Cut-and-Paste Synthetic Data Generation for Instance Segmentation

LINKÖPINGS UNIVERSITET

Asma Raeisi, Gabriel Eilertsen, Jonas Unger
Linköping University, Norrköping, Sweden
Email: {asma.raeisi, gabriel.eilertsen, jonas.unger}@liu.se

Introduction

Synthetic data is increasingly used for training and evaluating deep neural networks in computer vision tasks due to the difficulty of obtaining real-world ground truth data [1]. However, it presents challenges, the reality gap, where networks trained on synthetic data may not perform well on real-world data without additional fine-tuning [2]. This project aims to develop techniques for synthetic data generation, explore its effects on the reality gap, and investigate how synthesis and augmentation can be integrated as integral parts in a training feedback loop.

Methodology

Synthetic data generation process:

Object Extraction

- Randomly select one to five objects
- Create a binary mask for object shape
- Extract objects with masks using bitwise AND operation

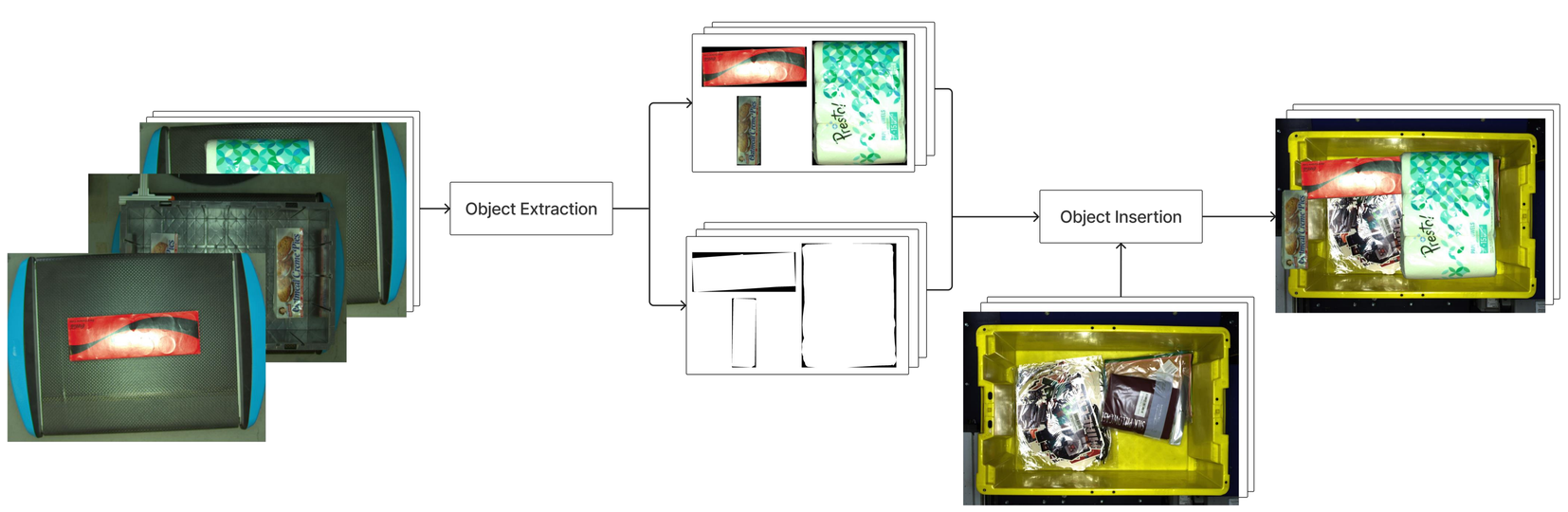
Object Insertion

Two blending methods:

- Objects with sharp edges
- Objects with blurred edges via Gaussian blur

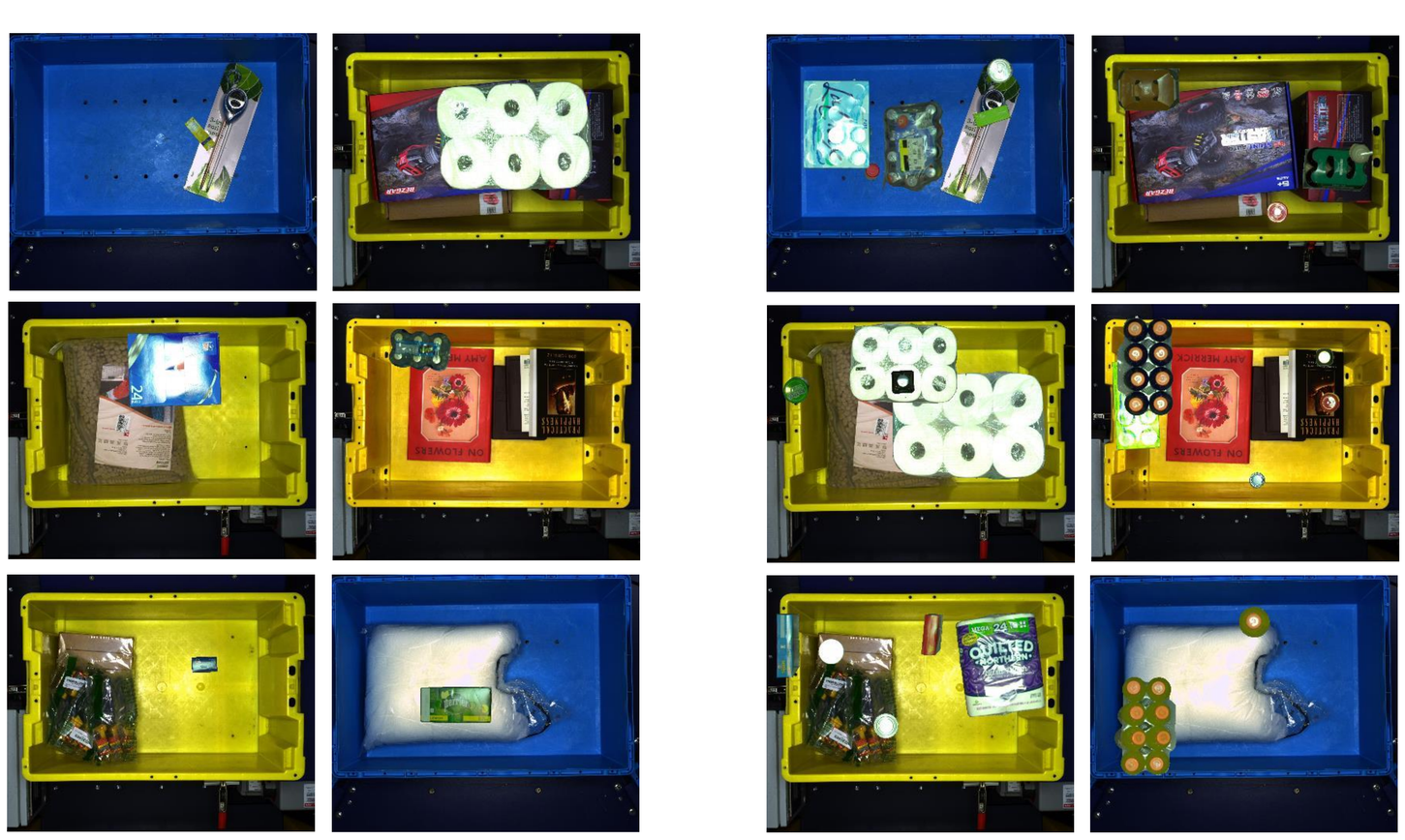
Performance evaluation:

- Instance segmentation using **Mask R-CNN** [3]



Results

Figures display synthetic images with sharp edge objects, featuring both single and multiple objects randomly placed on the background.



Conclusion

- Blurred edges didn't improve results
- Larger dataset led to longer training with minimal gains
- Integration of real and synthetic data improves generalization on real data

Future work:

Exploring sensor simulation using:

- **Computer graphics rendering**
- **Domain randomization**
- **Generative models**

References

[1] J. Tremblay, et al. “Training Deep Networks with Synthetic Data: Bridging The Reality Gap by Domain Randomization,” in Proceedings of the *IEEE conference on computer vision and pattern recognition workshops*, 2018, pp. 969–977.

[2] J. Tremblay, et al. “Deep Object Pose Estimation for Semantic Robotic Grasping of Household Objects*,” arXiv preprint arXiv*:1809.10790, 2018.

[3] K. He, et al, “Mask R-CNN,” in Proceedings of the *IEEE international conference on computer vision*, 2017, pp. 2961–2969.

[4] C. Mitash, et al. “ARMBENCH: an Object-centric Benchmark Dataset for Robotic Manipulation,” *arXiv preprint arXiv*:2303.16382, 2023.

- Tables 1 and 2 compare Mask R-CNN performance for instance segmentation across different scenarios.

Table 1: Model trained on real data

Model	mAP50	mAP75
Model in [4]	72	61
Aug-Model	78.48	57.52
Baseline	**80.10**	**73.07**

Table 2: Model trained on synthetic data

Model	mAP50	mAP75
SingleObj*sharp*	79.56	72.75
SingleObj*blur*	79.13	72.41
MultiObj*sharp*	78.47	71.51
MultiObj*blur*	78.25	71.52
LargeSet-MultiObj*sharp*	78.58	71.65
LargeSet-MultiObj*blur*	78.18	71.57
LargeSet-Real-SingleObj*sharp*	**80.31**	**73.30**

WASP WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM