Supplementary Results for Robotic Pick-and-Place With Uncertain Object Instance Segmentation and Shape Completion

Marcus Gualtieri and Robert Platt

Results for all tasks and datasets are included here for completeness. At the beginning of each section, we describe the task and the object datasets used for evaluation. Results are in simulation except under the subsection heading "Real robot experiments". Source code and links to object datasets are available online at `https://github.com/mgualti/GeomPickPlace`.

1 Bin Packing

The robot is presented with six novel objects which all must be packed into a rectangular bin such as to minimize the height of the final packing. *Train* refers to the same object models used for training all deep networks, *Test-1* refers to same-category novel objects, and *Test-2* refers to novel-category objects. Train and Test-1 categories are boat, bottle, box, car, dinosaur, mug, and wine glass. Test-2 categories are airplane, bowl, and stapler. Object models are from 3DNet [4].

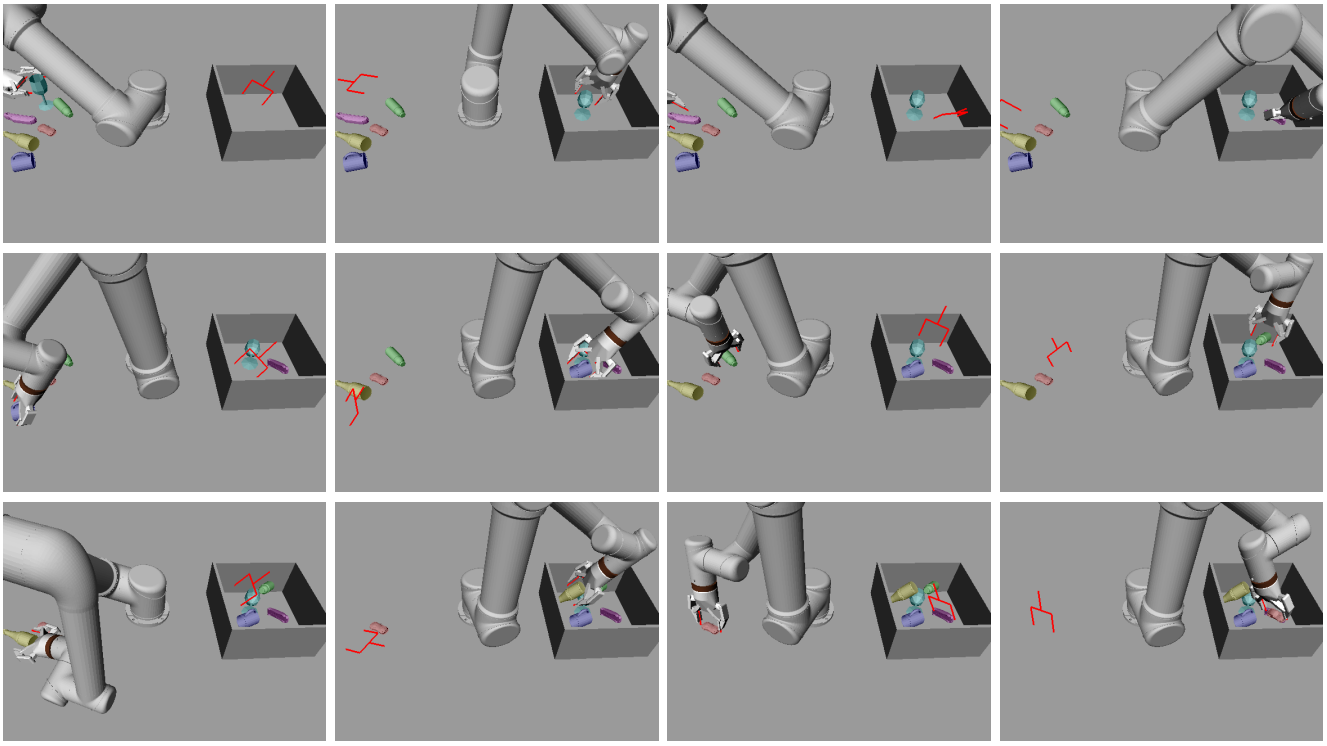


Figure 1: A sequence of picks and places for one episode of the bin packing task.

Figure 2: Objects used for bin packing on a real robot.

1.1 Perception ablation study

	GT Seg. & Comp.	GT Seg. (Train)	GT Seg. (Test-1)	Percep. (Train)	Percep. (Test-1)	GT Seg. & No Comp.
Place Execution Success	0.929 ± 0.008	0.767 ± 0.013	0.747 ± 0.013	0.718 ± 0.014	0.710 ± 0.014	0.508 ± 0.046
Regrasp Plan Found	0.957 ± 0.006	0.882 ± 0.009	0.939 ± 0.007	0.879 ± 0.009	0.941 ± 0.007	0.100 ± 0.009
Grasp Antipodal	0.931 ± 0.007	0.779 ± 0.013	0.761 ± 0.013	0.755 ± 0.013	0.736 ± 0.013	0.563 ± 0.047
Temporary Place Stable	1.000 ± 0.000	0.769 ± 0.122	1.000 ± 0.000	0.828 ± 0.071	0.826 ± 0.081	0.500 ± 0.500
Packing height of 5 (cm)	12.27 ± 0.315	12.36 ± 0.331	12.18 ± 0.306	12.37 ± 0.447	12.44 ± 0.307	–
Regrasp planning time (s)	35.62 ± 1.103	38.46 ± 1.115	38.68 ± 1.141	35.76 ± 1.059	35.05 ± 1.077	15.86 ± 1.482

Table 1: Perception ablation study for packing, Test-1. Showing average ± standard error over 200 episodes.

	GT Seg. & Comp.	GT Seg.	Percep.	GT Seg. & No Comp.
Place Execution Success	0.849 ± 0.011	0.459 ± 0.017	0.432 ± 0.017	0.304 ± 0.034
Regrasp Plan Found	0.878 ± 0.009	0.708 ± 0.013	0.718 ± 0.013	0.151 ± 0.010
Grasp Antipodal	0.854 ± 0.011	0.478 ± 0.017	0.457 ± 0.017	0.337 ± 0.036
Temporary Place Stable	1.000 ± 0.000	0.786 ± 0.114	0.167 ± 0.112	0.500 ± 0.500
Packing height of 5 (cm)	8.894 ± 0.173	7.734 ± 0.408	11.68 ± 0.741	–
Regrasp planning time (s)	36.85 ± 1.614	26.50 ± 0.891	25.22 ± 0.869	23.007 ± 2.145

Table 2: Perception ablation study for packing, Test-2. Showing average ± standard error over 200 episodes.

1.2 Regrasp cost comparison

	No Cost	Step Cost	GQ	MC	MC + GQ	CU	SP
Place Execution Success	0.651 ± 0.013	0.725 ± 0.012	0.748 ± 0.012	0.756 ± 0.012	**0.787** ± 0.011	0.712 ± 0.013	0.779 ± 0.012
Grasp Antipodal	0.737 ± 0.011	0.751 ± 0.012	0.794 ± 0.011	0.811 ± 0.011	**0.830** ± 0.010	0.743 ± 0.012	0.823 ± 0.010
Temporary Place Stable	0.784 ± 0.024	0.857 ± 0.097	0.845 ± 0.030	0.904 ± 0.028	0.883 ± 0.031	0.848 ± 0.054	**0.959** ± 0.018
Plan Length	2.665 ± 0.031	**2.038** ± 0.008	2.293 ± 0.021	2.222 ± 0.019	2.201 ± 0.018	2.105 ± 0.013	2.233 ± 0.019
Regrasp planning time (s)	**4.904** ± 0.230	7.201 ± 0.393	84.56 ± 0.827	90.10 ± 0.892	126.5 ± 1.029	72.00 ± 0.835	86.61 ± 1.040

Table 3: Cost comparison for bin packing for Test-1. Showing average ± standard error over 230 episodes.

	No Cost, SP	Step Cost, SP	GQ, SP	No Cost, MC + GQ	Step Cost, MC + GQ,	GQ, MC + GQ
Place Execution Success	3.0×10^{-13}	7.0×10^{-4}	3.2×10^{-2}	5.5×10^{-15}	1.1×10^{-4}	9.2×10^{-3}
Grasp Antipodal	1.5×10^{-08}	3.5×10^{-6}	2.6×10^{-2}	7.2×10^{-10}	3.2×10^{-7}	7.3×10^{-3}
Temporary Place Stable	4.6×10^{-06}	5.2×10^{-2}	1.1×10^{-3}	1.1×10^{-02}	3.9×10^{-1}	1.9×10^{-1}

Table 4: p-values for 1-tailed, unpaired, same-variance t-test for select comparisons to baseline for Test-1. For values less than 0.05 (shown in green), we accept the hypothesis that the treatment (2nd method in column heading) resulted in an improvement over the baseline (1st method in column heading).

	No Cost	Step Cost	GQ	MC	MC + GQ	CU	SP
Place Execution Success	0.412 ± 0.017	0.417 ± 0.017	0.395 ± 0.017	0.458 ± 0.017	0.422 ± 0.017	0.429 ± 0.017	**0.465** ± 0.017
Grasp Antipodal	0.484 ± 0.017	0.449 ± 0.017	0.450 ± 0.017	0.504 ± 0.017	0.472 ± 0.017	0.457 ± 0.017	**0.518** ± 0.017
Temporary Place Stable	0.704 ± 0.051	0.714 ± 0.125	0.533 ± 0.075	0.750 ± 0.083	**0.800** ± 0.082	0.778 ± 0.101	0.686 ± 0.080
Plan Length	2.514 ± 0.036	**2.094** ± 0.015	2.247 ± 0.024	2.167 ± 0.020	2.150 ± 0.019	2.118 ± 0.017	2.193 ± 0.022
Regrasp planning time (s)	**6.030** ± 0.237	8.484 ± 0.408	51.61 ± 1.113	58.56 ± 1.064	71.38 ± 1.333	50.92 ± 1.177	53.35 ± 1.159

Table 5: Cost comparison for bin packing for Test-2. Showing average ± standard error over 200 episodes.

	No Cost, SP	Step Cost, SP	GQ, SP
Place Execution Success	1.3×10^{-2}	2.3×10^{-2}	1.6×10^{-3}
Grasp Antipodal	7.6×10^{-2}	2.3×10^{-3}	2.5×10^{-3}
Temporary Place Stable	5.8×10^{-1}	5.8×10^{-1}	8.6×10^{-2}

Table 6: p-values for 1-tailed, unpaired, same-variance t-test for select comparisons to baseline for Test-2. For values less than 0.05 (shown in green), we accept the hypothesis that the treatment (2nd method in column heading) resulted in an improvement over the baseline (1st method in column heading).

1.3 Real robot experiments

	Step Cost	GQ	MC	SP
Place Success Rate	0.839 ± 0.027	0.833 ± 0.028	9.111 ± 0.021	**0.917** ± 0.021
Grasp Success Rate	0.883 ± 0.023	0.866 ± 0.024	**0.947** ± 0.016	0.933 ± 0.017
Number of Regrasps	**17**	21	27	30
Packing height of 6 (cm)	9.010 ± 1.119	7.445 ± 0.434	7.722 ± 0.390	7.339 ± 0.241
Packing height of 5 (cm)	7.333 ± 0.858	7.050 ± 0.650	7.588 ± 1.132	7.711 ± 0.880

Table 7: Packing performance over 30 episodes. Showing average ± standard error, where applicable.

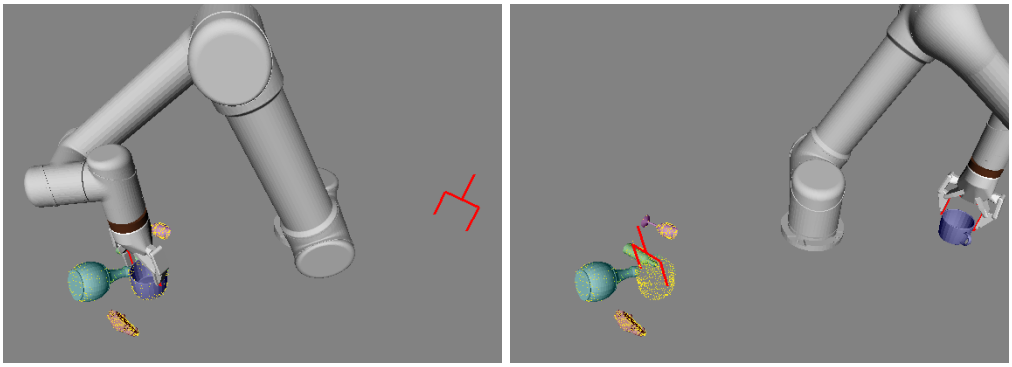
	Step Cost, MC	Step Cost, SP	GQ, MC	GQ, SP	Step Cost ∪ GQ, MC ∪ SP
Place Success Rate	1.9×10^{-2}	1.2×10^{-2}	1.4×10^{-2}	8.4×10^{-3}	7.9×10^{-4}
Grasp Success Rate	1.0×10^{-2}	3.8×10^{-2}	2.4×10^{-3}	1.1×10^{-2}	5.6×10^{-4}

Table 8: p-values for 1-tailed, unpaired, same-variance t-test for select comparisons. For values less than 0.05 (shown in green), we accept the hypothesis that the treatment (2nd method in column heading) resulted in an improvement over the baseline (1st method in column heading). The notation A ∪ B means the results for treatment A and B are aggregated.

2 Canonical Arrangement

For each episode, the robot must place any one of five objects into a given canonical goal pose. The canonical pose comes from an oracle arrangement planner which has access to the object model's reference

frame, so this task can only be evaluated in simulation. The object sets are the same as with the packing task (Train, Test-1, and Test-2).



2.1 Regrasp cost comparison

	No Cost	Step Cost	GQ	MC	MC + GQ	CU	SP
Place Execution Success	0.727 ± 0.010	0.777 ± 0.009	0.856 ± 0.008	0.852 ± 0.008	0.861 ± 0.008	0.830 ± 0.008	**0.913** ± 0.006
Grasp Antipodal	0.833 ± 0.007	0.824 ± 0.009	0.906 ± 0.006	0.902 ± 0.006	0.908 ± 0.006	0.857 ± 0.008	**0.951** ± 0.005
Temporary Place Stable	0.785 ± 0.015	0.623 ± 0.067	0.700 ± 0.031	0.852 ± 0.022	0.784 ± 0.030	0.885 ± 0.029	**0.967** ± 0.012
Plan Length	3.061 ± 0.029	**2.079** ± 0.009	2.273 ± 0.016	2.286 ± 0.016	2.220 ± 0.014	2.157 ± 0.013	2.239 ± 0.015
Regrasp planning time (s)	**2.462** ± 0.061	6.413 ± 0.353	62.19 ± 0.326	117.6 ± 0.724	121.1 ± 0.577	54.88 ± 0.366	61.54 ± 0.900

Table 9: Cost comparison for canonical task for Test-1. Showing average ± standard error over 2,000 episodes.

	No Cost, SP	Step, SP	GQ, SP
Place Execution Success	1.2×10^{-54}	1.9×10^{-33}	9.7×10^{-09}
Grasp Antipodal	1.4×10^{-38}	5.2×10^{-40}	3.8×10^{-09}
Temporary Place Stable	2.4×10^{-10}	2.9×10^{-15}	1.3×10^{-14}

Table 10: p-values for 1-tailed, unpaired, same-variance t-test for select comparisons to baseline for Test-1. For values less than 0.05 (shown in green), we accept the hypothesis that the treatment (2nd method in column heading) resulted in an improvement over the baseline (1st method in column heading).

	No Cost	Step Cost	GQ	MC	MC + GQ	CU	SP
Place Execution Success	0.446 ± 0.011	0.535 ± 0.012	0.520 ± 0.012	0.543 ± 0.012	0.566 ± 0.012	0.533 ± 0.012	**0.591** ± 0.011
Grasp Antipodal	0.585 ± 0.010	0.592 ± 0.011	0.612 ± 0.011	0.630 ± 0.011	0.650 ± 0.011	0.590 ± 0.011	**0.674** ± 0.010
Temporary Place Stable	0.690 ± 0.021	0.555 ± 0.046	0.608 ± 0.030	0.717 ± 0.032	0.621 ± 0.034	0.671 ± 0.036	**0.742** ± 0.027
Plan Length	3.265 ± 0.035	**2.323** ± 0.018	2.686 ± 0.025	2.501 ± 0.022	2.474 ± 0.021	2.419 ± 0.020	2.518 ± 0.023
Regrasp planning time (s)	**4.278** ± 0.156	14.84 ± 0.539	68.87 ± 0.657	99.36 ± 0.818	99.02 ± 0.819	60.05 ± 0.633	74.08 ± 0.732

Table 11: Packing performance for Test-2 over 2,000 episodes. Showing average ± standard error.

	No Cost, SP	Step, SP	GQ, SP
Place Execution Success	1.6×10^{-19}	2.9×10^{-4}	5.3×10^{-6}
Grasp Antipodal	9.0×10^{-10}	5.4×10^{-8}	2.0×10^{-5}
Temporary Place Stable	6.9×10^{-02}	1.3×10^{-4}	5.5×10^{-4}

Table 12: p-values for 1-tailed, unpaired, same-variance t-test for select comparisons to baseline for Test-1. For values less than 0.05 (shown in green), we accept the hypothesis that the treatment (2nd method in column heading) resulted in an improvement over the baseline (1st method in column heading).

2.2 SP Network architecture comparison

Here, we compare the PCN network architecture for grasp/place success prediction to the PointNetGPD network architecture [3]. For this comparison, the criteria for grasp/place success are the same for both methods, and the exact same datasets are used for training and evaluation. (Note the original version of PointNetGPD uses a different grasp quality metric [3].) The learning rate was optimized separately for all contingencies. Results for grasp and place success prediction accuracy and precision is shown in Table 13. Performance on the canonical task is shown in Tables 14 and 15.

	Accuracy (Train)	Accuracy (Test-1)	Precision (Test-1)	Accuracy (Test-2)	Precision (Test-2)
PCN	**0.8522**	**0.7735**	**0.9718**	0.5911	0.6530
PointNetGPD	0.8441	0.7674	0.9401	**0.5989**	**0.6938**
PCN	0.8471	**0.7790**	**0.9358**	**0.6536**	**0.8526**
PointNetGPD	**0.8572**	0.7398	0.9010	0.6392	0.8243

Table 13: **Top**. Grasp success prediction. **Bottom**. Place success prediction. For Test-1, precision was evaluated at threshold 0.95, i.e., the probability the grasp/place was labeled as positive was 0.95. For Test-2, the threshold was 0.90.

We see that both methods perform similarly on Test-1 while PointNetGPD predicts grasp success better for Test-2. This may be because PointNetGPD has fewer parameters and thus generalizes better.

	SP (PCN)	SP (PointNetGPD)
Place Execution Success	0.913 ± 0.006	**0.924** ± 0.006
Grasp Antipodal	0.951 ± 0.005	**0.967** ± 0.004
Temporary Place Stable	**0.967** ± 0.012	0.951 ± 0.014
Plan Length	**2.239** ± 0.015	2.275 ± 0.017
Regrasp planning time (s)	**61.54** ± 0.900	83.14 ± 0.519

Table 14: Canonical performance for Test-1 over 2,000 episodes. Showing average ± standard error.

	SP (PCN)	SP (PointNetGPD)
Place Execution Success	0.591 ± 0.011	**0.643** ± 0.011
Grasp Antipodal	0.674 ± 0.010	**0.731** ± 0.010
Temporary Place Stable	**0.742** ± 0.027	0.707 ± 0.026
Plan Length	**2.518** ± 0.023	2.560 ± 0.024
Regrasp planning time (s)	74.082 ± 0.732	**73.18** ± 0.689

Table 15: Canonical performance for Test-2 over 2,000 episodes. Showing average ± standard error.

3 Bottle Arrangement

The task is to place two bottles upright and stably onto two coasters, as in our prior work [2]. Bottles were selected from ShapeNET [1], 318 of which were for training and 100 were for testing/evaluation. Bottles were scaled uniformly between 9 and 22 cm height, and bottles too wide for the gripper were discarded.

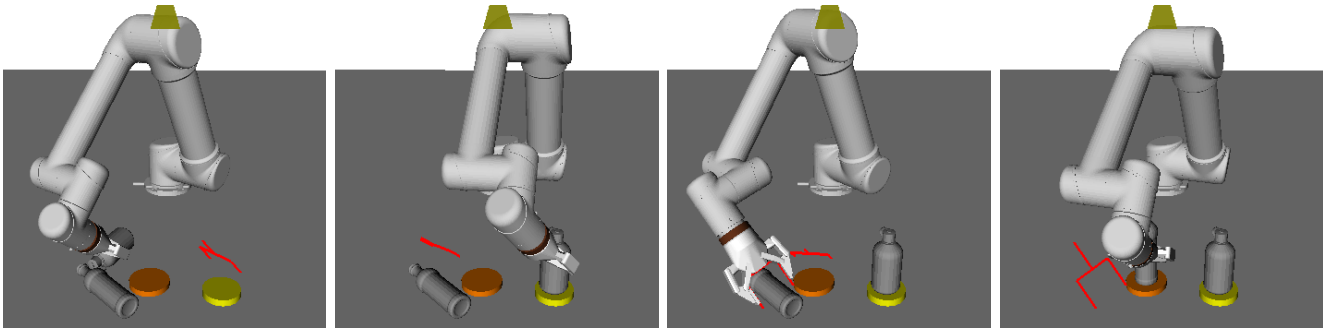


Figure 3: Example sequence of picks and places for one episode of the bottles task.

Figure 4: Bottles used for real robot experiments.

3.1 Perception ablation study

	GT Seg. & Comp.	GT Seg. (Train)	GT Seg. (Test)	Percep. (Train)	Percep. (Test)	GT Seg. & No Comp.
Place Execution Success	0.970 ± 0.005	0.868 ± 0.011	0.860 ± 0.011	0.853 ± 0.011	0.826 ± 0.012	0.234 ± 0.053
Regrasp Plan Found	0.996 ± 0.002	0.998 ± 0.001	0.999 ± 0.001	0.988 ± 0.003	0.984 ± 0.004	0.064 ± 0.008
Grasp Antipodal	0.996 ± 0.002	0.929 ± 0.008	0.914 ± 0.009	0.940 ± 0.008	0.904 ± 0.010	0.339 ± 0.062
Temporary Place Stable	1.000 ± 0.000	1.000 ± 0.000	1.000 ± 0.000	1.000 ± 0.000	1.000 ± 0.000	–
Plan Length	2.002 ± 0.002	2.010 ± 0.004	2.004 ± 0.003	2.006 ± 0.004	2.004 ± 0.003	2.031 ± 0.031
Regrasp planning time (s)	4.486 ± 0.100	1.680 ± 0.083	1.596 ± 0.065	1.598 ± 0.070	1.550 ± 0.066	1.144 ± 0.059

Table 16: Bottle arrangement performance over 500 episodes. Showing average ± standard error.

3.2 Regrasp cost comparison

	No Cost	Step Cost	GQ	MC	CU	SP
Place Execution Success	0.831 ± 0.012	0.824 ± 0.012	0.860 ± 0.011	0.867 ± 0.011	0.820 ± 0.012	**0.877** ± 0.011
Grasp Antipodal	0.910 ± 0.009	0.903 ± 0.010	0.960 ± 0.006	0.958 ± 0.006	0.896 ± 0.010	**0.966** ± 0.006
Temporary Place Stable	0.889 ± 0.043	–	0.926 ± 0.051	**1.000** ± 0.000	–	0.972 ± 0.028
Plan Length	2.122 ± 0.016	**2.000** ± 0.000	2.061 ± 0.011	2.024 ± 0.007	**2.000** ± 0.000	2.073 ± 0.012
Regrasp planning time (s)	**1.378** ± 0.032	1.442 ± 0.045	31.98 ± 0.181	33.33 ± 0.458	30.55 ± 0.253	32.97 ± 0.259

Table 17: Bottle arrangement performance over 500 episodes. Showing average ± standard error.

	No Cost, SP	Step Cost, SP	GQ, SP
Place Execution Success	2.0×10^{-3}	5.3×10^{-4}	1.9×10^{-1}
Grasp Antipodal	9.7×10^{-8}	7.2×10^{-9}	2.5×10^{-1}
Temporary Place Stable	7.6×10^{-2}	–	2.0×10^{-1}

Table 18: p-values for 1-tailed, unpaired, same-variance t-test for select comparisons to baseline. For values less than 0.05 (shown in green), we accept the hypothesis that the treatment (2nd method in column heading) resulted in an improvement over the baseline (1st method in column heading).

3.3 Real robot experiments

	Shape Completion	HSA [2]
Number of Objects Placed	**1.800** ± 0.074	1.667 ± 0.088
Task Success Rate	**0.800** ± 0.074	0.667 ± 0.088
Grasp Success Rate	0.948 ± 0.029	**0.983** ± 0.017
Place Success Rate	**1.000** ± 0.000	0.900 ± 0.040

Table 19: Bottle arrangement performance with shape completion with the "Step Cost" versus hierarchical spatial attention (HSA) [2]. Showing average ± standard error over 30 episodes.

4 Block Arrangement

The robot is to arrange five rectangular blocks from tallest to shortest according to the longest edge. Edge lengths of blocks were scaled uniformly at random between 2 and 7 cm. 5,000 blocks were generated for training, and 1,000 blocks were generated for testing/evaluation.

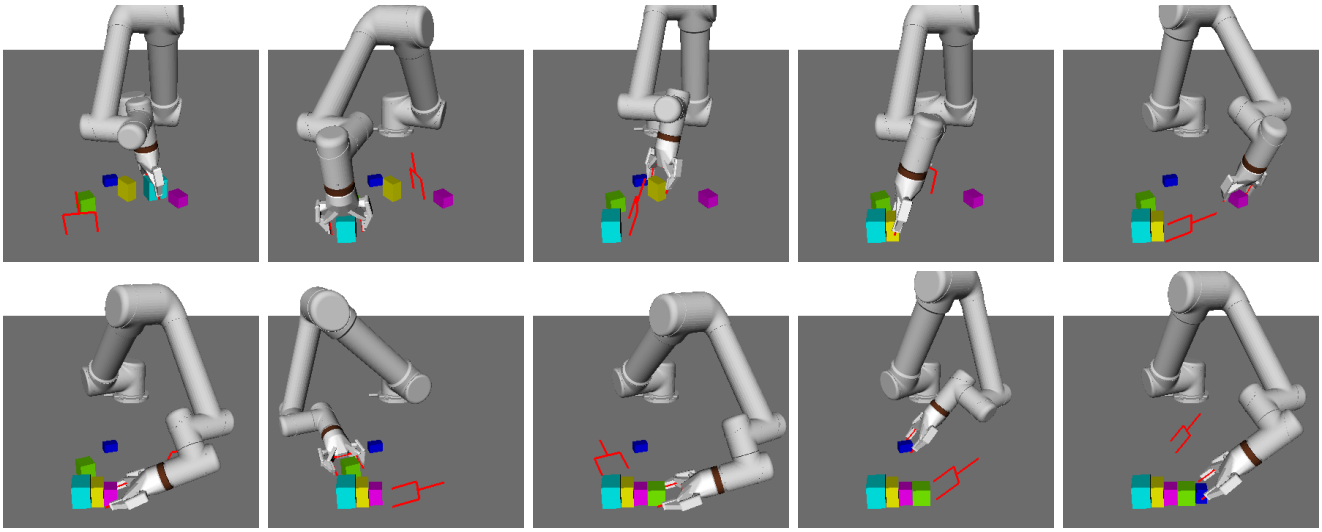


Figure 5: Example sequence of picks and places for the block arrangement task.

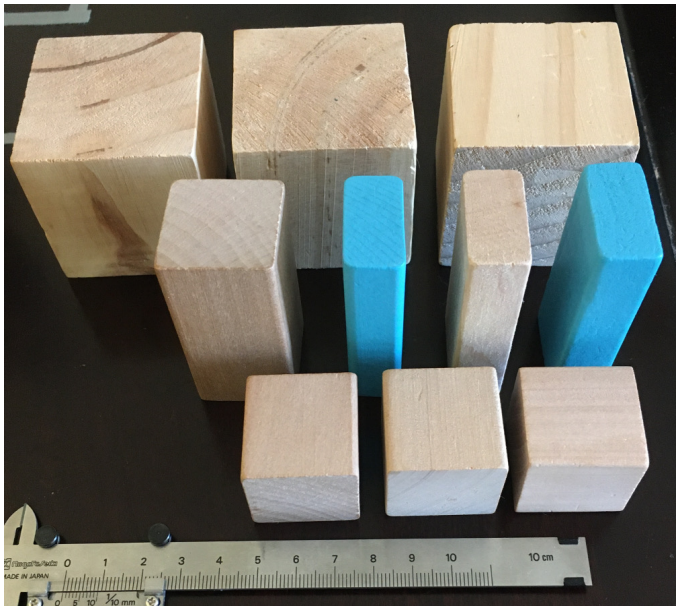


Figure 6: Objects used for block arrangement on a real robot.

4.1 Perception ablation study

	GT Seg. and Comp.	GT Seg. (Train)	GT Seg. (Test)	Percep. (Train)	Percep. (Test)	GT Seg. & No Comp.
Place Execution Success	1.000 ± 0.000	0.940 ± 0.008	0.940 ± 0.008	0.934 ± 0.008	0.936 ± 0.008	–
Plan Found	1.000 ± 0.000	1.000 ± 0.000	1.000 ± 0.000	1.000 ± 0.000	1.000 ± 0.000	0.000 ± 0.000
Grasp Antipodal	1.000 ± 0.000	0.943 ± 0.007	0.950 ± 0.007	0.951 ± 0.007	0.952 ± 0.007	–
Temporary Place Stable	1.000 ± 0.000	0.857 ± 0.143	1.000 ± 0.000	1.000 ± 0.000	1.000 ± 0.000	–
Plan Length	2.048 ± 0.010	2.014 ± 0.005	2.026 ± 0.007	2.020 ± 0.006	2.020 ± 0.006	–
Order Correct	0.911 ± 0.009	0.771 ± 0.014	0.752 ± 0.014	0.758 ± 0.014	0.743 ± 0.014	–
Longest End Up	0.969 ± 0.005	0.757 ± 0.014	0.736 ± 0.014	0.741 ± 0.014	0.732 ± 0.014	–
Regrasp planning time (s)	3.520 ± 0.126	2.797 ± 0.072	2.947 ± 0.087	2.696 ± 0.073	2.890 ± 0.082	0.709 ± 0.002

Table 20: Block arrangement performance over 200 episodes. Showing average ± standard error.

4.2 Regrasp cost comparison

	No Cost	Step Cost	GQ	MC	CU	SP
Place Execution Success	0.888 ± 0.010	0.917 ± 0.009	0.982 ± 0.004	0.969 ± 0.005	0.916 ± 0.009	**0.989** ± 0.003
Grasp Antipodal	0.950 ± 0.005	0.939 ± 0.008	0.999 ± 0.001	0.989 ± 0.003	0.934 ± 0.008	**1.000** ± 0.000
Temporary Place Stable	0.971 ± 0.006	**1.000** ± 0.000	**1.000** ± 0.000	**1.000** ± 0.000	**1.000** ± 0.000	**1.000** ± 0.000
Plan Length	3.854 ± 0.042	**2.004** ± 0.003	2.002 ± 0.002	2.180 ± 0.018	2.022 ± 0.007	2.008 ± 0.004
Regrasp planning time (s)	**1.624** ± 0.015	2.420 ± 0.038	26.88 ± 0.141	41.40 ± 0.303	23.10 ± 0.189	10.62 ± 0.325

Table 21: Block arrangement performance over 200 episodes. Showing average ± standard error.

	No Cost, SP	Step Cost, SP	GQ, SP
Place Execution Success	1.0×10^{-21}	9.4×10^{-15}	9.5×10^{-2}
Grasp Antipodal	3.9×10^{-13}	7.2×10^{-16}	1.6×10^{-1}
Temporary Place Stable	3.7×10^{-01}	–	–

Table 22: p-values for 1-tailed, unpaired, same-variance t-test for select comparisons to baseline. For values less than 0.05 (shown in green), we accept the hypothesis that the treatment (2nd method in column heading) resulted in an improvement over the baseline (1st method in column heading).

4.3 Real robot experiments

Number of Objects Placed	4.900 ± 0.100
Grasp Success Rate	0.981 ± 0.019
Plan Length	2.080 ± 0.085
Order Correct	0.980 ± 0.020
Longest End Up	0.959 ± 0.029

Table 23: Block arrangement performance using the step cost over 10 episodes. Showing average ± standard error.

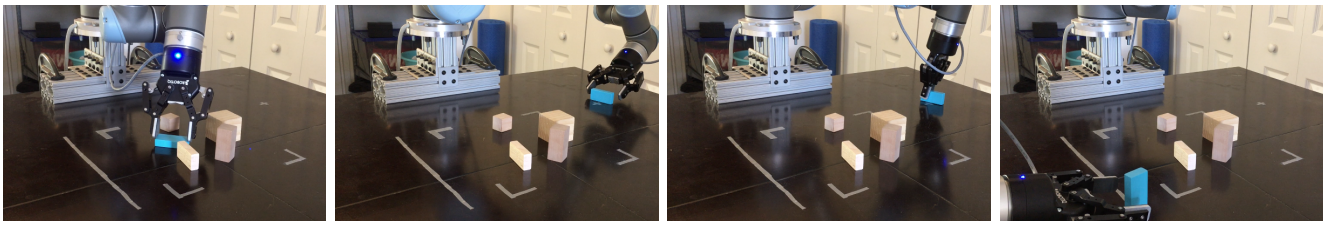


Figure 7: Example regrasp sequence with blocks.

References

[1] Angel Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, Jianxiong Xiao, Li Yi, and Fisher Yu. ShapeNet: An information-rich 3D model repository. Technical Report arXiv:1512.03012 [cs.GR], Stanford University — Princeton University — Toyota Technological Institute at Chicago, 2015.

[2] Marcus Gualtieri and Robert Platt. Learning manipulation skills via hierarchical spatial attention. *IEEE Transactions on Robotics*, 36(4):1067–1078, 2020.

[3] Hongzhuo Liang, Xiaojian Ma, Shuang Li, Michael Görner, Song Tang, Bin Fang, Fuchun Sun, and Jianwei Zhang. PointNetGPD: Detecting grasp configurations from point sets. In *2019 Int'l Conf. on Robotics and Automation*, pages 3629–3635. IEEE, 2019.

[4] Walter Wohlkinger, Aitor Aldoma, Radu Rusu, and Markus Vincze. 3DNet: Large-scale object class recognition from CAD models. In *IEEE Int'l Conf. on Robotics and Automation*, pages 5384–5391, 2012.