Learning Bimanual Scooping Policies for Food Acquisition Supplementary Material

This appendix contains food property details, classifier training details, experiment rollout visualizations, and a more in-depth analysis of failure modes. For more videos, please see our website.

A Food Property Analysis

The detailed food properties of 14 different types of food are presented in Table 3. Within the robust food category, we consider variations along brittleness and geometry. We also consider both brittle foods, e.g., cashews, and compliant foods, e.g., pasta. In addition, we experiment with a wide range of shapes, e.g., round grapes and irregularly shaped farfalle pasta. In our list, we also include foods with varied sizes, from extra small, e.g., peas, to larger items, e.g., strawberries, and we vary the thickness of food, e.g., thin snow peas vs thick broccoli.

For fragile food, we use firm tofu, cheesecakes, and jello pieces of different colors, shapes and stiffness in the experiments. Unlike the tofu and jello pieces, cheesecakes have an additional property of stickiness that makes it extremely difficult to be scooped without any residue in the environment. For jello pieces, we have two colors and shapes of red square and orange triangle. The latter one is less stiff. One reason for the total failure of Single baseline on orange triangle jello is that the triangle is not symmetric along certain axis. Therefore, Single baseline is able to cup part of the orange triangle jello into the spoon mouth but it falls out of the spoon during the Scooping Phase because of the asymmetric weight imbalance.

Table 3: **Food Property Analysis:** We report the food properties of 14 different food classes in terms of deformability, brittleness and geometry, including shape, size and thickness. In the column of size, **L** represents large food, **M** represents medium food, **S** represents small food and **XS** represents extra small.

Food Type	**Deformability**	**Brittleness**	**Geometry**		
			Shape	Size	Thickness
Broccoli	Robust	Compliant	Irregular	L	Thick
Grapes	Robust	Compliant	Round	M	Thick
Blueberry	Robust	Compliant	Round	S	Thick
Strawberry	Robust	Compliant	Irregular	L	Thick
Carrot	Robust	Compliant	Cylinder	M	Thick
Farfalle	Robust	Compliant	Irregular	M	Thin
Macaroni	Robust	Compliant	Irregular	S	Thick
Snow Pea	Robust	Compliant	Irregular	L	Thin
Cashews	Robust	Brittle	Irregular	S	Thick
Goldfish	Robust	Brittle	Irregular	S	Thick
Peas	Robust	Compliant	Round	XS	Thick
Tofu	Fragile	Compliant	Square	L	Thick
Cheesecake	Fragile	Compliant	Square	L	Thick
Orange Triangle Jello	Fragile	Compliant	Triangle	L	Thick
Red Square Jello	Fragile	Compliant	Square	L	Thick

B Training Details

Risk Classifier. We instantiate the Risk Classifier with a ResNet34 architecture trained on a hand-labelled dataset of 14 food classes labelled "Robust" or "Fragile". The dataset is composed of 600 overhead RGB images of all foods in Table 3 except Orange Triangle Jello and augmented by 8X by applying a series of standard label-preserving image transformations, including rotation, flipping, blurring, affine transformation, contrast changes, hue and saturation changes and addition of Gaussian noise. The applied augmentations not only enlarge the dataset but also enable Risk Classifier to work under various lighting conditions and be robust to small camera shift. With enough augmented data, we use Binary Cross-Entropy Loss and the Adam Optimizer [19] with a learning rate of 1e-4 and a weight decay of 1e-4 to train the Risk Classifier. We train on a NVIDIA GeForce GTX 1070 GPU for 15 epochs.

Figure 6: **Augmented Dataset Images**: We present examples of augmented overhead food images from the dataset used to train the Risk Classifier. The datasets used to train the Segmentation and Failure Classifier models were augmented with the same techniques.

Failure Classifier. Similar to Risk Classifier, we also use the ResNet34 [14] architecture to train a Failure Classifier to identify the breakage-imminent states during the pushing phase. The model is trained on the dataset of 30 rollouts, each containing 60 image frames, of the Pushing Phase of tofu with $\alpha = 1$ for maximum pushing distance. We augment the dataset by 8X with the same augmentations as with the Risk Classifier. We train the Failure Classifier with Binary Cross-Entropy Loss and the Adam optimizer [19] with a learning rate of 1e-4 and a weight decay of 1e-4. We train on a NVIDIA GeForce GTX 1070 GPU for 25 epochs. We find that the Failure Classifier is able to generalize to the three fragile food classes unseen during training with a success rate of 96.8% of 157 images over 13 scooping rollouts.

Augmentation	**Parameters**
LinearContrast	(0.95, 1.05)
Add	(-10, 10)
GammaContrast	(0.95, 1.05)
GaussianBlur	(0.0, 0.6)
MultiplySaturation	(0.95, 1.05)
AdditiveGaussianNoise	(0, 3.1875)
Flipud	0.5

Table 4: **Data Augmentation Parameters:** We report the augmentation techniques used to train all models in CARBS, along with their accompanying parameter values. All augmentations are used from the imgaug library [15].

C Hardware Design

For our bimanual scooping task, we find that the design of the pusher and the scooper greatly improved the efficiency and effectiveness of the bimanual primitive.

Inspired by the antique pushers used by children to push food onto the spoon, we 3D print a custom concave pusher that has approximately the same curvature as the mouth of the spoon. During the Pushing Phase, the concave surface of the pusher pushes the off-centered food items to the mouth of the spoon and groups multiple food items together towards the center while a flat pusher may cause potential spreading of food over the plate. Another critical design choice is the pusher's size. It approximately the same size of the spoon so that the food items that are already grouped to the center of the pusher with the pusher's concavity. Foods are cupped into the spoon mouth without leaving anything beyond the reachable range of the spoon.

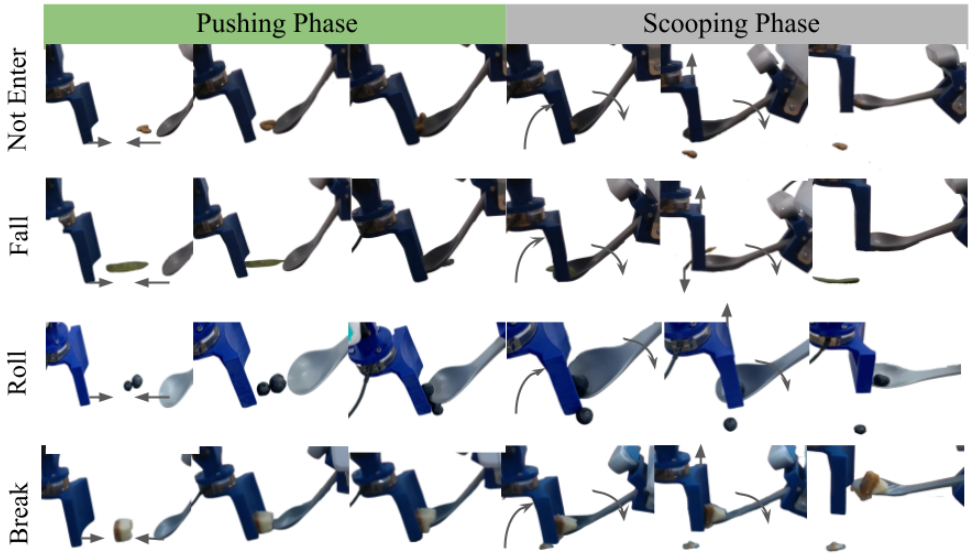

Figure 7: **Failure Mode Rollouts**: We present selected rollouts illustrating each of the four failure modes. The Not Enter failure shows a cashew becoming wedged between the pusher and scooper and then failing to enter the spoon mouth. The Fall failure shows the snow pea entering the spoon bowl, but falling out of the spoon off the side due to its irregular geometry. The Roll failure shows two blueberries rolling off each other and out of the trajectory of the scooper. Lastly, the final row shows a cheesecake piece breaking, leaving a piece of food on the plate after scooping.

Apart from the design of the pusher, we also experiment with different designs of the scooper spoon tool and mount. The two key components of the scooper design are the mounted angle of the spoon and the tilted angle of the camera. The spoon is angled 45 degrees off the vertical axis, to align the mouth of the spoon to be tangent to the plate workspace and avoid robot motion range constraints. A larger angle would make the front part of the spoon mouth too high, which increases the difficulty of scooping thin foods, e.g., snow peas, and extra small foods, e.g., peas, because they can slide under the spoon mouth. A smaller angle is prone to cause conflicts and break constraints of two arms during the Pushing and Scooping phase. The mounted camera is mounted at 30 degrees to capture the full view of the spoon mouth, the food in front of the spoon and the pusher during the Pushing Phase, which is necessary for the Failure Classifier.

D Failure Mode Analysis

We observe four failure modes during experiments with all three bimanual scooping strategies: Not Enter, Roll, Fall, and Break. We present visualizations of these failure modes in the Fig. 8.

Not Enter: As shown in Fig. 5, the failure mode Not Enter is the most common failure mode for the Single baseline. The food is in contact with the spoon during the Pushing Phase but it fails to be cupped inside the spoon mouth in the end. There are various failure cases in this failure modes, e.g., the cashews are pushed with too much force so it jumps out of the space between the pusher and the spoon. Another failure case of this failure mode in the experiments is that some small or thin food items are stuck between the spoon and the pusher, not able to roll into the spoon mouth.

Roll: Roll is a failure mode of round foods that are easy to roll in the environment. During the Pushing Phase, either the pusher or the spoon exerts a force on the food item. The food builds momentum and may roll in the environment. If the round foods are in contact with each other, they are also likely to roll against each other and roll out of the scooping trajectory. Therefore, the scooper will fail to pick up the foods.

Fall: After the Pushing Phase, only part of the food is cupped into the spoon mouth. Therefore, when the spoon rotates in the Scooping Phase, the food may fall out of the spoon because of its unstable position. In general, this type of failure mode is more common in Single baseline than

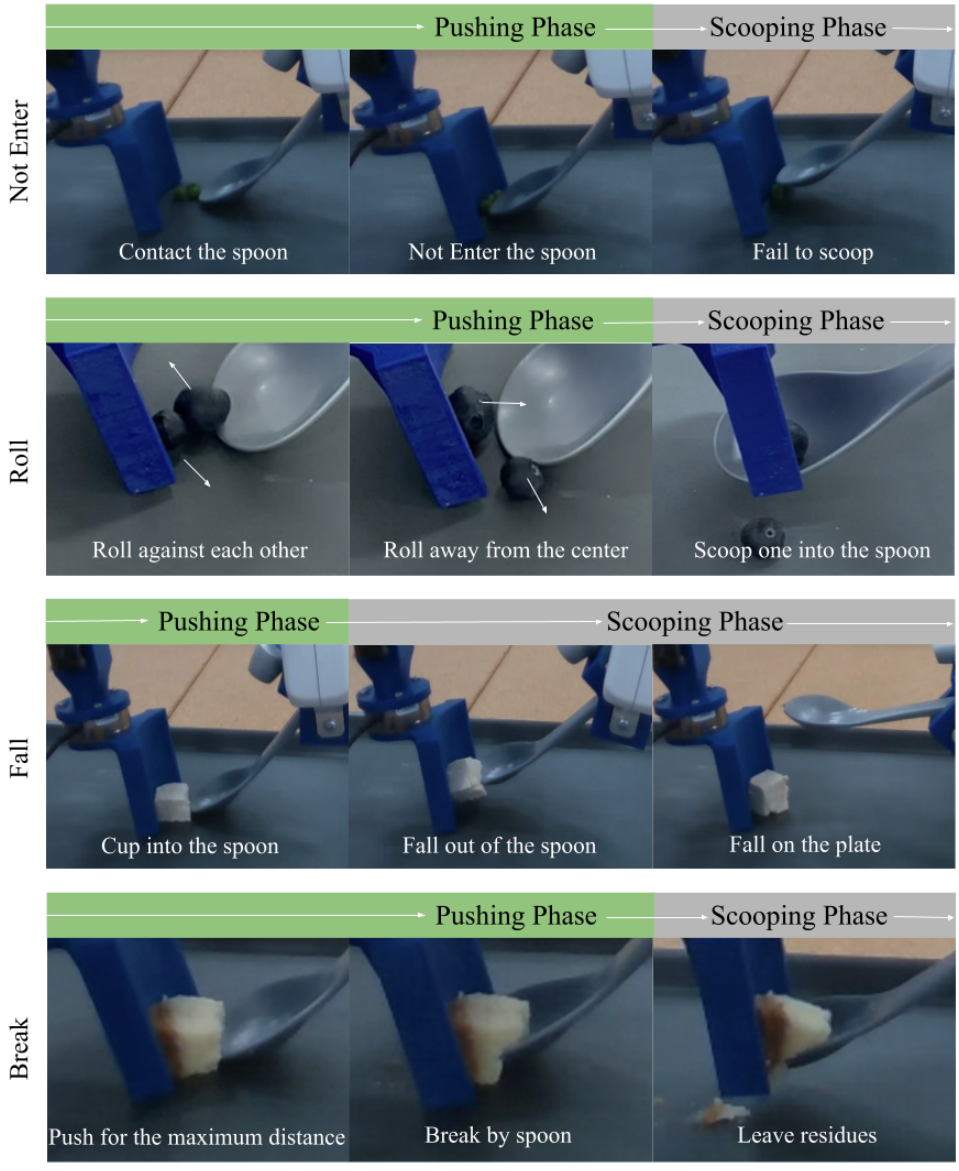

Figure 8: **Failure Mode Visualizations:**: We present four failure modes: Not Enter, Roll, Fall and Break. Break failures are only present with deformable foods, such as cheesecake shown here. For the robust food category, Not Enter is the most common failure mode.

other methods because the pusher in Single baseline acts like a static barrier without following the rotating motion of the spoon and fails to stabilize the food during the process.

Break: The failure mode of Break only happens to fragile foods because this category of foods are prone to deform and break while being squeezed under large forces from the spoon or the pusher.

E Additional Rollouts

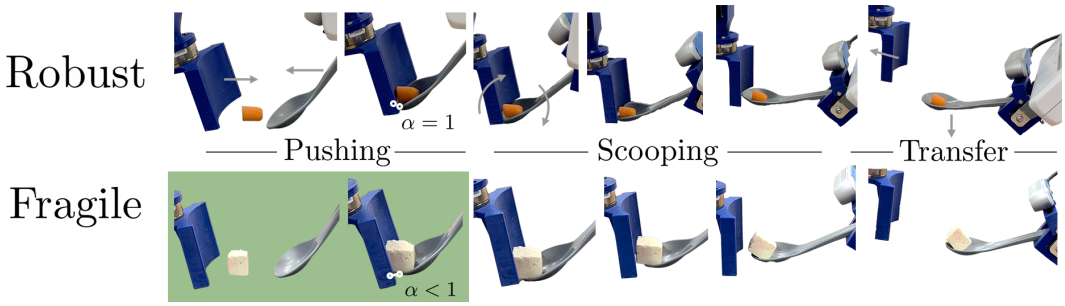

Figure 9: **Full Scooping Rollouts**: Full rollout visualizations for scooping a robust food (carrot) and a fragile food (tofu) with adaptive pushing distance α. Both foods employ the three phase bimanual scooping primitive, but differ in α choice. The carrot is a rigid food and uses and α for maximum stabilization during scooping, while the tofu is fragile and requires an $\alpha < 1$ to avoid breakage.

F Ablation Study of Stablizing Strategies

Our approach CARBS includes three stabilizing strategies, Angled Pushing, Adaptive Cupping and Pinning, to prevent potential failures while scooping food of various properties. To understand how these three stabilizing strategies impact performance, we compare CARBS to three ablation baselines, CARBS without Angled Pushing, CARBS without Adaptive Cupping and CARBS without Pinning, on five robust food classes. In CARBS without Angled Pushing, the angle of the pusher θ is set to 0 during the Pushing Phase. In CARBS without Adaptive Cupping, the angled pusher and scooper move in sequence rather than simultaneously so that the pusher and the spoon cannot cup the food together. Lastly in CARBS without Pinning, the pusher stays on the table during the Scooping phase rather than following the scooper up to pin the foods in the spoon.

For the ablation experiments, we consider 5 different types of foods: grape, blueberry, macaroni, snow pea and cashews. These 5 classes cover a variety of visual and physical properties, including brittleness and geometry. We consider cashews to represent brittle foods and macaroni for compliant foods. In terms of geometry, we include varied sizes from small to large, e.g., blueberry, grape and snow pea, and a wide range of shapes, e.g., round grape and irregularly shaped snow pea. Table 5 reports the success rate of scooping grape, blueberry, macaroni, snow pea, and cashews with ablations of CARBS's stabilizing strategies.

Based on the success rate, CARBS without Angled Pushing has worse performance on cashews because two cashews become wedged between the vertical pusher and scooper. The Angled Pushing stabilizing strategy was designed with this failure mode in mind, and encourages food items to roll over the lip of the spoon and into the spoon bowl. CARBS without Adaptive Cupping does not match the performance of our approach on 4 out of 5 food items, showing that Adaptive Cupping is critical for most food items. This is because in the Pushing phase, Adaptive Cupping cups the food to be centered with the spoon mouth and prevents the food rolling away. Lastly, CARBS without Pinning also performs worse than CARBS on all five food classes because when the pusher does not follow the scooper motion during the Scooping phase, food will easily fall out of the spoon bowl.

G Additional Experiments for Out of Distribution Food Items

In order to further test the generalization of our method, we extend our method to other unseen, scattered food items that also require a scooping mechanism in our daily life. The food tested in the Table 6 are cooked rice and couscous and they are grouped into a scoopable area on the workspace before scooping. Results in Table 6 suggests that CARBS achieves comparable performance for

Food Type	Success Rate			
	w/o *Angled Pushing*	w/o *Adaptive Cupping*	w/o *Pinning*	*CARBS*
Grape	**5/5**	3/5	3/5	**5/5**
Blueberry	**5/5**	3/5	4/5	**5/5**
Macaroni	**5/5**	4/5	4/5	**5/5**
Snow Pea	**4/5**	**4/5**	3/5	**4/5**
Cashews (2)	5/10	6/10	6/10	**7/10**

Table 5: **Ablation Study Results:** We report the per food item success rate over 5 trials of scooping robust foods with CARBS strategies and three ablation baselines: CARBS without Angled Pushing, CARBS without Adaptive Cupping, and CARBS without Pinning. As expected, we observe CARBS achieves best overall performance across all 5 food classes in the table. CARBS without Angled Pushing fails on cashews due to their propensity to become wedged between the vertical pusher and scooper, while CARBS without Adaptive Cupping only matches CARBS performance on one out of five food items. CARBS without Pinning performs worse than CARBS on all five food items. These results suggest that the combination of all three bimanual stabilizing strategies (Angled Pusher, Cupping Motion, and Pinning Motion) are indispensable to the generalization and robustness of our method for scooping various food items.

cooked rice and couscous with human baseline. Both *Human* and CARBS have larger weight loss in couscous than rice, indicating that it is more difficult to scoop couscous because they are much smaller and more easily pushed out of the pushing and scooping path. We do not consider any grouping strategies for scattered foods, and leave this problem as an interesting direction for future work.

Food Type	Avg. Weight Difference (%)	
	CARBS	*Human*
Rice (CARBS OOD)	10.339	**8.423**
Couscous (CARBS OOD)	31.666	**21.594**

Table 6: **OOD Food Results**: We report the weight loss of food items after scooping as a percentage of the original food weight, averaged across 5 scooping trials. We compare our method CARBS to human baselines over 2 unseen scattered food rice and couscous. We observe that both CARBS and *Human* could scoop a majority amount of scattered food on the plate.

H Force Sensor Analysis

Apart from the visual servoing system for bimanual scooping, we also consider using a force torque sensor mounted to the pusher. We conduct experiments to test if force sensor readings give additional information about the deformability of the food to distinguish between robust and fragile foods, as well as whether a fragile food is in a breakage-imminent state. We present the force readings along three axes during the three phases: Pushing, Scooping and Transfer, for three food items, including grape, strawberry and tofu.

Figure 10 demonstrates that there are no distinct differences of forces between robust and fragile foods, suggesting that tactile information alone would not be sufficient for classifying food fragility. We also present the forces during two rollouts of scooping tofu. One of these rollouts has breakage at the transition between the Pushing and Scooping phases, while the other has no breakage throughout. The difference of forces in Figure 11 is also too small to identify which one has the breakage, and when the breakage occurred.

In conclusion, we find that the force-torque readings alone are too noisy to be used to replace the Risk Classifier and distinguish between robust and fragile food items. Even within fragile food scooping rollouts, we are unable to identify breakage failures from force-torque sensing alone, suggesting the need for a vision-based system for our Failure Classifier. Although past food acquisition works have used tactile sensing [10], we hypothesize the multi-object interactions present in bimanual scooping, such as the pusher scraping against the plate, adds too much noise to the force readings to be used to identify food states.

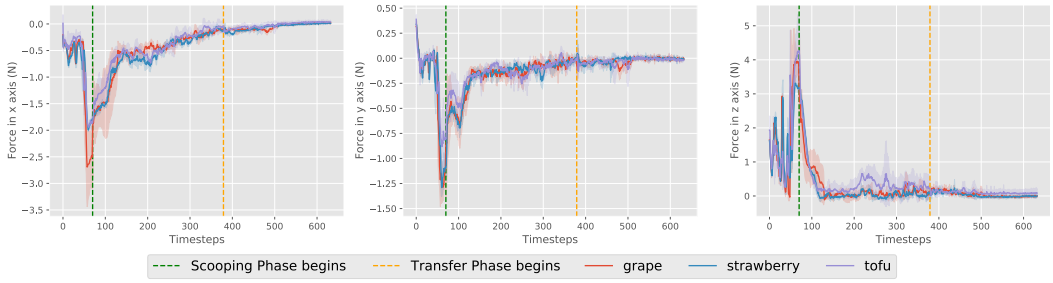

Figure 10: **Sensored forces of 3 food items**: We report the per food item sensored force in the pusher over 3 trials of scooping foods along the x, y and z axes in three phases(Pushing, Scooping, Transfer). The confidence interval in the plot is 95% and there is no obvious difference between the grape, strawberry and tofu.

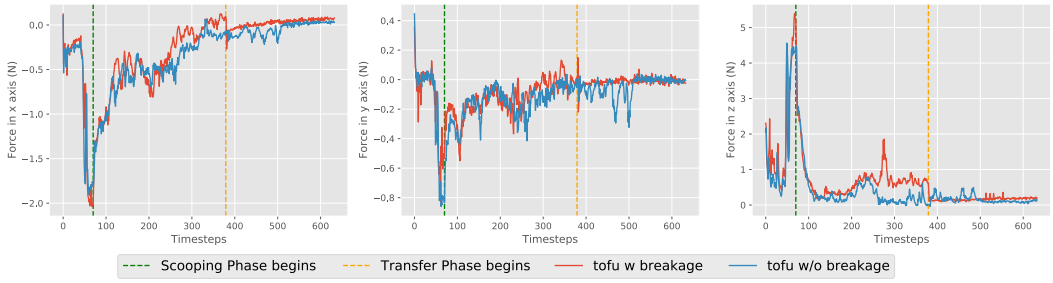

Figure 11: **Sensored forces from tofu rollouts**: We report the sensored force in the pusher along the x, y and z axes in three phases (Pushing, Scooping, Transfer) for tofu rollouts. The breakage happens around the timestep of the green line where the transition from pushing to scooping starts. We cannot identify the breakage failures purely from the force torque sensing.