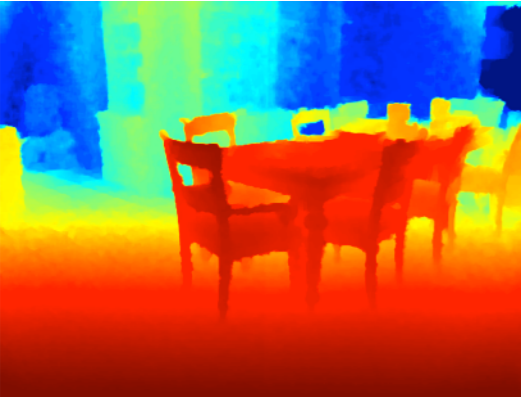


Instance Segm.



# Overview

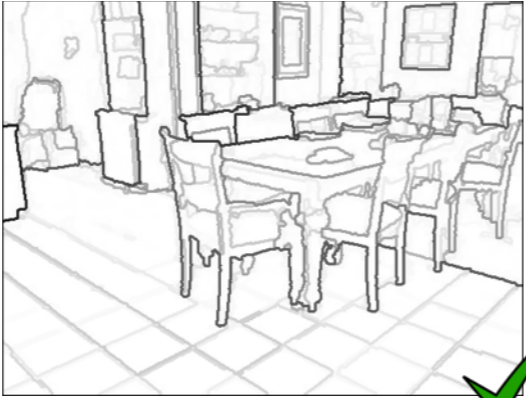
## Input



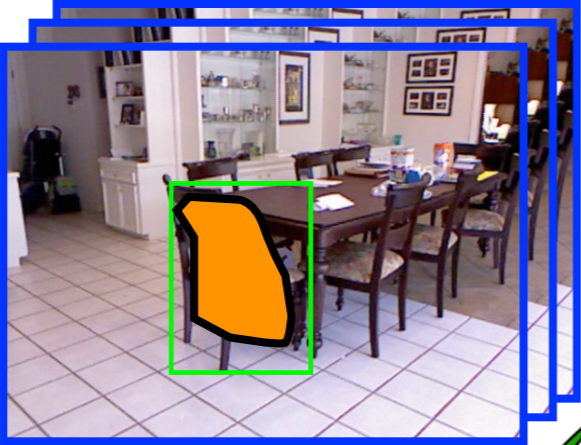
Color and Depth Image Pair

✓ State-of-the-art

## Re-organization

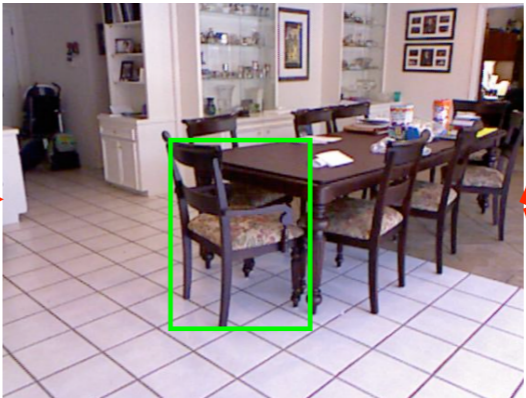


Contour Detection ✓



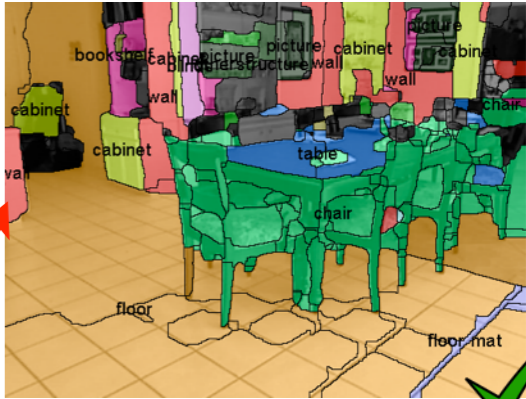
Region Proposal Generation ✓

## Recognition

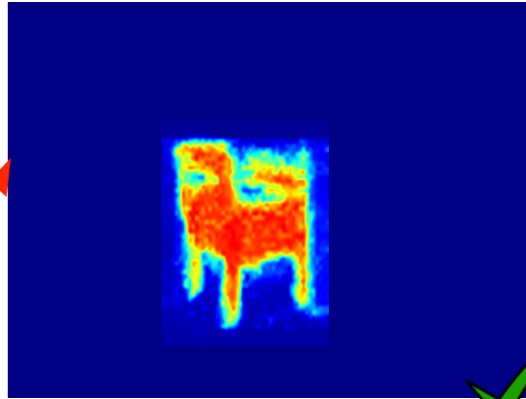


Object Detection ✓

## Extensions



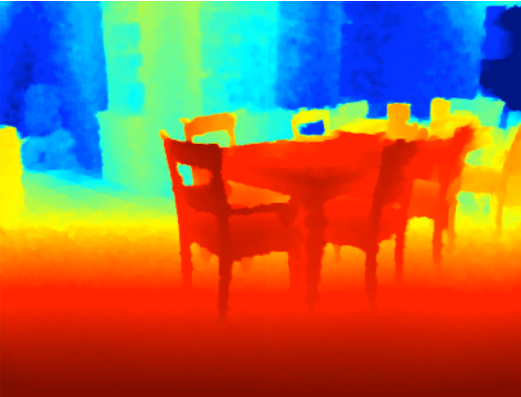
Semantic Segm. ✓



Instance Segm. ✓

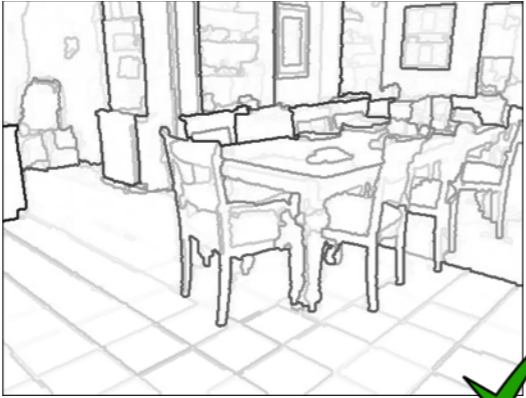
# Overview

## Input



Color and Depth Image Pair

## Re-organization

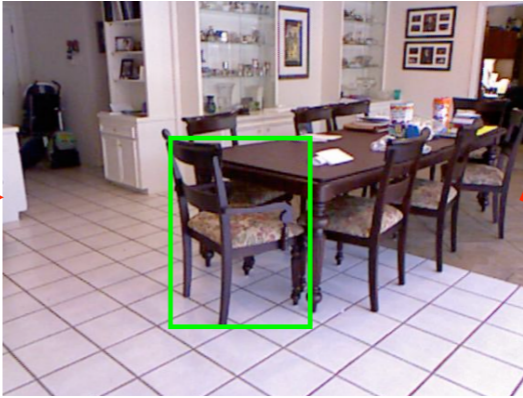


Contour Detection



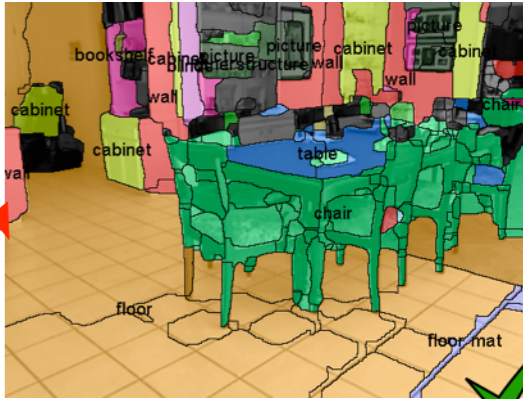
Region Proposal Generation

## Recognition

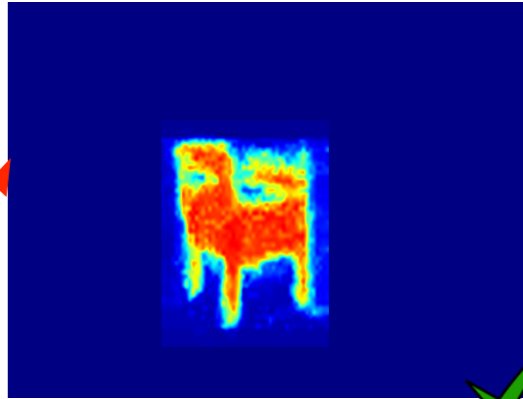


Object Detection

## Extensions



Semantic Segm.



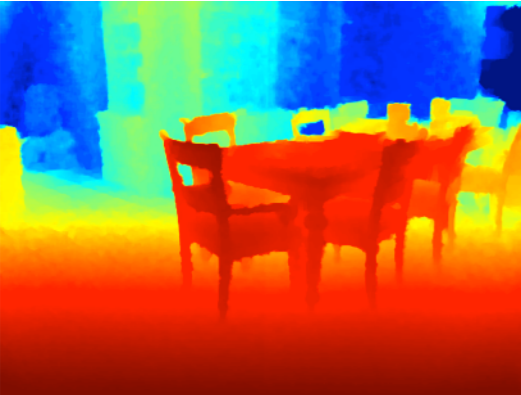
Instance Segm.

✓ State-of-the-art

Code available online!

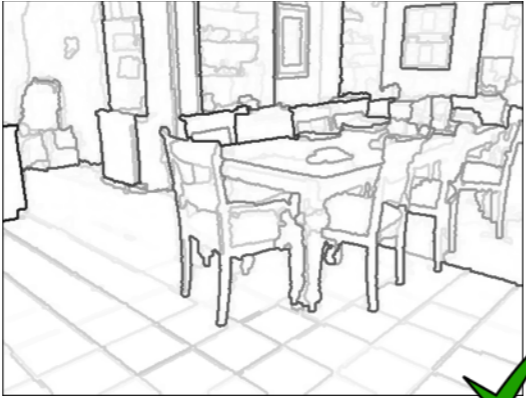
# Overview

## Input

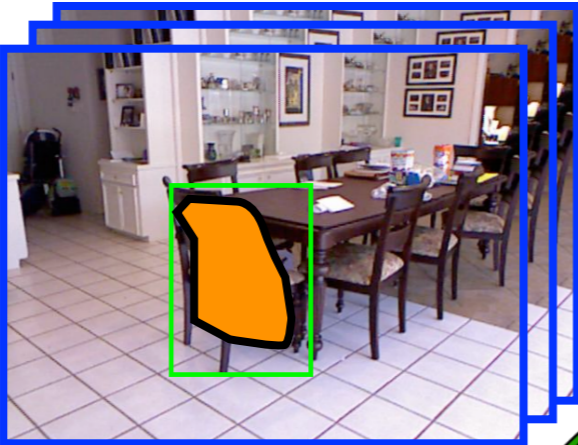


Color and Depth Image Pair

## Re-organization

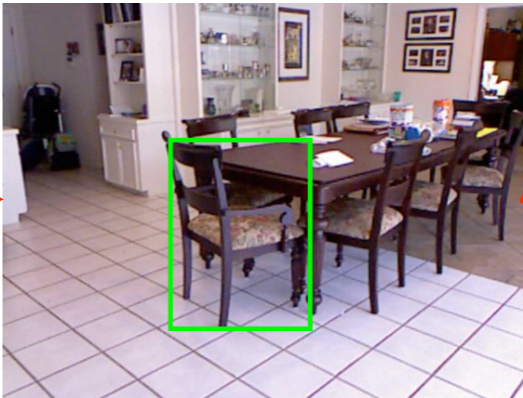


Contour Detection



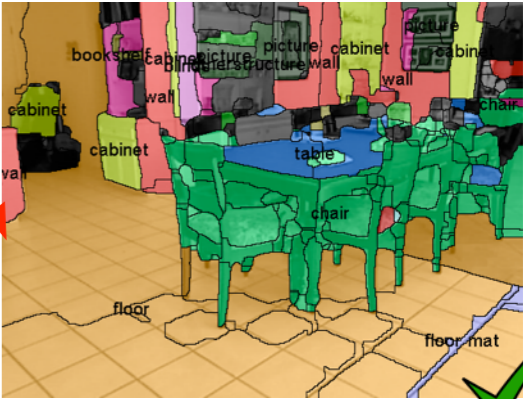
Region Proposal Generation

## Recognition

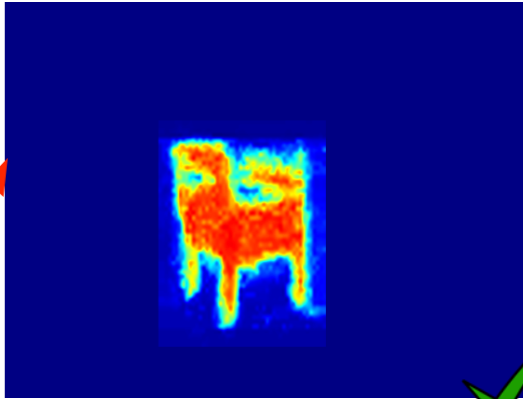


Object Detection

## Extensions



Semantic Segm.



Instance Segm.

✓ State-of-the-art

Code available online!

# Thank You