Synthesizing Training Data for Object Detection in Indoor Scenes

Georgios Georgakis
Arsalan Mousavian
Alexander C. Berg
Jana Kosecka

GEORGE MASON UNIVERSITY

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

Object Detection in Indoor Scenes

- Given an RGB-D image ***I***, localize all bounding boxes ***{x,y,w,h}***, of object instances of interest ***c*** - instance recognition.

Fetch these objects

GEORGE MASON UNIVERSITY

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

Challenges

Viewpoint variation

Clutter and scale variations

Occlusions

- State-of-the-art object detectors SSD / Faster R-CNN – deep learning.
- These methods require large amounts of training data.
- Manual annotation is time consuming.

- Liu et al. “Single shot multibox detector”, ECCV 2016.
- Ren et al. “Faster R-CNN: Towards real-time object detection with region proposal networks”, NIPS 2015

GEORGE MASON UNIVERSITY

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

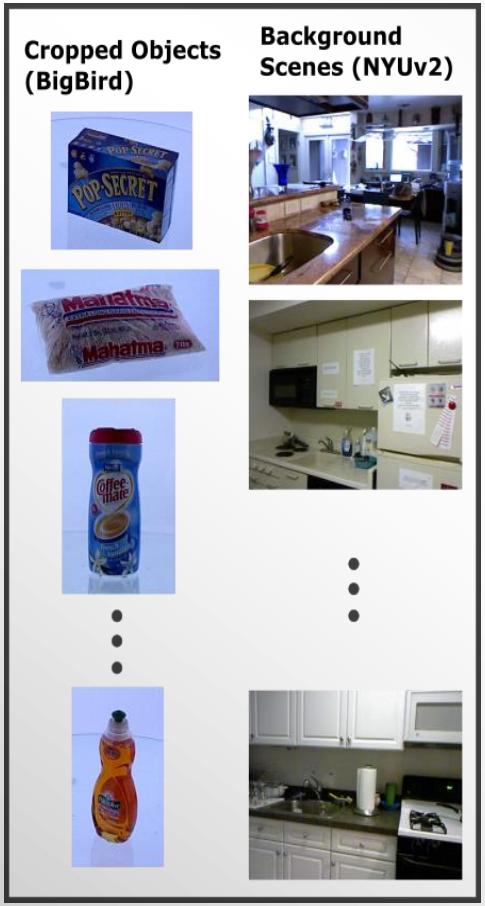
Question

How can we train SSD and Faster R-CNN with the least amount of manual annotation?

GEORGE MASON UNIVERSITY

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

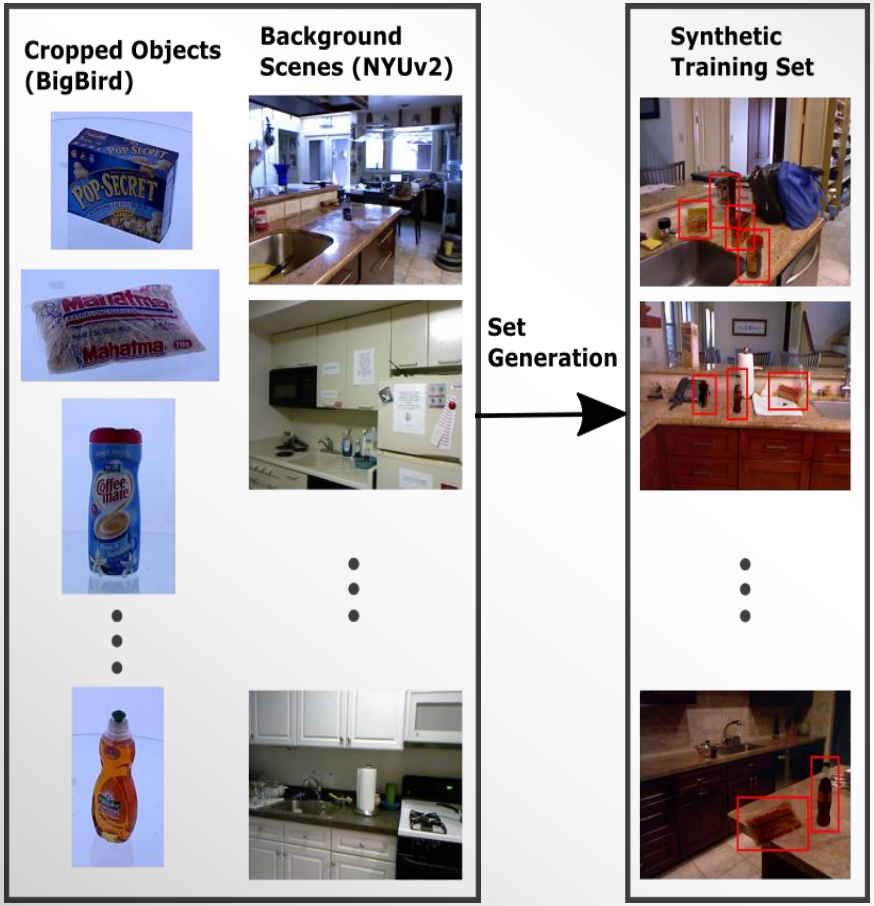
System Overview



GEORGE MASON UNIVERSITY

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

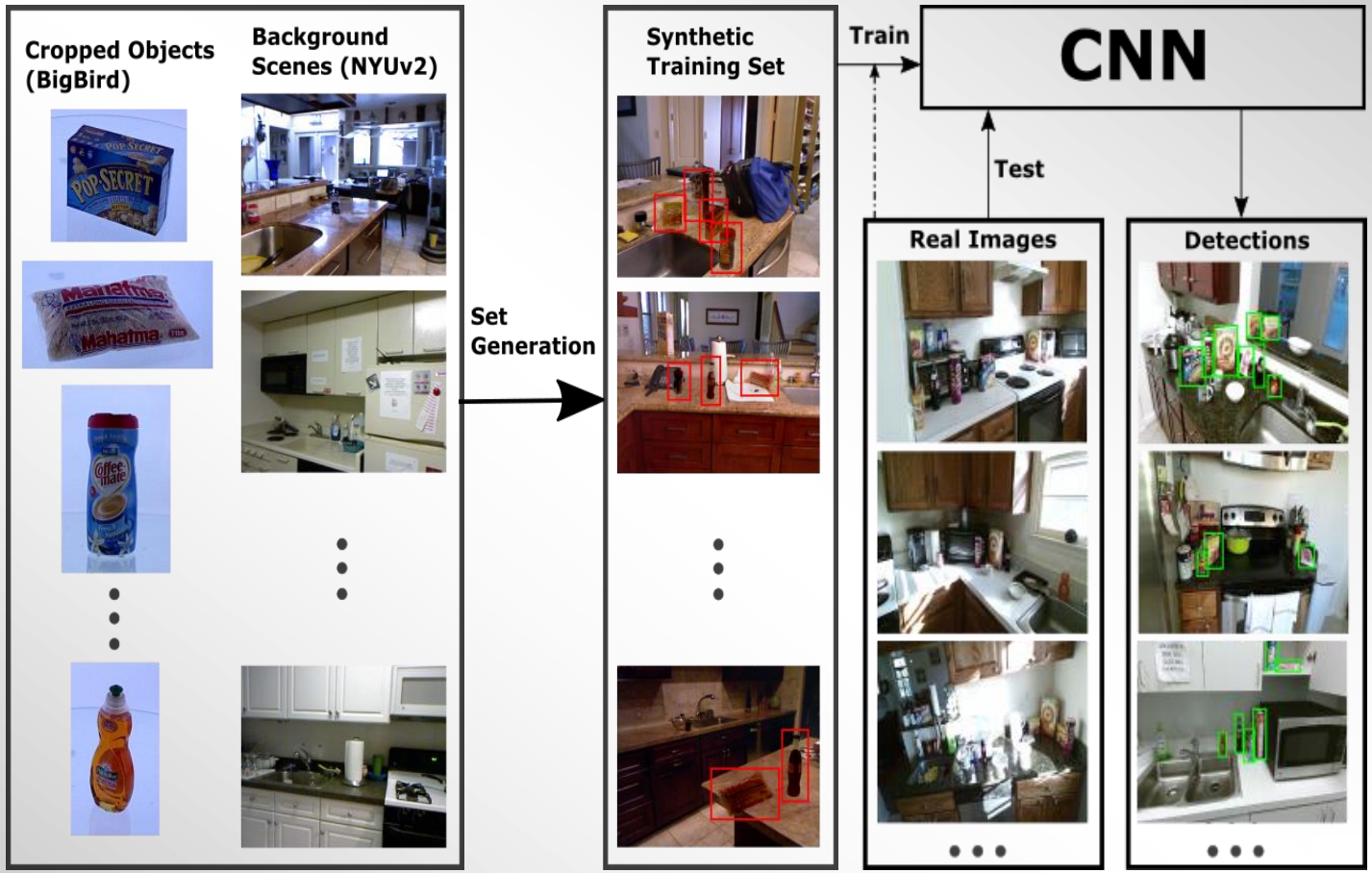
System Overview



GEORGE MASON UNIVERSITY

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

System Overview



GEORGE MASON UNIVERSITY

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

Synthetic Set Generation

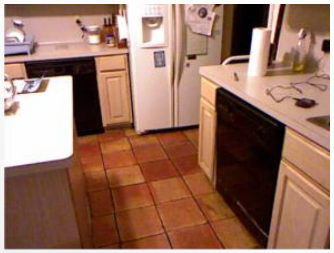
- Important observations:
  - Household objects are found on support surfaces.
  - The distance from camera defines the scale.
  - Objects in natural images have smooth boundary discontinuities with the background.

- Three simple constraints for increased realism:
  - *Selective positioning*, *Selective scaling*, and *Blending*.

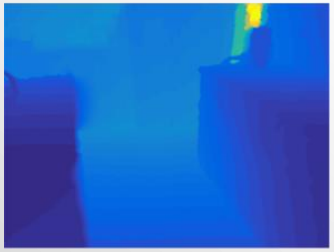
GEORGE MASON UNIVERSITY

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

Automatic Generation



RGB

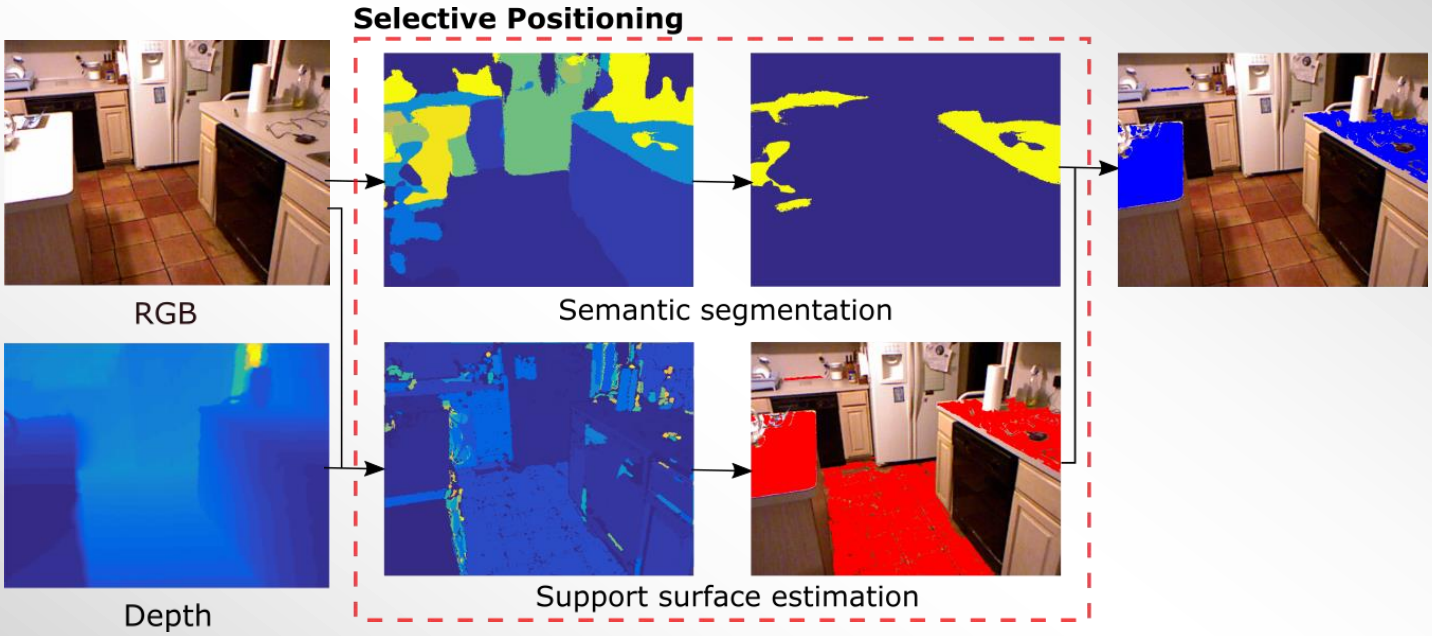


Depth

GEORGE MASON UNIVERSITY

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

Automatic Generation



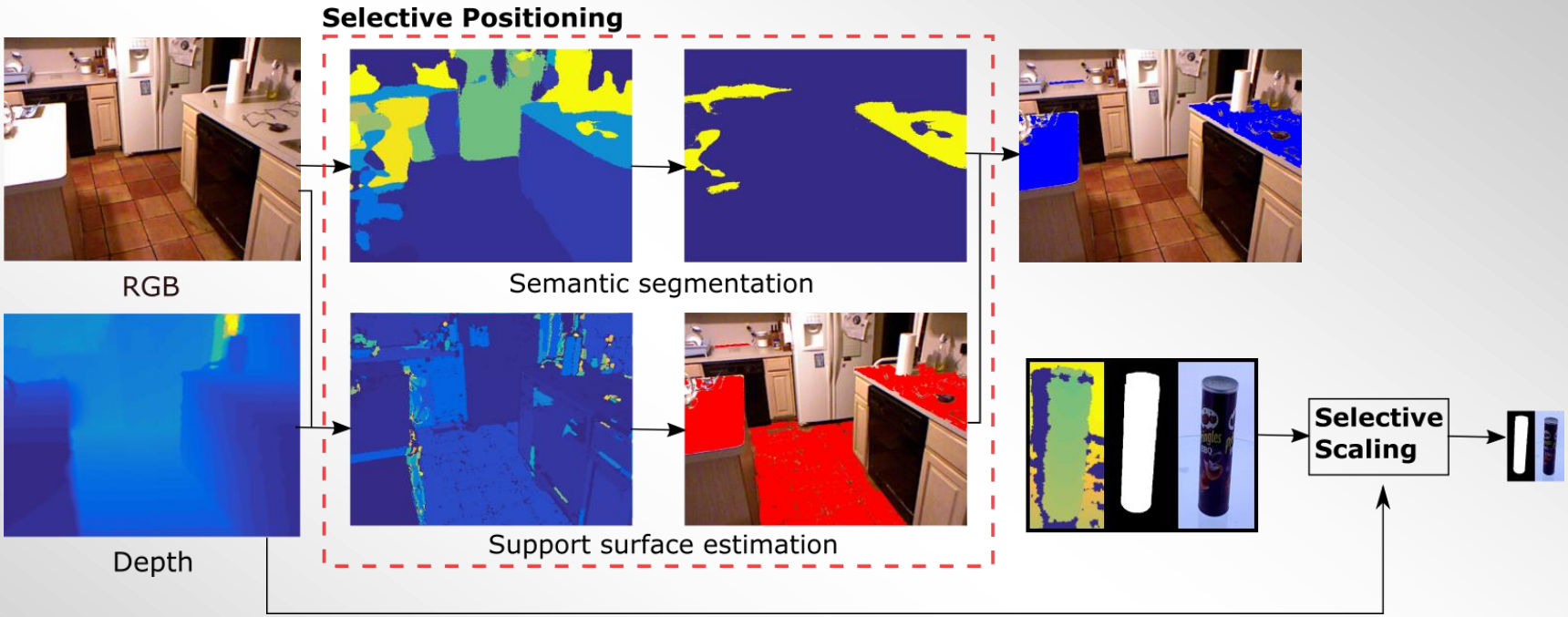
Semantic segmentation state-of-the-art approach:

- Mousavian et al. “Joint semantic segmentation and depth estimation with deep convolutional networks”, 3DV 2016.

GEORGE MASON UNIVERSITY

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

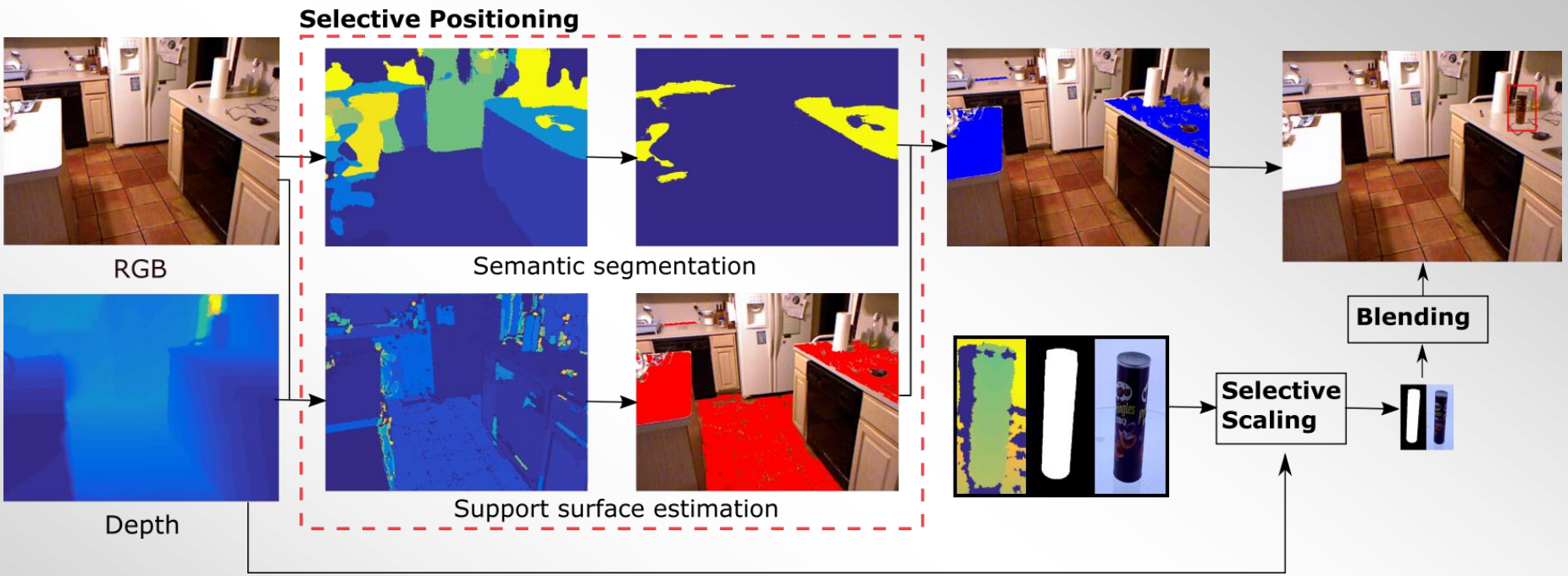
Automatic Generation



GEORGE MASON UNIVERSITY

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

Automatic Generation



Off-the-shelf blending technique:

- Tanaka et al. “Seamless image cloning by a closed form solution of a modified Poisson problem”, SIGGRAPH Asia 2012.

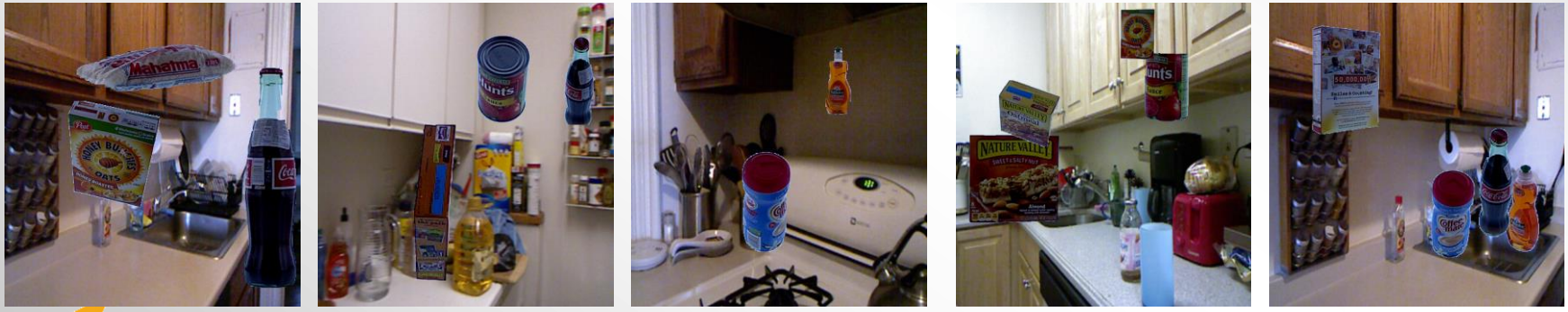
GEORGE MASON UNIVERSITY

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

Composited Examples

Our approach

Randomized approach



GEORGE MASON UNIVERSITY

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

Experiments

- GMU-Kitchens dataset
  - Georgakis et al. "Multiview RGB-D dataset for object instance detection", 3DV 2016
- Synth to Real
  - Informatively generated synthetic sets (~20000 images) outperform random sets.
- Synth+Real to Real
  - Full real training set ~4000 images.
  - Augmenting existing few real annotations (400 real + 20000 synthetic) can yield comparable or superior results.

Train Set	SSD / Faster R-CNN
Real to Real	
1. Real Scenes	65.6 / 82.5
Synthetic to Real	
2. RP-SI-RS	23.2 / 42.1
3. RP-BL-RS	22.2 / 44.2
4. SP-SI-SS	19.5 / 48.6
5. SP-BL-SS	**33.5 / 51.7**
Synthetic+Real to Real	
6. 1% real	60.3 / 69.3
7. 10% real	71.6 / 79.2
8. 50% real	**74.1** / 83.8
9. 100% real	73.7 / **85.0**

GEORGE MASON UNIVERSITY

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

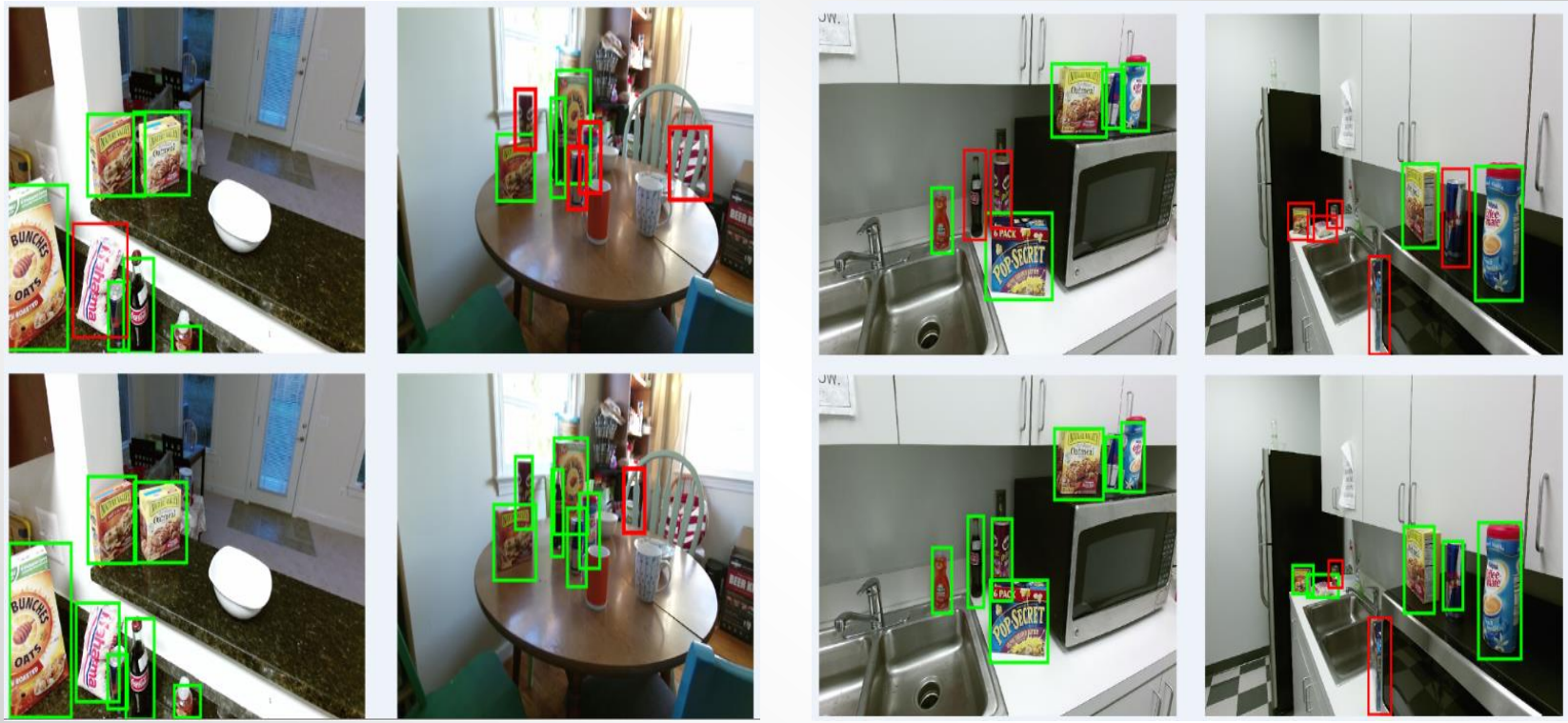
Qualitative Results

Faster R-CNN

SSD

Real

Real + Synth



GEORGE MASON UNIVERSITY

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

Summary

- Synthetic training data are more effective when they are generated with semantic and geometric information.

- Object detectors can be trained with significantly less annotated data using our proposed synthetic data augmentation.

GEORGE MASON UNIVERSITY

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL