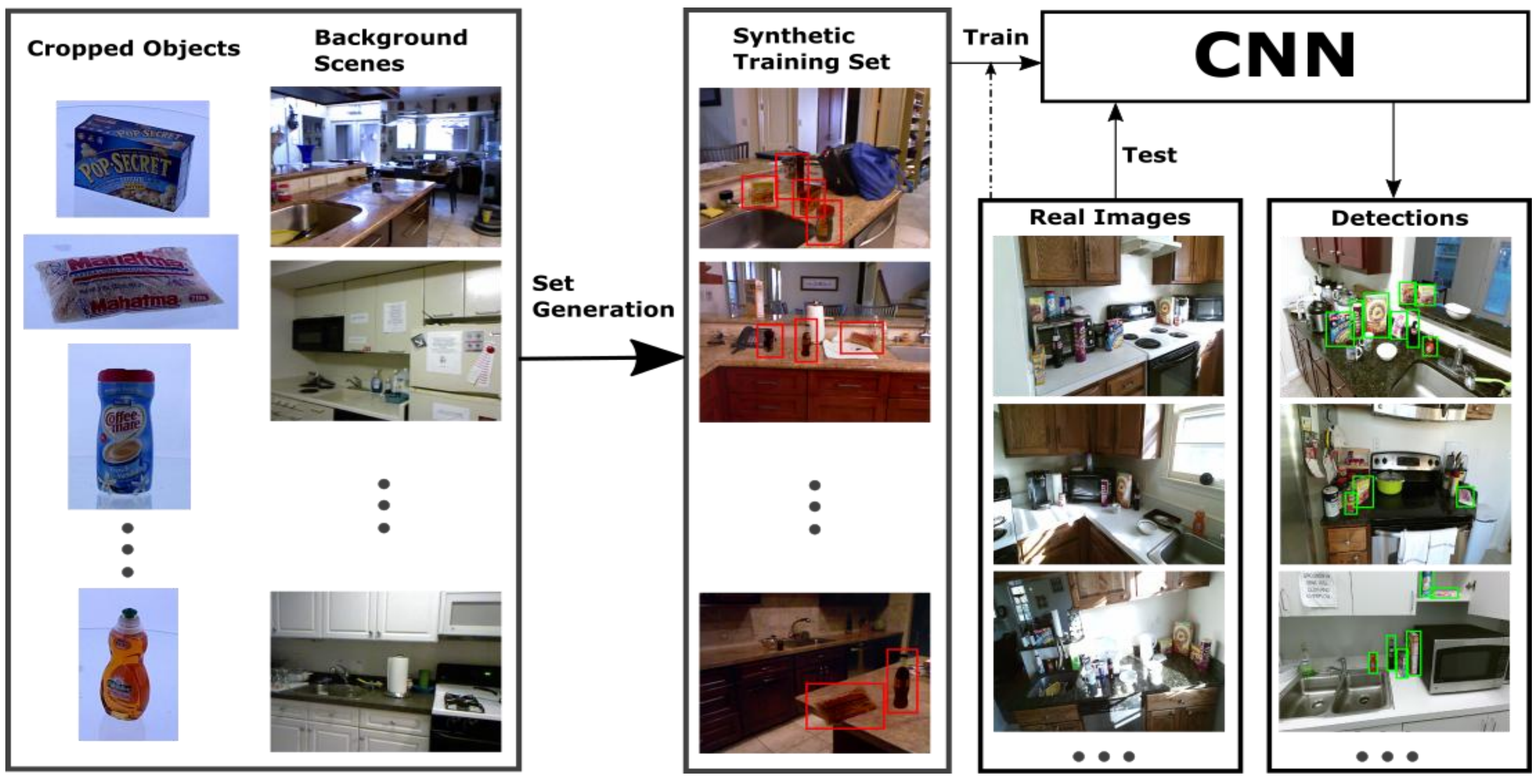
Synthesizing Training Data for Object Detection in Indoor Scenes

Georgios Georgakis[1], Arsalan Mousavian[1], A. C. Berg[2], Jana Kosecka[1]
[1]George Mason University, [2]University of North Carolina Chapel Hill

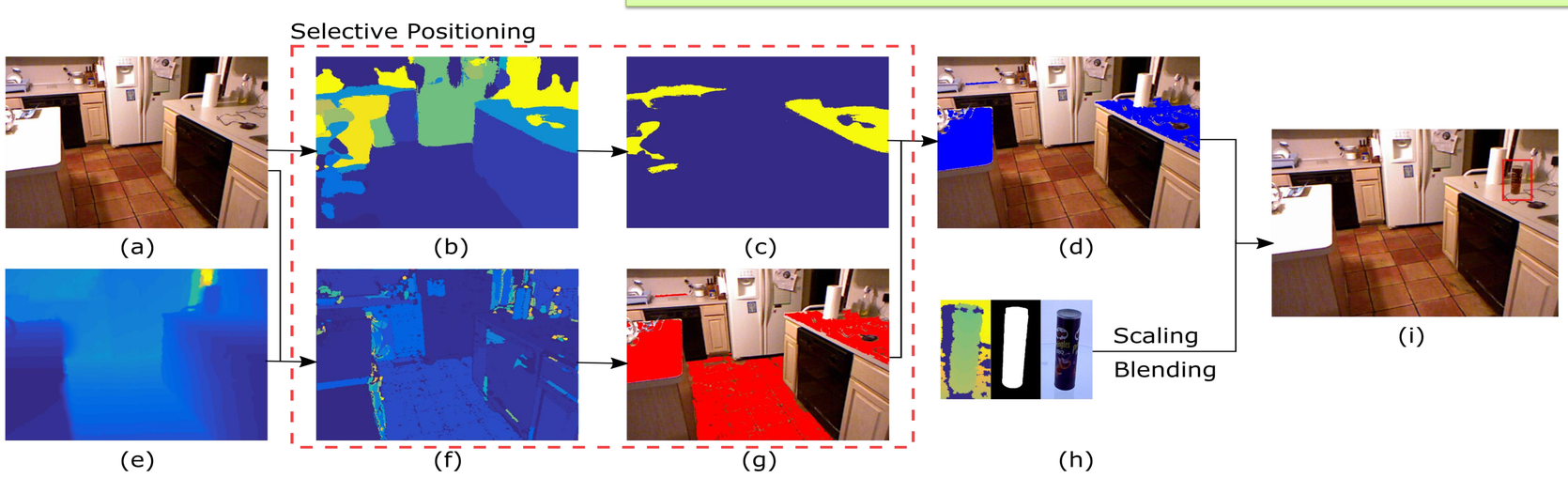
Overview



Problem:
- Training of object detectors requires large amounts of annotated data.
- Manual annotation of bounding boxes is very time-consuming.

Contributions:
- Automated generation approach exploiting geometry and semantic information.
- Offer insights on how training data should be generated in the future.
- Extensive evaluation of state-of-the-art object detectors.

Synthetic Set Generation



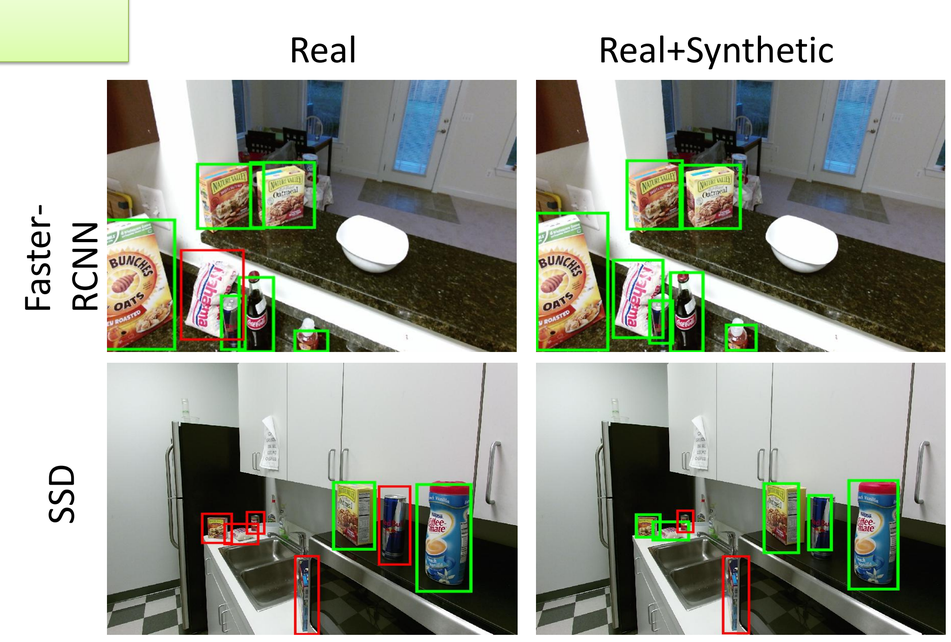
(a) RGB
(b) Semantic segmentation
(c) Counters and tables predictions
(d) Regions for object placement
(e) Depth
(f) Plane extraction
(g) Support Surfaces
(h) Object RGB, mask, and depth
(i) Final blending result

Experiments and Results

Train Set	GMU-Kitchens	WRGB-D
Real to Real		
1. **Real Scenes**	65.6 / 82.5	90.2 / 97.9
Synthetic to Real		
2. **RP-SI-RS**	23.2 / 42.1	78.2 / 52.0
3. **RP-BL-RS**	22.2 / 44.2	72.6 / 70.1
4. **SP-SI-SS**	19.5 / 48.6	**79.5** / 73.4
5. **SP-BL-SS**	**33.5 / 51.7**	77.5 / **78.1**
Synthetic+Real to Real		
6. **1% real**	60.3 / 69.3	88.9 / 96.2
7. **10% real**	71.6 / 79.2	90.5 / 97.8
8. **50% real**	**74.1** / 83.8	90.7 / **98.2**
9. **100% real**	73.7 / **85.0**	**90.8 / 98.2**
Synthetic to Synthetic		
10. **RP-SI-RS**	**90.8 / 99.5**	**90.5 / 97.4**
11. **SP-BL-SS**	83.8 / 85.7	90.1 / 96.2

TABLE I
AVERAGE PRECISION RESULTS (SSD / FASTER-RCNN).

1) Real to Real: Trained only on the real manually annotated data.
2) Synthetic to Real: Object detectors are trained on various synthetic sets to determine the effect of positioning and scaling.
3) Synthetic + Real to Real: Combination of synthetic and varying proportions of real data.
4) Synthetic to Synthetic: Synthetic data for training and testing to observe the reduction of over-fitting.



Synthetic sets:
- Random/Selective Positioning, (**RP**)/(**SP**)
- Random/Selective Scaling, (**RS**)/(**SS**)
- Simple superimposition/Blending, (**SI**)/(**BL**)

Conclusions

- Object detectors can be trained with significantly less annotated data using our proposed synthetic data augmentation.
- When the object placements are based on semantic and geometric information the synthetic training data are more effective.

Acknowledgments: We acknowledge support from NSF NRI grant 1527208.

References

1. X. Peng, B. Sun, K. Ali, K. Saenko, Learning Deep Object Detectors from 3d Models , ICCV, 2015.
2. A. Singh, J. Sha, K. Narayan, T.Achim, P. Abeel, A Large-scale 3d Database of Object Instances, ICRA 2014.
3. G. Georgakis, Md. Reza, A. Mousavian, P. H. Le, J. Kosecka, Multiview RGB-D dataset for Object Instance Detection, 3DV, 2016.
4. H. Su, R. Charles, Y. Li, L. Guibas, Render for CNN: Viewpoint Estimation in Images Using CNNS Trained with Rendered 3D Model Views, ICCV, 2015