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

<table>
<thead>
<tr>
<th>Input</th>
<th>Re-organization</th>
</tr>
</thead>
</table>

**Color and Depth Image Pair**
<table>
<thead>
<tr>
<th>Overview</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Input</strong></td>
</tr>
<tr>
<td><img src="image1.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image2.png" alt="Image" /></td>
</tr>
<tr>
<td><strong>Color and Depth Image Pair</strong></td>
</tr>
<tr>
<td><strong>Re-organization</strong></td>
</tr>
<tr>
<td><img src="image3.png" alt="Image" /></td>
</tr>
<tr>
<td><strong>Contour Detection</strong></td>
</tr>
</tbody>
</table>
Overview

Input

Re-organization

Color and Depth Image Pair

Contour Detection

Region Proposal Generation
Overview

Input

Re-organization

Recognition

Color and Depth Image Pair

Contour Detection

Region Proposal Generation

Object Detection

Recognition
Overview

Input

Re-organization

Recognition

Extensions

Color and Depth Image Pair

Contour Detection

Region Proposal Generation

Object Detection

Semantic Segm.
Overview

Input

Re-organization

Recognition

Extensions

Contour Detection

Region Proposal Generation

Object Detection

Semantic Segm.

Instance Segm.

Color and Depth Image Pair
Overview

Input

Re-organization

Recognition

Extensions

State-of-the-art
Recognition, Reconstruction & Reorganization
Recognition, Reconstruction & Reorganization
Overview

Input

Re-organization

Recognition

Extensions

Contour Detection

Region Proposal Generation

Object Detection

Semantic Segm.

Instance Segm.
Reorganization
Reorganization
Reorganization

Contour Detection
Reorganization

Contour Detection

Region Proposal
Reorganization

Related Work

Contour Detection

Region Proposal
Reorganization

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

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

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**
- 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

Re-organization

Recognition

Extensions

Color and Depth Image Pair

Contour Detection

Region Proposal Generation

Object Detection

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

Input Depth Image

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
# Using Local Gradients for Contour Detection

<table>
<thead>
<tr>
<th>Method</th>
<th>max F</th>
</tr>
</thead>
<tbody>
<tr>
<td>gPb-UCM</td>
<td>63.15</td>
</tr>
<tr>
<td>Silberman et al.</td>
<td>65.77</td>
</tr>
<tr>
<td>Dollar et al.</td>
<td>68.96</td>
</tr>
<tr>
<td><strong>Our (gPb-UCM + our cues)</strong></td>
<td>68.66</td>
</tr>
<tr>
<td><strong>Our (Dollar et al. + our cues++)</strong></td>
<td><strong>70.36</strong></td>
</tr>
</tbody>
</table>

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
Using Local Gradients for Contour Detection

Use with gPb-UCM

<table>
<thead>
<tr>
<th>Method</th>
<th>max F</th>
</tr>
</thead>
<tbody>
<tr>
<td>gPb-UCM (RGB)</td>
<td>63.15</td>
</tr>
<tr>
<td>Silberman et al. (RGB-D)</td>
<td>65.77</td>
</tr>
<tr>
<td>Dollar et al. (RGB-D)</td>
<td>68.96</td>
</tr>
<tr>
<td><strong>Our</strong> (gPb-UCM + our cues)</td>
<td>68.66</td>
</tr>
<tr>
<td><strong>Our</strong> (Dollar et al. + our cues++)</td>
<td><strong>70.36</strong></td>
</tr>
</tbody>
</table>

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
Using Local Gradients for Contour Detection

Use with gPb-UCM

Use with Dollar et al.’s structured edges

<table>
<thead>
<tr>
<th>Method</th>
<th>max F</th>
</tr>
</thead>
<tbody>
<tr>
<td>gPb-UCM</td>
<td>63.15</td>
</tr>
<tr>
<td>Silberman et al.</td>
<td>65.77</td>
</tr>
<tr>
<td>Dollar et al.</td>
<td>68.96</td>
</tr>
<tr>
<td><strong>Our (gPb-UCM + our cues)</strong></td>
<td>68.66</td>
</tr>
<tr>
<td><strong>Our (Dollar et al. + our cues++)</strong></td>
<td>70.36</td>
</tr>
</tbody>
</table>

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
Using Local Gradients for Contour Detection

Use with gPb-UCM

Use with Dollar et al.'s structured edges

<table>
<thead>
<tr>
<th>Method</th>
<th>max F</th>
<th>Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>gPb-UCM</td>
<td>63.15</td>
<td>RGB</td>
</tr>
<tr>
<td>Silberman et al.</td>
<td>65.77</td>
<td>RGB-D</td>
</tr>
<tr>
<td>Dollar et al.</td>
<td>68.96</td>
<td>RGB-D</td>
</tr>
<tr>
<td><strong>Our (gPb-UCM + our cues)</strong></td>
<td>68.66</td>
<td>RGB-D</td>
</tr>
<tr>
<td><strong>Our (Dollar et al. + our cues++)</strong></td>
<td><strong>70.36</strong></td>
<td><strong>RGB-D</strong></td>
</tr>
</tbody>
</table>

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
Results

RGB

gPb-UCM(RGB)

D

This Work (RGB-D)
Results

RGB

gPb-UCM(RGB)

This Work (RGB-D)

Less distracted by albedo
Results

RGB

gPb-UCM(RGB)

Less distracted by albedo

Higher Recall

D

This Work (RGB-D)
Results

- RGB
- gPb-UCM (RGB)
- This Work (RGB-D)

Less distracted by albedo
Higher Recall
Higher Precision
Results

RGB

Less distracted by albedo

Higher Recall

Higher Precision

More Complete Objects
Results

Less distracted by albedo

Higher Recall

Higher Precision

More Complete Objects
Results

RGB | Depth | Contours
---|---|---
![RGB Image](image1) | ![Depth Image](image2) | ![Contours Image](image3)
![RGB Image](image4) | ![Depth Image](image5) | ![Contours Image](image6)
![RGB Image](image7) | ![Depth Image](image8) | ![Contours Image](image9)
![RGB Image](image10) | ![Depth Image](image11) | ![Contours Image](image12)
Results

RGB | Depth | Contours | Contour Labels
--- | --- | --- | ---
![RGB Image](image1) | ![Depth Image](image2) | ![Contours Image](image3) | ![Contour Labels](image4)

--- | --- | --- | ---
![RGB Image](image5) | ![Depth Image](image6) | ![Contours Image](image7) | ![Contour Labels](image8)

--- | --- | --- | ---
![RGB Image](image9) | ![Depth Image](image10) | ![Contours Image](image11) | ![Contour Labels](image12)

--- | --- | --- | ---
![RGB Image](image13) | ![Depth Image](image14) | ![Contours Image](image15) | ![Contour Labels](image16)
Results

RGB | Depth | Contours | Contour Labels
--- | --- | --- | ---
![Image](image1) | ![Image](image2) | ![Image](image3) | ![Image](image4)
![Image](image5) | ![Image](image6) | ![Image](image7) | ![Image](image8)
![Image](image9) | ![Image](image10) | ![Image](image11) | ![Image](image12)

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
---|---|---|---
![RGB Image](image1)
![Depth Image](image2)
![Contour Image](image3)
![Label Image](image4)

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

Re-organization

Recognition

Extensions

Color and Depth Image Pair

Contour Detection

Region Proposal Generation

Object Detection

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

Contour Signal → Multi-scale Combinatorial Grouping → Feature Based Pruning / Reranking → k candidates

Region Proposal Generation

Generalize MCG to RGB-D Images

Improved RGB-D Contour → Multi-scale Combinatorial Grouping → Feature Based Pruning / Reranking → k candidates

Region Proposal Generation

Generalize MCG to RGB-D Images

Improved RGB-D Contour → Multi-scale Combinatorial Grouping → Features from RGB-D image → k candidates

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

Improved RGB-D Contour → Multi-scale Combinatorial Grouping → Features from RGB-D image → $k$ candidates

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

- Improved RGB-D Contour
- Multi-scale Combinatorial Grouping
- Features from RGB-D image
- k candidates

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

35 Objects Classes from NYUD2

Average Jaccard Index over Classes

- Lin et al. NMS [RGBD]
- Lin et al. All [RGBD]
- MCG [color contours, color features]
- MCG [depth contours, color features]
- MCG [depth contours, depth features]

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

![Graph showing results](image)

Lin et al. NMS [RGBD]
Lin et al. All [RGBD]
MCG [color contours, color features]
MCG [depth contours, color features]
MCG [depth contours, depth features]

Examples

GT Mask

Best Proposal @500

GT Mask

Best Proposal @500
Overview

Input

Re-organization

Recognition

Extensions

Color and Depth Image Pair

Contour Detection

Region Proposal Generation

Object Detection

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

Region Classification

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

Related Work [RGB]

Sliding Window

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

Region Classification

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

Related Work [RGB]

Sliding Window

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

Region Classification

Uijlings et al., IJCV 13, Selective Search for Object Recognition
Classifying bottom-up bounding box proposals using rich feature descriptors

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

Related Work [RGB]

Sliding Window

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

Region Classification

Uijlings et al., IJCV 13, Selective Search for Object Recognition
Classifying bottom-up bounding box proposals using rich feature descriptors

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

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.
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

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

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

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)
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)
3. Compute CNN features
4. Classify regions

alexnet, finetuned-alexnet

Object Detection

Related Work - RCNN

R-CNN: Regions with CNN features

1. Input image
2. Extract region proposals (~2k)
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

<table>
<thead>
<tr>
<th>finetuned?</th>
<th>layer</th>
<th>synthetic?</th>
<th>mAP</th>
</tr>
</thead>
</table>
## Object Detection

### Experiments

<table>
<thead>
<tr>
<th></th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>finetuned?</td>
<td>DPM</td>
</tr>
<tr>
<td>layer</td>
<td>RGB</td>
</tr>
<tr>
<td>synthetic?</td>
<td></td>
</tr>
<tr>
<td>mAP</td>
<td>8.4</td>
</tr>
</tbody>
</table>
## Object Detection Experiments

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>finetuned?</td>
<td>DPM</td>
<td>DPM</td>
</tr>
<tr>
<td>layer</td>
<td>RGB</td>
<td>RGBD</td>
</tr>
<tr>
<td>synthetic?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mAP</td>
<td>8.4</td>
<td>21.7</td>
</tr>
</tbody>
</table>
## Object Detection

### Experiments

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>finetuned?</td>
<td>DPM</td>
<td>DPM</td>
<td>CNN</td>
</tr>
<tr>
<td>layer</td>
<td>RGB</td>
<td>RGBD</td>
<td>fc6</td>
</tr>
<tr>
<td>synthetic?</td>
<td>RGB</td>
<td>RGBD</td>
<td>RGB</td>
</tr>
<tr>
<td>mAP</td>
<td>8.4</td>
<td>21.7</td>
<td>16.4</td>
</tr>
</tbody>
</table>
## Object Detection Experiments

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>finetuned?</td>
<td>yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>layer</td>
<td>DPM</td>
<td>DPM</td>
<td>CNN</td>
<td>CNN</td>
</tr>
<tr>
<td></td>
<td>fc6</td>
<td>fc6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>synthetic?</td>
<td>RGB</td>
<td>RGBD</td>
<td>RGB</td>
<td>RGB</td>
</tr>
<tr>
<td>mAP</td>
<td>8.4</td>
<td>21.7</td>
<td>16.4</td>
<td>19.7</td>
</tr>
</tbody>
</table>
## Object Detection Experiments

<table>
<thead>
<tr>
<th>finetuned?</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DPM</td>
<td>DPM</td>
<td>CNN</td>
<td>CNN</td>
<td>CNN</td>
</tr>
<tr>
<td>layer</td>
<td>fc6</td>
<td>fc6</td>
<td>fc6</td>
<td>fc6</td>
<td>fc6</td>
</tr>
<tr>
<td>synthetic?</td>
<td>RGB</td>
<td>RGBD</td>
<td>RGB</td>
<td>RGB</td>
<td>disparity</td>
</tr>
<tr>
<td>mAP</td>
<td>8.4</td>
<td>21.7</td>
<td>16.4</td>
<td>19.7</td>
<td>11.3</td>
</tr>
</tbody>
</table>
## Object Detection

### Experiments

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>finetuned?</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>layer</td>
<td>DPM</td>
<td>DPM</td>
<td>CNN</td>
<td>CNN</td>
<td>CNN</td>
<td>CNN</td>
</tr>
<tr>
<td></td>
<td>fc6</td>
<td>fc6</td>
<td>fc6</td>
<td>fc6</td>
<td>fc6</td>
<td>fc6</td>
</tr>
<tr>
<td>synthetic?</td>
<td>RGB</td>
<td>RGBD</td>
<td>RGB</td>
<td>RGB</td>
<td>disparity</td>
<td>disparity</td>
</tr>
<tr>
<td>mAP</td>
<td>8.4</td>
<td>21.7</td>
<td>16.4</td>
<td>19.7</td>
<td>11.3</td>
<td>20.1</td>
</tr>
</tbody>
</table>
## Object Detection
### Experiments

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>finetuned?</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
<td>E</td>
<td>F</td>
</tr>
<tr>
<td></td>
<td>DPM</td>
<td>DPM</td>
<td>CNN</td>
<td>CNN</td>
<td>CNN</td>
<td>CNN</td>
</tr>
<tr>
<td>layer</td>
<td>fc6</td>
<td>fc6</td>
<td>fc6</td>
<td>fc6</td>
<td>fc6</td>
<td>fc6</td>
</tr>
<tr>
<td>synthetic?</td>
<td>RGB</td>
<td>RGBD</td>
<td>RGB</td>
<td>RGB</td>
<td>disparity</td>
<td>disparity</td>
</tr>
<tr>
<td>mAP</td>
<td>8.4</td>
<td>21.7</td>
<td>16.4</td>
<td>19.7</td>
<td>11.3</td>
<td>20.1</td>
</tr>
</tbody>
</table>
## Object Detection

### Experiments

<table>
<thead>
<tr>
<th>finetuned?</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DPM</td>
<td>DPM</td>
<td>CNN</td>
<td>CNN</td>
<td>CNN</td>
<td>CNN</td>
<td>CNN</td>
<td>CNN</td>
</tr>
<tr>
<td>layer</td>
<td>fc6</td>
<td>fc6</td>
<td>fc6</td>
<td>fc6</td>
<td>fc6</td>
<td>fc6</td>
<td>fc6</td>
<td>fc6</td>
</tr>
<tr>
<td>synthetic?</td>
<td></td>
<td></td>
<td>RGB</td>
<td>RGBD</td>
<td>RGB</td>
<td>disparity</td>
<td>disparity</td>
<td>HHA</td>
</tr>
<tr>
<td>mAP</td>
<td>8.4</td>
<td>21.7</td>
<td>16.4</td>
<td>19.7</td>
<td>11.3</td>
<td>20.1</td>
<td>25.2</td>
<td>26.1</td>
</tr>
</tbody>
</table>
## Object Detection

### Experiments

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>finetuned?</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>layer</td>
<td>DPM</td>
<td>DPM</td>
<td>CNN</td>
<td>CNN</td>
<td>CNN</td>
<td>CNN</td>
<td>CNN</td>
<td>CNN</td>
<td>CNN</td>
</tr>
<tr>
<td></td>
<td>fc6</td>
<td>fc6</td>
<td>fc6</td>
<td>fc6</td>
<td>fc6</td>
<td>fc6</td>
<td>fc6</td>
<td>fc6</td>
<td>pool5</td>
</tr>
<tr>
<td>synthetic?</td>
<td>RGB</td>
<td>RGBD</td>
<td>RGB</td>
<td>RGB</td>
<td>disparity</td>
<td>disparity</td>
<td>HHA</td>
<td>HHA</td>
<td>HHA</td>
</tr>
<tr>
<td>mAP</td>
<td>8.4</td>
<td>21.7</td>
<td>16.4</td>
<td>19.7</td>
<td>11.3</td>
<td>20.1</td>
<td>25.2</td>
<td>26.1</td>
<td>21.9</td>
</tr>
</tbody>
</table>
# Object Detection Experiments

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
</tr>
</thead>
<tbody>
<tr>
<td>finetuned?</td>
<td></td>
<td></td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>layer</td>
<td>DPM</td>
<td>DPM</td>
<td>CNN</td>
<td>CNN</td>
<td>CNN</td>
<td>CNN</td>
<td>CNN</td>
<td>CNN</td>
<td>CNN</td>
<td>CNN</td>
</tr>
<tr>
<td></td>
<td>fc6</td>
<td>fc6</td>
<td>fc6</td>
<td>fc6</td>
<td>fc6</td>
<td>fc6</td>
<td>fc6</td>
<td>fc6</td>
<td>pool5</td>
<td>fc7</td>
</tr>
<tr>
<td>synthetic?</td>
<td>RGB</td>
<td>RGBD</td>
<td>RGB</td>
<td>RGB</td>
<td>disparity</td>
<td>disparity</td>
<td>HHA</td>
<td>HHA</td>
<td>HHA</td>
<td>HHA</td>
</tr>
<tr>
<td>mAP</td>
<td>8.4</td>
<td>21.7</td>
<td>16.4</td>
<td>19.7</td>
<td>11.3</td>
<td>20.1</td>
<td>25.2</td>
<td>26.1</td>
<td>21.9</td>
<td>25.3</td>
</tr>
</tbody>
</table>
## Object Detection Experiments

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
</tr>
</thead>
<tbody>
<tr>
<td>finetuned?</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>layer</td>
<td>DPM</td>
<td>DPM</td>
<td>CNN</td>
<td>CNN</td>
<td>CNN</td>
<td>CNN</td>
<td>CNN</td>
<td>CNN</td>
<td>CNN</td>
<td>CNN</td>
<td>CNN</td>
</tr>
<tr>
<td></td>
<td>fc6</td>
<td>fc6</td>
<td>fc6</td>
<td>fc6</td>
<td>fc6</td>
<td>fc6</td>
<td>fc6</td>
<td>fc6</td>
<td>pool5</td>
<td>fc7</td>
<td>fc6</td>
</tr>
<tr>
<td>synthetic?</td>
<td>RGB</td>
<td>RGBD</td>
<td>RGB</td>
<td>RGB</td>
<td>disparity</td>
<td>disparity</td>
<td>HHA</td>
<td>HHA</td>
<td>HHA</td>
<td>HHA</td>
<td>RGB +HHA</td>
</tr>
<tr>
<td>mAP</td>
<td>8.4</td>
<td>21.7</td>
<td>16.4</td>
<td>19.7</td>
<td>11.3</td>
<td>20.1</td>
<td>25.2</td>
<td>26.1</td>
<td>21.9</td>
<td>25.3</td>
<td>32.5</td>
</tr>
</tbody>
</table>
## Object Detection
### Experiments

<table>
<thead>
<tr>
<th>finetuned?</th>
<th>layer</th>
<th>synthetic?</th>
<th>mAP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>DPM</td>
<td>yes</td>
<td>8.4</td>
</tr>
<tr>
<td>B</td>
<td>DPM</td>
<td></td>
<td>21.7</td>
</tr>
<tr>
<td>C</td>
<td>CNN</td>
<td>yes</td>
<td>16.4</td>
</tr>
<tr>
<td>D</td>
<td>CNN</td>
<td></td>
<td>19.7</td>
</tr>
<tr>
<td>E</td>
<td>CNN</td>
<td>yes</td>
<td>11.3</td>
</tr>
<tr>
<td>F</td>
<td>CNN</td>
<td></td>
<td>20.1</td>
</tr>
<tr>
<td>G</td>
<td>CNN</td>
<td>yes</td>
<td>25.2</td>
</tr>
<tr>
<td>H</td>
<td>CNN</td>
<td></td>
<td>26.1</td>
</tr>
<tr>
<td>I</td>
<td>CNN</td>
<td>yes</td>
<td>21.9</td>
</tr>
<tr>
<td>J</td>
<td>CNN</td>
<td></td>
<td>25.3</td>
</tr>
<tr>
<td>K</td>
<td>CNN</td>
<td></td>
<td>32.5</td>
</tr>
</tbody>
</table>

### Test Set

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

Our
## Object Detection

### Experiments

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

### Test Set

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

**Our** results are shown in bold.
Overview

Input

Re-organization

Recognition

Extensions

Color and Depth Image Pair

Contour Detection

Region Proposal Generation

Object Detection

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, ...

<table>
<thead>
<tr>
<th></th>
<th>Silberman et al. ECCV 12</th>
<th>Ren et al. CVPR 12</th>
<th>Gupta et al. CVPR 13</th>
<th>Gupta et al. (13) + RGB-D DPM</th>
<th>Gupta et al. (13) + Our Obj Det.</th>
</tr>
</thead>
<tbody>
<tr>
<td>fWavacc</td>
<td>38.2</td>
<td>37.6</td>
<td>43.4</td>
<td>45.2</td>
<td>47</td>
</tr>
<tr>
<td>Avacc</td>
<td>19</td>
<td>20.5</td>
<td>24.3</td>
<td>27.3</td>
<td>28.6</td>
</tr>
<tr>
<td>Mean (MaxIU)</td>
<td>-</td>
<td>21.4</td>
<td>27.9</td>
<td>29.6</td>
<td>31.3</td>
</tr>
<tr>
<td>Pixel Acc</td>
<td>54.6</td>
<td>49.3</td>
<td>57.9</td>
<td>59</td>
<td>60.3</td>
</tr>
<tr>
<td>Obj Avg</td>
<td>18.4</td>
<td>21.1</td>
<td>26.4</td>
<td>31.1</td>
<td>35.1</td>
</tr>
</tbody>
</table>

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

Re-organization

Recognition

Extensions

Color and Depth Image Pair

Contour Detection

Region Proposal Generation

Object Detection

Semantic Segm.

Instance Segm.

Input

Re-organization

Recognition

Extensions
Object Detection

**Task**

**Metric**

**Average Precision**

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

\[ \text{I/U} = \frac{\text{pred}}{\text{gt}} \]

assign true positive, ...

to compute precision and recall

\[ \text{AP} = \text{area under the Precision recall curve} \]
Instance Segmentation Task

Hariharan et al., ECCV14, Simultaneous Localization and Detection
Instance Segmentation

Task

Hariharan et al., ECCV14, Simultaneous Localization and Detection
Instance Segmentation Task

Hariharan et al., ECCV14, Simultaneous Localization and Detection
Instance Segmentation

Task

Hariharan et al., ECCV14, Simultaneous Localization and Detection
Instance Segmentation

Task

Hariharan et al., ECCV14, Simultaneous Localization and Detection
Instance Segmentation

Task

Hariharan et al., ECCV14, Simultaneous Localization and Detection
Instance Segmentation

Task

Hariharan et al., ECCV14, Simultaneous Localization and Detection
Instance Segmentation

Task

Hariharan et al., ECCV14, Simultaneous Localization and Detection
Instance Segmentation

Task

Hariharan et al., ECCV14, Simultaneous Localization and Detection
<table>
<thead>
<tr>
<th>Task</th>
<th>Metric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hariharan et al., ECCV'14, Simultaneous Localization and Detection</td>
<td></td>
</tr>
</tbody>
</table>
Instance Segmentation

**Task**

**Metric**

$\text{AP}_r$

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

$\text{I/U} = \frac{\text{pred}}{+} + \frac{\text{gt}}{+}$

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

$\text{AP}_r = \text{area under the Precision-Recall curve}$

Hariharan et al., ECCV14, Simultaneous Localization and Detection
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

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

$p + \delta_1$
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

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 \textit{(region average precision, AP\textsubscript{r})}
Instance Segmentation

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

Test Set

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>bathtub</th>
<th>bed</th>
<th>book shelf</th>
<th>box</th>
<th>chair</th>
<th>counter</th>
<th>desk</th>
<th>door</th>
<th>dresser</th>
<th>garbage</th>
<th>bin</th>
<th>lamp</th>
<th>monitor</th>
<th>night stand</th>
<th>pillow</th>
<th>sink</th>
<th>sofa</th>
<th>table</th>
<th>television</th>
<th>toilet</th>
</tr>
</thead>
<tbody>
<tr>
<td>box</td>
<td>14</td>
<td>6</td>
<td>40</td>
<td>4</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td>3</td>
<td>15</td>
<td>27</td>
<td>33</td>
<td>1</td>
<td>40</td>
<td>11</td>
<td>6</td>
<td>9</td>
<td>14</td>
<td>3</td>
<td>35</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>region</td>
<td>28</td>
<td>32</td>
<td>55</td>
<td>9</td>
<td>1</td>
<td>27</td>
<td>21</td>
<td>9</td>
<td>20</td>
<td>29</td>
<td>37</td>
<td>26</td>
<td>48</td>
<td>39</td>
<td>33</td>
<td>31</td>
<td>31</td>
<td>10</td>
<td>34</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>fg mask</td>
<td>28</td>
<td>15</td>
<td>60</td>
<td>9</td>
<td>1</td>
<td>29</td>
<td>5</td>
<td>7</td>
<td>23</td>
<td>33</td>
<td>38</td>
<td>31</td>
<td>55</td>
<td>39</td>
<td>32</td>
<td>32</td>
<td>36</td>
<td>11</td>
<td>37</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>our</td>
<td>32</td>
<td>19</td>
<td>66</td>
<td>10</td>
<td>2</td>
<td>36</td>
<td>33</td>
<td>10</td>
<td>23</td>
<td>34</td>
<td>38</td>
<td>36</td>
<td>53</td>
<td>43</td>
<td>32</td>
<td>34</td>
<td>41</td>
<td>14</td>
<td>37</td>
<td>51</td>
<td></td>
</tr>
</tbody>
</table>
## Instance Segmentation

### Performance (region average precision, AP$_r$)

#### Test Set

<table>
<thead>
<tr>
<th></th>
<th>bath</th>
<th>bed</th>
<th>book shelf</th>
<th>box</th>
<th>chair</th>
<th>counter</th>
<th>desk</th>
<th>door</th>
<th>dresser</th>
<th>garbage</th>
<th>bin</th>
<th>lamp</th>
<th>monitor</th>
<th>night stand</th>
<th>pillow</th>
<th>sink</th>
<th>sofa</th>
<th>table</th>
<th>television</th>
<th>toilet</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>box</strong></td>
<td>14</td>
<td>6</td>
<td>40</td>
<td>4</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td>3</td>
<td>15</td>
<td>27</td>
<td>33</td>
<td>1</td>
<td>40</td>
<td>11</td>
<td>6</td>
<td>9</td>
<td>14</td>
<td>3</td>
<td>35</td>
<td>12</td>
</tr>
<tr>
<td><strong>region</strong></td>
<td>28</td>
<td>32</td>
<td>55</td>
<td>9</td>
<td>1</td>
<td>27</td>
<td>21</td>
<td>9</td>
<td>20</td>
<td>29</td>
<td>37</td>
<td>26</td>
<td>48</td>
<td>39</td>
<td>33</td>
<td>31</td>
<td>10</td>
<td>34</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td><strong>fg mask</strong></td>
<td>28</td>
<td>15</td>
<td>60</td>
<td>9</td>
<td>1</td>
<td>29</td>
<td>5</td>
<td>7</td>
<td>23</td>
<td>33</td>
<td>38</td>
<td>31</td>
<td>55</td>
<td>39</td>
<td>32</td>
<td>32</td>
<td>36</td>
<td>11</td>
<td>37</td>
<td>38</td>
</tr>
<tr>
<td><strong>our</strong></td>
<td>32</td>
<td>19</td>
<td>66</td>
<td>10</td>
<td>2</td>
<td>36</td>
<td>33</td>
<td>10</td>
<td>23</td>
<td>34</td>
<td>38</td>
<td>36</td>
<td>53</td>
<td>43</td>
<td>32</td>
<td>34</td>
<td>41</td>
<td>14</td>
<td>37</td>
<td>51</td>
</tr>
<tr>
<td><strong>box AP</strong></td>
<td>37</td>
<td>44</td>
<td>71</td>
<td>33</td>
<td>1</td>
<td>43</td>
<td>44</td>
<td>15</td>
<td>24</td>
<td>30</td>
<td>39</td>
<td>37</td>
<td>53</td>
<td>40</td>
<td>35</td>
<td>36</td>
<td>54</td>
<td>24</td>
<td>38</td>
<td>47</td>
</tr>
</tbody>
</table>
## Instance Segmentation

### Performance

(region average precision, $AP^r$)

### Test Set

<table>
<thead>
<tr>
<th></th>
<th>mean bath tub</th>
<th>bed</th>
<th>book shelf</th>
<th>box</th>
<th>chair</th>
<th>counter</th>
<th>desk</th>
<th>door</th>
<th>dresser</th>
<th>garbage bin</th>
<th>lamp</th>
<th>monitor</th>
<th>night stand</th>
<th>pillow</th>
<th>sink</th>
<th>sofa</th>
<th>table</th>
<th>television</th>
<th>toilet</th>
</tr>
</thead>
<tbody>
<tr>
<td>box</td>
<td>14</td>
<td>6</td>
<td>40</td>
<td>4</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td>3</td>
<td>15</td>
<td>27</td>
<td>33</td>
<td>1</td>
<td>40</td>
<td>11</td>
<td>6</td>
<td>9</td>
<td>14</td>
<td>3</td>
<td>35</td>
</tr>
<tr>
<td>region</td>
<td>28</td>
<td>32</td>
<td>55</td>
<td>9</td>
<td>1</td>
<td>27</td>
<td>21</td>
<td>9</td>
<td>20</td>
<td>29</td>
<td>37</td>
<td>26</td>
<td>48</td>
<td>39</td>
<td>33</td>
<td>31</td>
<td>10</td>
<td>34</td>
<td>40</td>
</tr>
<tr>
<td>fg mask</td>
<td>28</td>
<td>15</td>
<td>60</td>
<td>9</td>
<td>1</td>
<td>29</td>
<td>5</td>
<td>7</td>
<td>23</td>
<td>33</td>
<td>38</td>
<td>31</td>
<td>55</td>
<td>39</td>
<td>32</td>
<td>36</td>
<td>11</td>
<td>37</td>
<td>38</td>
</tr>
<tr>
<td>our</td>
<td>32</td>
<td>19</td>
<td>66</td>
<td>10</td>
<td>2</td>
<td>36</td>
<td>33</td>
<td>10</td>
<td>23</td>
<td>34</td>
<td>38</td>
<td>36</td>
<td>53</td>
<td>43</td>
<td>32</td>
<td>34</td>
<td>14</td>
<td>37</td>
<td>51</td>
</tr>
<tr>
<td>box AP</td>
<td>37</td>
<td>44</td>
<td>71</td>
<td>33</td>
<td>1</td>
<td>43</td>
<td>44</td>
<td>15</td>
<td>24</td>
<td>30</td>
<td>39</td>
<td>37</td>
<td>53</td>
<td>40</td>
<td>35</td>
<td>36</td>
<td>24</td>
<td>38</td>
<td>47</td>
</tr>
</tbody>
</table>

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

Input

Color and Depth Image Pair
<table>
<thead>
<tr>
<th>Overview</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Input</strong></td>
<td><strong>Re-organization</strong></td>
</tr>
<tr>
<td><img src="image1.png" alt="Input Image" /></td>
<td></td>
</tr>
<tr>
<td><img src="image2.png" alt="Color and Depth" /></td>
<td></td>
</tr>
</tbody>
</table>

**Color and Depth Image Pair**
## Overview

<table>
<thead>
<tr>
<th>Input</th>
<th>Re-organization</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Color and Depth Image Pair" /></td>
<td><img src="image2.png" alt="Contour Detection" /></td>
</tr>
</tbody>
</table>

### Contour Detection
Overview

Input

Color and Depth Image Pair

Re-organization

Contour Detection

Region Proposal Generation
Overview

Input

Re-organization

Recognition

Color and Depth Image Pair

Contour Detection

Region Proposal Generation

Object Detection
Overview

Input

Re-organization

Recognition

Extensions

Contour Detection

Region Proposal Generation

Object Detection

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

Re-organization

Color and Depth Image Pair

Region Proposal Generation

Contour Detection

Recognition

Object Detection

Extensions

Semantic Segm.

Instance Segm.

State-of-the-art
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

Re-organization

Recognition

Extensions

Contour Detection

Region Proposal Generation

Object Detection

Semantic Segm.

Instance Segm.

Color and Depth Image Pair

State-of-the-art

Code available online!

Thank You