Transfer depends on Acquisition: Analyzing Manipulation Strategies for Robotic Feeding

Daniel Gallenberger*, Tapomayukh Bhattacharjee†, Youngsun Kim‡ and Siddhartha S. Srinivasa§
Paul G. Allen School of Computer Science and Engineering
University of Washington
Seattle, USA
Email: *danielga@cs.washington.edu, †tapo@cs.washington.edu, ‡yskim@cs.washington.edu, §siddh@cs.washington.edu

***Abstract*—Successful robotic assistive feeding depends on reliable bite acquisition and easy bite transfer. The latter constitutes a unique type of robot-human handover where the human needs to use the mouth. This places a high burden on the robot to make the transfer easy. We believe that the ease of transfer not only depends on the transfer action but also is tightly coupled with the way a food item was acquired in the first place. To determine the factors influencing good bite transfer, we designed both skewering and transfer primitives and developed a robotic feeding system that uses these manipulation primitives to feed people autonomously. First, we determined the primitives' success rates for bite acquisition with robot experiments. Next, we conducted user studies to evaluate the ease of bite transfer for different combinations of skewering and transfer primitives. Our results show that an intelligent food item dependent skewering strategy improves the bite acquisition success rate and that the choice of skewering location and the fork orientation affects the ease of bite transfer significantly.**

***Index Terms*—assistive feeding; deformable object manipulation; bite acquisition; bite transfer**

I. Introduction

Eating is an activity of daily living (ADL) and losing the ability to self-feed can be devastating [1]. According to a survey in 2010, around 1.0 million adults in the United States required the assistance of another person to help them eat [2]. Conditions such as cerebrovascular diseases like strokes [3], Parkinson's, arthritis, multiple sclerosis [4], spinal cord injuries [5], bilateral amputations, and many others can render individuals unable to eat on their own accord. Instead, they depend on a caregiver to feed them every morsel of every meal every day [6]. In addition to positively impacting the self-worth of people with disabilities [7], [8], independent dining would have a considerable effect on caregiver hours because feeding is one of the most time-consuming tasks for caregivers [9], [10]. Also, dining together with other people is a cornerstone of society and provides a personal link to the wider community [11]–[13] but the presence of caregivers during dinner with friends or relatives may pose a privacy concern [14].

Some commercial powered feeding systems currently on the market [15]–[18] address this problem using a robotic arm that scoops food with a spoon. These systems have lacked widespread acceptance probably because of limited mobility and minimal autonomy demanding a time-consuming food preparation process in specialized food containers [15], [16], [19]. A general-purpose robotic system attached to a wheelchair with increased mobility can perform in situ feeding tasks in addition to general tasks such as opening doors and picking up items. Feeding is challenging in because it involves complex bite acquisition strategies for food with a variety of physical characteristics. Bite transfer is also challenging because the food needs to be positioned and oriented relative to the mouth in a way conducive to easy bite transfer. We address this challenge by employing our key insight that depending on the physical properties of a food item, the manipulation strategies for easy bite transfer may be dependent on the strategies for reliable bite acquisition.

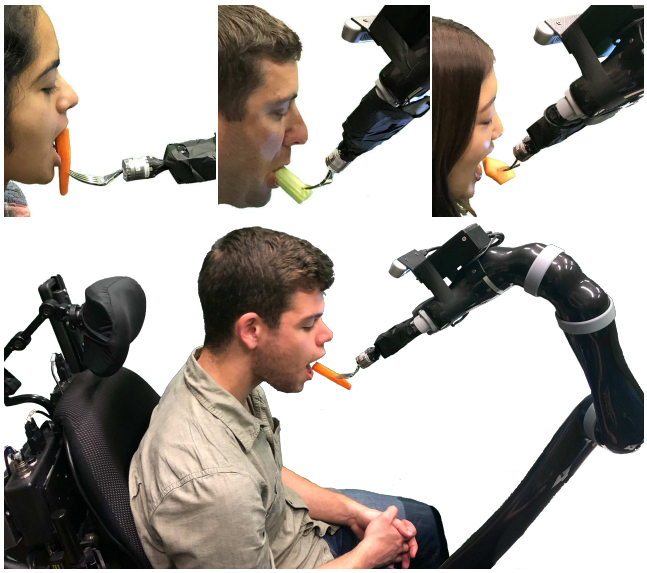
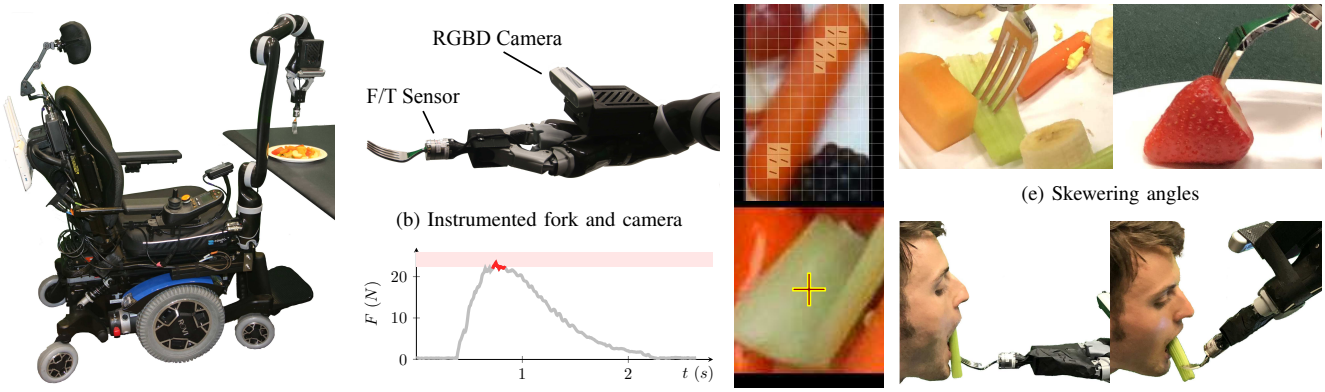


Fig. 1: Robotic feeding using various manipulation strategies.

Using a fixed strategy for bite acquisition is not ideal in realistic situations because food items come in various shapes, and have physical properties that are difficult to model. Eggs, for example, may require skewering at a position where the white and the yolk are simultaneously skewered to prevent any one part from falling off [20]. Bananas may require angled skewering approach angles to prevent them from slipping due to gravity when lifting off. Our robotic system uses fine manipulation planning and leverages the complementary capabilities

This work was funded by the National Institute of Health R01 (#R01EB019335), National Science Foundation CPS (#1544797), National Science Foundation NRI (#1637748), the Office of Naval Research, the RCTA, Amazon, and Honda.



(a) Our autonomous feeding system (b) Instrumented fork and camera (c) Skewering forces (d) Positioning (e) Skewering angles (f) Bite transfer angles

Fig. 2: Our feeding system autonomously skewers food items and feeds people with different strategies using multiple sensing modalities. The system uses a food item dependent force threshold (shaded region in Fig.2(c)) to skewer a food item.

of multiple sensing modalities such as vision and haptics to automatically acquire bites for a variety of food items. Our system decides which discrete manipulation primitive to instantiate such as a food item dependent skewering approach angle as well as select its continuous parametrization such as skewering position, rotation and how much force to apply. We analyzed the effect of these variations on bite acquisition success rate.

Bite transfer may also need different food item dependent strategies because an easy bite transfer strategy may require a robot to orient the food item such that a user can take a bite without opening the mouth excessively [20]. One can think of feeding as a handover of food from the fork or spoon to the mouth. However, unlike traditional hand-to-hand transfers, hand-to-mouth transfers pose a greater burden to the robot initiator because of fewer degrees-of-freedom to orient the mouth to receive a bite. Therefore, we developed our system to transfer a bite using different discrete handover primitives and compared these primitives for the ease of bite transfer using studies with 25 human participants.

Our results show that a skewering strategy based on a food item's shape, size and physical properties outperforms the baseline approach of skewering at the center in terms of the bite acquisition success rate, especially for long, slippery, and heterogeneous food items. Our human participant studies show that transfer depends on acquisition. Angled skewering combined with angled transfer performed significantly better than vertical skewering combined with horizontal transfer for easy bite transfer (See Fig.1). Also, people tend to avoid hitting the tines of the fork while biting a long and slender carrot and thus, where a robot skewers an item can affect the ease of bite transfer as well.

Our contributions can be summarized below as:

1) We developed a feeding system that can acquire solid food items and feed a user autonomously using multiple complementary modalities of vision and haptics.
2) We designed various discrete manipulation primitives and their continuous parametrizations for reliable bite acquisition and ease of bite transfer, and empirically compared the different strategies with human participants.
3) We created a new dataset [21] of food items with masks of skewering positions and rotations for effective bite acquisition and bite transfer.

II. Related Work

An autonomous robotic feeding system encompasses the fields of manipulation, perception, and human-robot interaction. There is hardly any work investigating the human-robot interaction aspects of varied manipulation strategies used for autonomous robotic feeding, but here, we present a review of work related to manipulation and perception.

A. Food Manipulation

Though there are few studies on manipulation of solid food items for assistive feeding, related work on solid food manipulation focus on the either packaging or cooking applications [22], [23]. However, food manipulation in the context of assistive feeding is different from food manipulation in other contexts and it has its unique manipulation, perception, and human-robot interaction challenges.

1) Acquisition: Gemici et al. [24] study food manipulation in the context of food preparation for cooking, but their system uses a spatula and a gripper for grasping the food items which require different manipulation strategies than skewering. There is also considerable work on industrial robotic food manipulation but Chua et al. [25] noted that intrusive gripping methods are generally not used in this field because they could potentially damage the food items and force feedback is crucial to manipulate non-rigid food items. Park et al. [26] developed a semi-solid food acquisition system for assistive feeding using a general purpose manipulator to scoop yogurt with a spoon. They also developed an anomaly detection framework for assistive feeding using multiple sensing modalities [27], [28]. The system that comes closest to ours in terms of bite acquisition is the work done by Herlant [19], which also uses a fork to skewer solid food items autonomously. However, our objective to improve bite transfer calls for additional capabilities, such as identifying single food items on the plate and choosing skewering positions and fork rotations with bite transferability in mind.

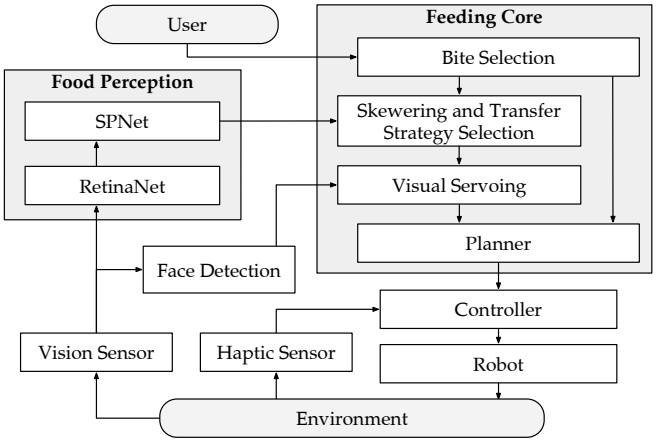


Fig. 3: Our system uses multimodal sensing modalities to sense, perceive, plan, and control the robot to skewer food items and transfer them to a user.

Interestingly, there is also a vast body of literature on grasping that is related to food manipulation. However, unlike most of the work on grasping which focuses on directly manipulating a rigid object, our application demands indirect tool-mediated manipulation of deformable objects. Having said that, finding a good grasping location is conceptually similar to finding good skewering locations. A common approach for finding good grasping locations is using learning based methods to learn good grasping locations on a 3d model and match them to a perceived object, as seen in [29]–[31], or using images as in [32], [33], or from real robot grasping trials [34]. However, most of these approaches may have challenges with deformable objects and may be difficult to use for manipulating food items because the food may be damaged and the physical characteristics may change after one skewering attempt. Additional interesting results in manipulation of difficult items include the research on articulated objects by Katz et al. [35] and the work on learning elasticity parameters by Frank et al. [36].

2) Transfer: Although it is hard to find related studies on analyzing manipulation strategies for robotic bite transfer, similarities can be drawn to the studies on robot-human handovers. Research on robot-human handovers often focuses on rigid objects that are gripped with fingers [37]–[39] which is different from tool-based handovers, where a robot transfers a bite using a fork. The feeding handover situation poses an additional challenge of transferring to a mouth with fewer degrees of freedom. Canal et al. [40] explore bite transfer in the context of a personalization framework. A related, more general idea is found in [41], where the human preference for "default orientations" of objects and the importance of grasp type were identified. Aleotti et al. [42] built on this by orienting items in a way that makes grasping easier and confirmed it with a human user study. In our paper, we want to extend the research domain of robot handovers to the use case of assistive feeding, where items are not grasped and handed to a person's hand, but skewered with a fork and transferred to a human's mouth.

B. Food Perception

Food perception for an autonomous robotic feeding system requires classification and detection of food items on a plate and segmentation for skewering position and rotation masks.

With the latest success in deep supervised learning, image classification and object detection research have been under active development. Researchers have proposed many networks for object classification such as AlexNet [43], GoogLeNet [44], VGG Net [45], ResNet [46], DenseNet [47], Feature Pyramid Network (FPN) [48], and a variety of work on food detection [49]–[51]. There have also been variants that prioritize speed such as MobileNet [52] and SqeezeNet [53]. On top of the evolution of the computing power, comprehensive datasets such as ImageNet [54], PASCAL VOC [55], and COCO [56] datasets continue to drive this research area.

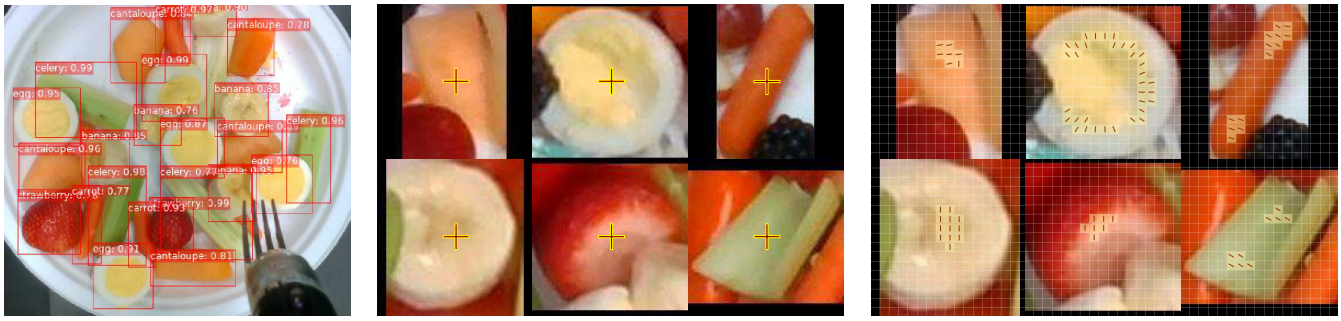
For object detection, algorithms such as Overfeat [57], R-CNN series [58]–[62], Yolo [63], SSD [64], RetinaNet [65], and DetNet [66] have made great strides. Among the state-of-the-art object detectors, we chose RetinaNet for food item detection and recognition, mainly because RetinaNet is faster and lighter than two-stage object detectors such as Faster R-CNN, but more accurate than other single stage networks such as SSD or YOLO [65].

Another area focusing on image segmentation or semantic segmentation infers tight masks for each object in a scene. For image segmentation, researchers developed algorithms such as Vanilla FCN [67], U-Net [68], SegNet [69], and segmentation with two-way DenseNet [70]. Mask RCNN [71] and BlitzNet [72] generate fine masks on top of object detection layers. [71] showed that the independent masks from Mask RCNN with the Fully Convolutional Network (FCN) branch achieved the highest performance. We adopt this idea and implement two sub-branches for our skewering position and rotation masks for bite acquisition, which sits on top of our RetinaNet object detection network.

III. An Autonomous Assistive Feeding System

We developed an autonomous robotic feeding system that uses newly developed sensing, perception, planning, and control modules to acquire a bite from a plate and feed it to a person using various manipulation primitives. Our system consists of a 6 DoF JACO robotic arm [73] mounted on a powered ROVI wheelchair [74] to mimic similar setups used in real homes. The robotic arm has 2 fingers that grab an instrumented fork (forque, see Fig.2(b)) using a custom built 3D printed fork holder. The system uses visual and haptic modalities to perform the feeding task. For haptic feedback, we instrumented the forque with a 6-axis ATI Nano25 Force-Torque sensor [75]. We use haptic sensing to control the end effector forces during skewering and to detect if food acquisition was successful as well as if the fork hits something unexpectedly to improve safety. For visual feedback, we mounted a custom built wireless RGBD camera on the robot's wrist by using the Intel RealSense D415 camera and the Intel Joule 570x for wireless transmission.

We designed the system such that it perceives a food item on a plate using the perception methods described in



(a) RetinaNet (b) Baseline (c) SPNet

Fig. 4: Our system uses *RetinaNet* to detect food items. After detecting food items, our system can either use a baseline method to estimate a skewering location in the center or SPNet for estimating masks of skewering locations and rotations.

Section IV-A, servoes to it using the visual modality, acquires the bite using the haptic modality, and then feeds it to a person by detecting the face and servoing to it using the visual modality. Our system runs on Ubuntu 16.04.5 and uses ROS Kinetic [76] for communication between modules and visualization, AIKIDO [77] for planning and executing trajectories, and PyTorch [78] and Dlib C++ Library [79] for perception methods. Our face perception works with $\pm 80°$ roll, $\pm 30°$ pitch, $\pm 38°$ yaw, and is reasonably robust to mouth occlusion. We developed various ROS nodes which communicate with each other for the feeding task. The feeding node, which uses information from the other nodes, decides the sequence of actions to control the robot.

Fig.3 shows the flow of information in the system. First, the user selects a type of food for the next bite. Different interfaces depending on the abilities of the user are conceivable for this interaction, but since interface design is not the focus of this work, we used a terminal. The robot responds by moving its camera above the plate, perceiving the environment and checking if the selected food can be found. If it does find a suitable food item, a visual servoing control loop begins: While carefully moving the fork and camera closer to the desired food item, the system continually perceives the plate, tracks the selected food item and estimates its continuous parametrization of fork position and rotation using our perception module. We implemented the visual servoing procedure to help increase skewering precision in the presence of manipulation and perception uncertainties in realistic scenarios such as a non-rigid wheelchair base.

Once the fork tines touch the food item, the controller uses haptic feedback to skewer it. Once the force values exceed a threshold, the controller stops the skewering motion and the system continues with planning for the bite transfer step. This threshold depends on the type of food item and was adapted from the average force data obtained from the human experiments in [20]. On an average, approaching the food takes around 3 seconds and skewering it takes 1.3 seconds. Food perception takes anywhere between 120ms and 450ms depending on the number of items on the plate.

For bite transfer, we re-use our closed loop visual servoing capability to approach the user's face using discrete manipulation primitives of transfer angles. The system, while carefully moving the fork and food close to the user, continually re-perceives the face and adjusts its trajectory. In this case, visual servoing not only improves precision but also allows the robot to handle some degree of neck movement on the user's part. This is important because some disabilities involve involuntary spastic movements which may make the care-recipient move unpredictably. In our study (Section VII) the robotic system stopped the approach before the fork touched the mouth because of safety concerns. However, the haptic sensing modality enables the system to continue moving towards a user's mouth until it feels a slight touch based on haptic feedback from the physical interaction.

We run our bite acquisition and transfer experiments using this autonomous robotic feeding system. For bite acquisition, we designed our system to be able to use two discrete skewering primitives based on the skewering approach angles, *vertical* and *angled*, which are parameterized by the perception outputs. Similarly for bite transfer, our system can use two discrete transfer primitives, *horizontal* and *angled*. We also designed our system to use different continuous parametrizations of these manipulation primitives (See Section IV). However, the system design is independent of the specific sensors or robot used and we expect our methods to generalize to any robot equipped with haptic and visual sensing capabilities.

IV. Skewering Manipulation Primitives

For each of the discrete skewering manipulation primitives, we designed their continuous parametrizations not only for successful bite acquisition but also with easy bite transfer in mind. For example, the system can estimate an intelligent skewering position (*where*) and fork rotation (*how*) based on physical characteristics of a food item. We implement and train two neural networks to give the robot autonomous capabilities to answer these questions.

A. Skewering Primitives: Where and How to Skewer?

Perceiving food items is an essential initial step of bite acquisition. Our goal is to localize food items and estimate their pose in 3D space from the input RGBD stream. In a realistic situation, food items may be small, placed in a cluttered plate, and need to be perceived in real-time.

For detecting food items, we choose RetinaNet for the reasons mentioned in Section II-B. For each bounding box

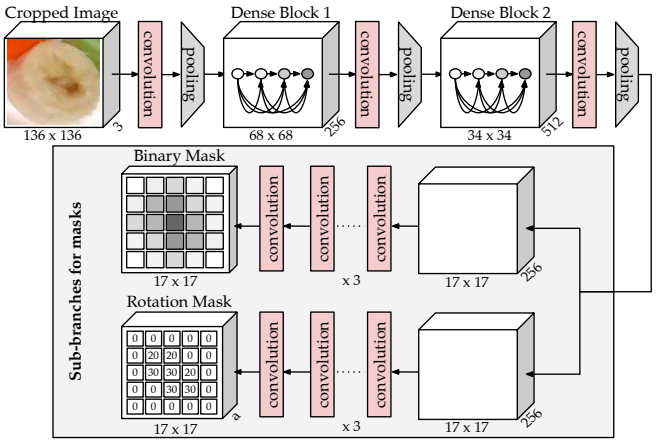


Fig. 5: SPNet architecture with two dense blocks and two sub-branches for binary and rotation masks.

generated by the object detector, our *Baseline* approach sets the midpoint of the bounding box as its skewering position. However, food items are of different shapes and physical characteristics. Therefore, we propose the *Skewering Position Network (SPNet)*, a network that estimates the skewering locations and rotations that could result in reliable bite acquisition and easy bite transfer for each bounding box (see Fig.5). SPNet consists of base convolutional layers and two sub-branches for binary masks and rotation masks. Compared with a single location and rotation per bounding box, masks can represent the probability distribution of all the possible skewering locations and corresponding rotations. We developed SPNet running on top of RetinaNet instead of extending Mask RCNN because we do not need pixel-wise mask generation and masks over the entire input scene – SPNet runs only for the food item we want to acquire.

The base convolutional layers are composed of multiple Convolution-Activation-Pooling sets and reduce the width and height of the input (136×136) to the size of binary and rotation masks (17×17). We developed two kinds of base convolutional layers. The simple network has only three Convolution-Activation-Pooling sets, and we adopt the dense block structure from DenseNet in order to increase the representation power of the network. We placed two dense blocks with different numbers of inner layers in the network. A shallower version has 3 and 6 inner layers in each dense block and a deeper version has 6 and 12 inner layers respectively.

The binary mask is one of the two branches on top of the base convolutional layers. After four convolutions without pooling, the final mask consists of 17×17 grid cells. Each cell represents the probability score of the skewering location at the center of the cell. We use the binary cross entropy loss for training this mask. The rotation mask consists of $17 \times 17 \times a$ grid cells, where a is the angle resolution represented by the number of classes for the discretized angles between $[0, \pi)$. We discretized angles because the rotation classification was more reliable than rotation regression for our early test cases. Furthermore, with discretization, we can handle the rotation free annotation as a class among other specific rotations. For example, when the angle resolution is 18, there are a total of 19 classes: class 0 represents any rotation while class $1 \sim 18$ denote specific angles, $(class\_id - 1) \times 10°$. We use the cross entropy loss to train the rotation for each cell.

B. Food Manipulation Dataset

We created a new dataset of food items to train RetinaNet and SPNet [21]. For the data collection, we maximized the variety of viewpoints and the selection and placement of food items in the scenes to increase the detection performance. We collected total 478 images including 349 real images by taking photos of plates of food items and 129 synthetic images which were rendered using the Unreal Engine.

We collected a total of 3,722 bounding boxes composed of 560 bananas (15%), 576 cantaloupes (15.5%), 556 carrots (14.9%), 636 celeries (17.1%), 597 eggs (16%), and 797 strawberries (21.4%). For the real images, we used an application, LabelImg [80], to manually record bounding boxes. For the synthetic images, we modified an Unreal Engine plugin, UnrealCV [81], so that it can automatically generate PASCAL VOC style annotations for random scenes. We generated a total of 2,954 masks for real images by using a new labeling application we developed. From the images and bounding box annotations, the application generates cropped images and overlays a 17×17 grid on the images so that a user can select skewerable grids and set a rotation value for a group of adjacent cells. Our dataset and code for the labeling application are available at [21], [82].

We categorized the six food items into three categories based on their shape and size: small, long, and round. Cantaloupes, bananas, and strawberries were in the *small* category, carrots and celeries were in the *long* category, and eggs were in the *round and heterogeneous* category. We generated the masks with specific rules for each category for effective feeding, so that SPNet learns our category-dependent strategies (see Fig.4), like skewering long items at their ends with the tines perpendicular to the item's long axis. We identified these skewering rules using the insights presented in [20].

C. Skewering Primitive Selection Performance

We tested multiple versions of SPNet with varying dense block sizes and angle resolutions. All the network variants including the simple SPNet were trained with the dataset until the training loss stabilized, and for the binary masks, they reached a similar test accuracy of 96.5%. However, the F1 score of the dense block SPNet was 11.76% higher than the simple SPNet. The recall of the dense block SPNet was higher than that of the simple SPNet while their precisions were similar.

All networks showed usable performance for the rotation masks. Among the various discretizations of angles such as 9, 18, 36, 90, and 180, we get the best performance when we discretized $[0, 180)$ into 18 angles. Thus, we chose the SPNet variant with shallow dense blocks and 18 angles resolution for bite acquisition. The performance of SPNet varies depending on the category of food items. See Table I for details.

The "small" category was the easiest one since it is rotation free and the skewering positions in the category are symmetric

TABLE I: SPNet performance per category of food items

Category	Mask Accuracy	Mask Precision	Mask Recall	Mask F1 Score	Rotation Error
S*	**0.974**	0.739	0.664	**0.693**	-1.000
L*	**0.974**	0.691	**0.551**	0.604	6.986
RH*	0.934	0.687	0.666	0.675	4.710

* S = Small, L = Long, and RH = Round and Heterogeneous.

and placed around the center of the masks. SPNet performed better for this category compared to the other categories.

The "long" category was more difficult than the "small" category. We generally labeled two groups of skewering positions per mask near each end, and the rotation of a group of skewering positions was perpendicular to the long edge of the item. SPNet showed 97.4% accuracy, but the recall of this category was lower than that of others. The main reason for the low recall is probably due to the random shapes of celeries.

The "round and heterogeneous" category only included eggs which were labeled with skewering positions along the edge of yolk based on inputs from previous human studies. SPNet performed well for this category and interestingly showed a high F1 score. Although the accuracy of binary masks was slightly lower than others, it was reasonable for our robot experiments.

V. Bite Acquisition Experiments

We performed experiments to determine the success rate of bite acquisition using various discrete manipulation primitives and their continuous parametrizations. We used our autonomous robotic feeding system (See Section III) for these experiments. We performed our experiments with 6 food items: bananas, cantaloupes, carrots, celeries, hard-boiled eggs, and strawberries. We selected these food items based on their varied shape, size, and compliance, which may affect bite acquisition [20]. Carrots and celeries are long and slender, cantaloupes and strawberries are small and round, bananas are soft and slippery, and hard-boiled eggs are heterogeneous with both yolk and white. Furthermore, caregivers mentioned that these food items are among the regular salad items that patients eat.

Our experiment consisted of the robotic arm autonomously acquiring food items with different skewering strategies: the baseline method, which skewers at the center of the food item, SPNet, and BLD (Bite Location Detector), the state of the art method used by Herlant [19].

Using each of these strategies, our autonomous robotic system skewered 2 plates of 5 pieces of each of these 6 food items, thus totalling to 3 skewering position strategies $\times$ 2 plates $\times$ 5 pieces $\times$ 6 food items $=$ 180 trials. Since both the baseline and SPNet methods are capable of letting the user select the food item to acquire, our system attempted skewering each item only once. This is in contrast to BLD, which doesn't let the user choose and thus is free to retry food items. We restrict BLD to 30 attempts to match the number of trials used for the other methods.

For two of the food items, bananas, and strawberries, we also compared two skewering approach angle strategies: *vertical* and *angled*. We had two discrete manipulation primitives for the *angled* primitive, one with horizontal tines and one with vertical. Bananas often slip off the fork, so we tilted the fork by 45 degrees to orient the tines more horizontally. When skewering strawberries, the tines tend to slip on the rounded surface without penetrating it. Therefore, we tilted the fork in the other direction to so that the tines skewered the strawberry vertically. Using SPNet, we skewered 2 plates of 5 items for bananas and strawberries, resulting in 20 additional trials which we compared with the vertical performance of the above experiment.

For each trial, the robotic arm perceived food items from above the plate, chose one of the food items to skewer, and lifted it off the plate. If it stayed for at least 5 seconds, we labelled the bite acquisition attempt as successful. After each trial, we removed the food item from the fork and discarded it.

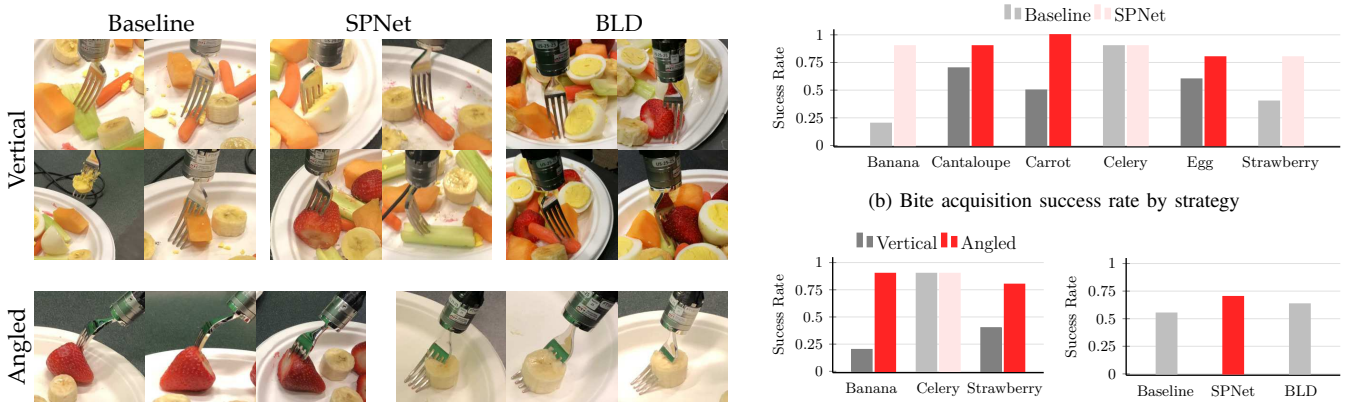
VI. Bite Acquisition Results

We found that SPNet generally outperformed the baseline approach, especially for cantaloupes, carrots, and eggs (See Fig.6(b)). The total success rate was 0.55 for the baseline, 0.7 for SPNet and 0.633 for BLD.

The results of baseline and SPNet are significantly different with $p < 0.05$ using the t-test $(t(29) = 2.19)$. In the case of carrots, this difference can be attributed to their long and slender shape: If the fork tines are not oriented perpendicular to the long axis of the carrot, they tend to slip off the curved surface. The intelligent orientation of the fork with respect to the food item solves this issue. Likewise, for eggs with the heterogeneous mixture of yolk and white, skewering those simultaneously prevents the yolk from falling off during lift-off, and thus leads to successful bite acquisition.

In comparison, the state-of-the art method BLD performed better than the baseline approach and slightly worse than SPNet. The current experiment did not find statistically significant differences with SPNet. However, it should be noted that BLD is free to choose which item to try next and skewer multiple items whereas the baseline method and SPNet are constrained to try to acquire each food item only one time. This implies that BLD could try skewering easier items multiple times and ignore difficult items.

Our second experiment compared the performance of vertical skewering with angling to adapt to the properties of the food item. Angling the tines more horizontally for bananas results in a success rate of 0.9, compared to 0.2 with a vertical fork, which is significantly different at $p < 0.01$ using a t-test $(t(9) = 3.15)$. For strawberries, angling the tines more vertically results in a success rate of 0.8, compared to 0.4 with a vertical fork, which is significantly different at $p < 0.05$ $(t(9) = 1.83)$. Generally, tilting the fork to adapt for food type specific difficulties results in big a improvement in success rate from 0.3 to 0.85 with a statistical significance of $p < 0.01$ $(t(19) = 3.52)$. Using SPNet together with fork angling for specific items results in a success rate of 0.88.



(a) A selection of skewering trials using different strategies

(b) Bite acquisition success rate by strategy

(c) Success rate by angle

(d) Success rate overall

Fig. 6: SPNet outperformed the baseline approach particularly for long food items. For strawberries and bananas, angled skewering improved the success rate significantly.

VII. Bite Transfer Study

For the next set of experiments, our objective was to find a set of manipulation primitives that makes biting off the fork easier. We hypothesized that users would prefer different bite transfer strategies for different food items and that choosing an appropriate combination of transfer and acquisition strategy would affect the ease of bite transfer.

Insight about the former was gained from a human study [20], which found that some people, when tasked with feeding a mannequin, tilted the fork for specific food items in order to orient the food and allow for easy bite transfer. Fig.7(a) illustrates the difference in the human participants' transfer angles. This told us that the shape of the food item may call for a different transfer angle. The workspace of our robot did not allow us to evaluate all of these angles, so we compared the horizontal approach with an angled approach tilted by $45°$, which was the maximum we could reach reliably. We believe that it is not so much angling the fork during transfer as it is the relative orientation of the food item with respect to the mouth that affects bite transfer. Thus, higher angular configurations can be achieved by not only angling the fork at transfer but also by picking up the food item at an angle in the first place, thus compounding the effect. Therefore we added a third strategy combination which consists of picking up the food item in an angled way and approaching the face at $45°$. We abbreviate these three combinations of strategies with *VS-HT* (vertical skewering-horizontal transfer), *VS-AT* (vertical skewering-angled transfer), and *AS-AT* (angled skewering-angled transfer).

To test the second hypothesis, we also employed another insight found by [20]: Some people skewered the food in a way that would make it easy for a recipient to take a bite without hitting the tines and therefore chose their bite acquisition method to improve bite transfer. To analyze the hypothesis with our system, we used SPNet to skewer a long food item at its ends. We designed the study such that the robot skewered the items at different ends and brought the food item to the user. This procedure resulted in the tines being closer or farther from the recipient during transfer.

To investigate the impact of these strategies on the the ease of bite transfer, we performed a study with 25 human participants from 18-37 years of age, under our organization's Institutional Review Board. 8 out of the 25 participants had experience feeding other people, 2 out of 25 were fed as an adult by someone else, and 7 out of 25 participants were female. We presented the participants with plates of 2 or 3 food items which our feeding system skewered and brought to their mouths autonomously using different manipulation primitives. We asked the participants to take a bite and rate how easy it was to take the bite off the fork, specifically if they had to strain themselves or move in an uncomfortable way. The recipients could either eat the food or bite it off the fork and discard it. The robot used a different strategy for each item on the plate and we required the participants to rank these strategies with respect to each other. We randomized the order in which the robot applied these different manipulation primitives and used SPNet for all of them since it had the highest acquisition success rate.

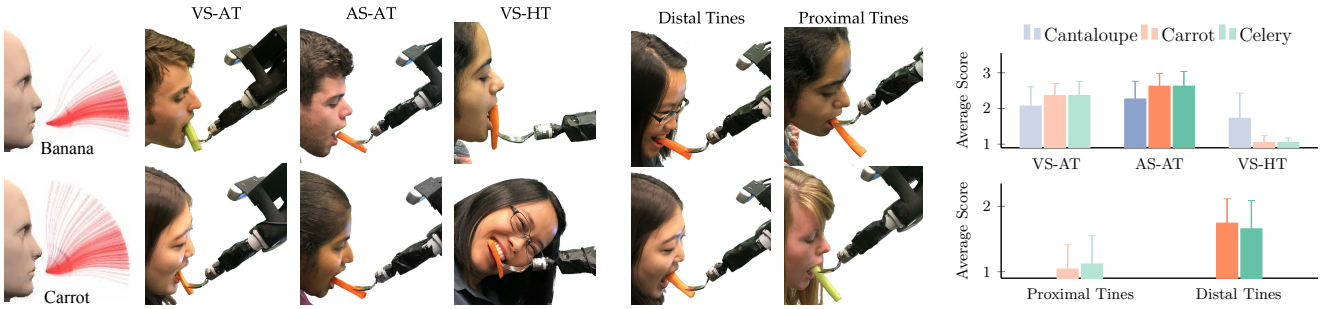
The study began by comparing the first 3 strategies: *VS-HT*, *VS-AT*, and *AS-AT* (See Fig.7(b)). Our robot fed the human users 2 plates of cantaloupes, carrots, and celeries with 1 food item for each strategy per plate. We selected these items based on their varied shapes and sizes.

Next, we evaluated the effect of the proximity of tines during bite transfer by having the participants eat one plate each with 2 carrots and 2 celeries using the vertical skewering and angled transfer, *VS-AT* strategy but skewered either at the near end or far end of the food item (see Fig.7(c)).

For each participant, we performed a total of 22 autonomous feeding trials (3 angular strategies $\times$ 2 plates $\times$ 3 food types + 2 far-near strategies $\times$ 2 plates).

VIII. Bite Transfer Results

For the ranked angular strategies, we analyzed the participants' ranking with the Friedman test [83] followed by a Nemenyi post-hoc analysis [83]. On average, angled skewering combined with angled transfer (*AS-AT*) resulted in significantly easier bite transfer than vertical skewering with horizontal transfer *VS-HT* ($p < 0.001$ with $df = 22$, $\alpha =$



(a) Human data (b) Combinations of acquisition and transfer strategies (c) Tine positioning strategies (d) Rankings for different strategies

Fig. 7: Humans feed using different transfer angles [20]. For ease of bite transfer, participants preferred manipulation primitives for bite transfer that depended on primitives for bite acquisition. VS-AT: *Vertical Skewering - Angled Transfer*, AS-AT: *Angled Skewering - Angled Transfer*, VS-HT: *Vertical Skewering - Horizontal Transfer*.

$0.001,\ q = 6.6 > q_{crit} = 6.065$) and vertical skewering with angled transfer *VS-AT* ($p < 0.001$ with $df = 22,\ \alpha = 0.001,\ q = 7.9 > q_{crit} = 6.065$) for long food items. Also, on an average for the food items in our study, angled skewering with angled transfer *AS-AT* makes a significant difference in the ease of bite transfer compared to vertical skewering and angled transfer *VS-AT* ($p < 0.001$ with $df = 22,\ \alpha = 0.001,\ q = 6.167 > q_{crit} = 6.065$). Thus, how a food item is skewered significantly impacts the bite transfer process. Interestingly, on a per food item basis, this is particularly true for celeries which were long in shape ($p < 0.001$ with $df = 22,\ \alpha = 0.001,\ q = 7.9 > q_{crit} = 6.065$). For carrots, both skewering and transfer strategies had an effect. For example, comparison between *VS-HT* and *AS-AT* ($p < 0.001$ with $df = 22,\ \alpha = 0.001,\ q = 6.1 > q_{crit} = 6.065$), as well as between *VS-AT* and *AS-AT* ($p < 0.001$ with $df = 22,\ \alpha = 0.001,\ q = 7.9 > q_{crit} = 6.065$) resulted in significant differences. However, for smaller and more cubic-like shaped cantaloupes, the current experiment did not find statistically significant differences between any of three manipulation primitives. This implies that for small food items, a simple strategy like vertical skewering with horizontal transfer may work quite well.

For the far-near strategies, we implemented Wilcoxon's Signed Rank test [83] to analyze the participants' ranking. On average, participants preferred to bite when the tines were distal compared to proximal for both carrots and celeries with statistical significance ($p < 0.001$ with $\alpha = 0.001,\ z = 5.35 > z_{crit} = 3.291$) with continuity correction. This implies that for long food items the choice of proximal vs. distal positioning of the tines affected the ease of bite transfer, and thus where a food item is skewered during bite acquisition significantly impacts the ease of bite transfer.

IX. Discussion

Assistive feeding is a problem with many facets. We saw multiple strategies in which bite acquisition significantly affected the ease of bite transfer. There could be other factors that may influence the perception of ease of bite transfer such as how close or far the food item is from the mouth or personal preferences for a food item. Responses to a qualitative question about the reasoning behind their ranking revealed some interesting insights. Some participants mentioned that for certain strategies taking the bite off the fork was difficult because of the curvature of the fork tines. This leads us to think that not only is the relative positioning between the food item and the mouth crucial but also factors that affect the physical interaction between the fork, food item, and the mouth are important. As found by Herlant [19], good food transfer timing is a crucial component of assistive feeding. While we developed a module to detect open and closed mouth states, a proper solution including the nuances of social dining falls outside the scope of this paper.

Note, the need to adapt the robot's motion to the user's face movement as well as the eye-in-hand camera system, which is inherently different from human body anatomy, necessitates the need for the robot to be positioned in certain pre-determined configurations during feeding to properly perceive the environment. However, the current system can be improved in future with more autonomous capabilities for realistic feeding situations such as generalizing perception to unseen food items, automated categorization of manipulation primitives based on food item characteristics, as well as using learning from demonstration techniques for annotations.

It is also important to note that this study was conducted with able-bodied participants who were able to move their neck slightly to be able to take a bite off the fork. This is however not a limiting factor for most of the target population because according to our interactions with occupational therapists and therapeutic recreational specialists with expertise in feeding, people with bilateral amputations and spinal cord injury have the necessary neck movement to be able to use the system as it is. However, note that our system assumes no cognitive or swallowing impairment on the user's side. Another qualitative study by Martinsen et al. [14] on caregiver-assisted feeding found that the objective of the interaction was to replicate a meal experience from before the disability. This also supports the premise of using insights learned from able-bodied peoples eating patterns to inform the way an assistive feeding device should behave [19]. In the future, we plan to investigate the capabilities of the system for both people with disabilities with similar range of neck movements as well as people with more severe motor impairments and no neck movements where the system would need to bring the food item close enough to actually touch the mouth.

REFERENCES

[1] C. Jacobsson, K. Axelsson, P. O. Österlind, and A. Norberg, "How people with stroke and healthy older people experience the eating process," *Journal of clinical nursing*, vol. 9, no. 2, pp. 255–264, 2000.

[2] M. W. Brault, "Americans with disabilities: 2010," *Current population reports*, vol. 7, pp. 70–131, 2012.

[3] M. Unosson, A. C. Ek, P. Bjurulf, H. H. von Schenck, and J. O. Larsson, "Feeding dependence and nutritional status after acute stroke." *Stroke*, vol. 25 2, pp. 366–71, 1994.

[4] A. Bäckström, A. Norberg, and B. O. Norberg, "Feeding difficulties in long-stay patients at nursing homes. caregiver turnover and caregivers' assessments of duration and difficulty of assisted feeding and amount of food received by the patient." *International journal of nursing studies*, vol. 24 1, pp. 69–76, 1987.

[5] B. Martinsen, I. Harder, and F. Biering-Sorensen, "The meaning of assisted feeding for people living with spinal cord injury: a phenomenological study." *Journal of advanced nursing*, vol. 62 5, pp. 533–40, 2008.

[6] L. Perry, "Assisted feeding," *Journal of advanced nursing*, vol. 62, no. 5, pp. 511–511, 2008.

[7] S. D. Prior, "An electric wheelchair mounted robotic arm-a survey of potential users," *Journal of medical engineering & technology*, vol. 14, no. 4, pp. 143–154, 1990.

[8] C. A. Stanger, C. Anglin, W. S. Harwin, and D. P. Romilly, "Devices for assisting manipulation: a summary of user task priorities," *IEEE Transactions on rehabilitation Engineering*, vol. 2, no. 4, pp. 256–265, 1994.

[9] J. S. Kayser-Jones and E. S. Schell, "The effect of staffing on the quality of care at mealtime." *Nursing outlook*, vol. 45 2, pp. 64–72, 1997.

[10] A. Chio, A. Gauthier, A. Vignola, A. Calvo, P. Ghiglione, E. Cavallo, A. Terreni, and R. Mutani, "Caregiver time use in als," *Neurology*, vol. 67, no. 5, pp. 902–904, 2006.

[11] D. Marshall, "Food as ritual, routine or convention," *Consumption Markets & Culture*, vol. 8, no. 1, pp. 69–85, 2005.

[12] T. Delormier, K. L. Frohlich, and L. Potvin, "Food and eating as social practice–understanding eating patterns as social phenomena and implications for public health," *Sociology of Health & Illness*, vol. 31, no. 2, pp. 215–228, 2009.

[13] M. Visser, *The rituals of dinner: The origins, evolution, eccentricities, and meaning of table manners*. Open Road Media, 2015.

[14] B. Martinsen, I. Harder, and F. Biering-Sorensen, "The meaning of assisted feeding for people living with spinal cord injury: a phenomenological study," *Journal of Advanced Nursing*, vol. 62, no. 5, pp. 533–540, 2008.

[15] "Obi," https://meetobi.com/,[Online; Retrieved on 25th January, 2018].

[16] "My spoon," https://www.secom.co.jp/english/myspoon/food.html,[Online; Retrieved on 25th January, 2018].

[17] "Meal-mate," https://www.made2aid.co.uk/productprofile?productId=8&company=RBF%20Healthcare&product=Meal-Mate,[Online; Retrieved on 25th January, 2018].

[18] "Meal buddy," https://www.performancehealth.com/meal-buddy-system,[Online; Retrieved on 25th January, 2018].

[19] L. V. Herlant, "Algorithms, Implementation, and Studies on Eating with a Shared Control Robot Arm," Ph.D. dissertation, 2016.

[20] T. Bhattacharjee, G. Lee, H. Song, and S. S. Srinivasa, "Towards robotic feeding: Role of haptics in fork-based food manipulation," *IEEE Robotics and Automation Letters*, 2019.

[21] D. Gallenberger, T. Bhattacharjee, Y. Kim, and S. Srinivasa, "A dataset of food items with skewering location and rotation masks," 2018. [Online]. Available: https://doi.org/10.7910/DVN/GHV7XZ

[22] M. Bollini, J. Barry, and D. Rus, "Bakebot: Baking cookies with the pr2," in *The PR2 workshop: results, challenges and lessons learned in advancing robots with a common platform, IROS*, 2011.

[23] M. Beetz, U. Klank, I. Kresse, A. Maldonado, L. Msenlechner, D. Pangercic, T. Rhr, and M. Tenorth, "Robotic roommates making pancakes," in *2011 11th IEEE-RAS International Conference on Humanoid Robots*, Oct 2011, pp. 529–536.

[24] M. C. Gemici and A. Saxena, "Learning haptic representation for manipulating deformable food objects," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2014, pp. 638–645.

[25] P. Chua, T. Ilschner, and D. Caldwell, "Robotic manipulation of food products–a review," *Industrial Robot: An International Journal*, vol. 30, no. 4, pp. 345–354, 2003.

[26] D. Park, Y. K. Kim, Z. M. Erickson, and C. C. Kemp, "Towards assistive feeding with a general-purpose mobile manipulator," *CoRR*, vol. abs/1605.07996, 2016. [Online]. Available: http://arxiv.org/abs/1605.07996

[27] D. Park, Z. Erickson, T. Bhattacharjee, and C. C. Kemp, "Multimodal execution monitoring for anomaly detection during robot manipulation," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, May 2016, pp. 407–414.

[28] D. Park, H. Kim, Y. Hoshi, Z. Erickson, A. Kapusta, and C. C. Kemp, "A multimodal execution monitor with anomaly classification for robot-assisted feeding," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Sep. 2017, pp. 5406–5413.

[29] J. Mahler, F. T. Pokorny, B. Hou, M. Roderick, M. Laskey, M. Aubry, K. Kohlhoff, T. Kröger, J. Kuffner, and K. Goldberg, "Dex-net 1.0: A cloud-based network of 3d objects for robust grasp planning using a multi-armed bandit model with correlated rewards," in *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 1957–1964.

[30] J. Mahler, J. Liang, S. Niyaz, M. Laskey, R. Doan, X. Liu, J. A. Ojea, and K. Goldberg, "Dex-net 2.0: Deep learning to plan robust grasps with synthetic point clouds and analytic grasp metrics," 2017.

[31] J. Mahler, M. Matl, X. Liu, A. Li, D. Gealy, and K. Goldberg, "Dex-net 3.0: Computing robust robot suction grasp targets in point clouds using a new analytic model and deep learning," *arXiv preprint arXiv:1709.06670*, 2017.

[32] A. Saxena, J. Driemeyer, and A. Y. Ng, "Robotic grasping of novel objects using vision," *The International Journal of Robotics Research*, vol. 27, no. 2, pp. 157–173, 2008. [Online]. Available: https://doi.org/10.1177/0278364907087172

[33] J. Redmon and A. Angelova, "Real-time grasp detection using convolutional neural networks," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, May 2015, pp. 1316–1322.

[34] L. Pinto and A. Gupta, "Supersizing self-supervision: Learning to grasp from 50k tries and 700 robot hours," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, May 2016, pp. 3406–3413.

[35] D. Katz, Y. Pyuro, and O. Brock, "Learning to manipulate articulated objects in unstructured environments using a grounded relational representation," 06 2008.

[36] B. Frank, R. Schmedding, C. Stachniss, M. Teschner, and W. Burgard, "Learning the elasticity parameters of deformable objects with a manipulation robot," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Oct 2010, pp. 1877–1883.

[37] A. Agah and K. Tanie, "Human interaction with a service robot: mobile-manipulator handing over an object to a human," in *Proceedings of International Conference on Robotics and Automation*, vol. 1, April 1997, pp. 575–580 vol.1.

[38] A. Edsinger and C. C. Kemp, "Human-robot interaction for cooperative manipulation: Handing objects to one another," in *RO-MAN 2007 - The 16th IEEE International Symposium on Robot and Human Interactive Communication*, Aug 2007, pp. 1167–1172.

[39] M. Huber, M. Rickert, A. Knoll, T. Brandt, and S. Glasauer, "Human-robot interaction in handing-over tasks," in *RO-MAN 2008 - The 17th IEEE International Symposium on Robot and Human Interactive Communication*, Aug 2008, pp. 107–112.

[40] G. Canal, G. Alenyà, and C. Torras, "Personalization framework for adaptive robotic feeding assistance," in *Social Robotics*, A. Agah, J.-J. Cabibihan, A. M. Howard, M. A. Salichs, and H. He, Eds. Cham: Springer International Publishing, 2016, pp. 22–31.

[41] M. Cakmak, S. S. Srinivasa, M. K. Lee, J. Forlizzi, and S. Kiesler, "Human preferences for robot-human hand-over configurations," in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sep. 2011, pp. 1986–1993.

[42] J. Aleotti, V. Micelli, and S. Caselli, "Comfortable robot to human object hand-over," in *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*, Sept 2012, pp. 771–776.

[43] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105.

[44] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," Sept. 2014.

[45] K. Simonyan and A. Zisserman, "Very deep convolutional networks for Large-Scale image recognition," Sept. 2014.

[46] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. cv-foundation.org, 2016, pp. 770–778.

[47] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," Aug. 2016.

[48] T. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 00, July 2017, pp. 936–944. [Online]. Available: doi.ieeecomputersociety.org/10.1109/CVPR.2017.106

[49] H. Hassannejad, G. Matrella, P. Ciampolini, I. De Munari, M. Mordonini, and S. Cagnoni, "Food image recognition using very deep convolutional networks," in *Proceedings of the 2Nd International Workshop on Multimedia Assisted Dietary Management*, ser. MADiMa '16. New York, NY, USA: ACM, 2016, pp. 41–49. [Online]. Available: http://doi.acm.org/10.1145/2986035.2986042

[50] A. Singla, L. Yuan, and T. Ebrahimi, "Food/non-food image classification and food categorization using pre-trained googlenet model," in *Proceedings of the 2Nd International Workshop on Multimedia Assisted Dietary Management*, ser. MADiMa '16. New York, NY, USA: ACM, 2016, pp. 3–11. [Online]. Available: http://doi.acm.org/10.1145/2986035.2986039

[51] K. Yanai and Y. Kawano, "Food image recognition using deep convolutional network with pre-training and fine-tuning," in *2015 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, June 2015, pp. 1–6.

[52] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," Apr. 2017.

[53] F. N. Iandola, M. W. Moskewicz, K. Ashraf, S. Han, W. J. Dally, and K. Keutzer, "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and <1mb model size," *CoRR*, vol. abs/1602.07360, 2016. [Online]. Available: http://arxiv.org/abs/1602.07360

[54] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, Dec 2015. [Online]. Available: https://doi.org/10.1007/s11263-015-0816-y

[55] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, Jun 2010. [Online]. Available: https://doi.org/10.1007/s11263-009-0275-4

[56] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. Lawrence Zitnick, and P. Dollár, "Microsoft COCO: Common objects in context," May 2014.

[57] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "OverFeat: Integrated recognition, localization and detection using convolutional networks," Dec. 2013.

[58] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," Nov. 2013.

[69] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional Encoder-Decoder architecture for image segmentation," Nov. 2015.

[59] R. Girshick, "Fast R-CNN," Apr. 2015.

[60] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time object detection with region proposal networks," June 2015.

[61] J. Dai, Y. Li, K. He, and J. Sun, "R-FCN: Object detection via region-based fully convolutional networks," May 2016.

[62] Z. Li, C. Peng, G. Yu, X. Zhang, Y. Deng, and J. Sun, "Light-Head R-CNN: In defense of Two-Stage object detector," Nov. 2017.

[63] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, Real-Time object detection," June 2015.

[64] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot MultiBox detector," Dec. 2015.

[65] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," Aug. 2017.

[66] Z. Li, C. Peng, G. Yu, X. Zhang, Y. Deng, and J. Sun, "DetNet: A backbone network for object detection," Apr. 2018.

[67] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*. cv-foundation.org, 2015, pp. 3431–3440.

[68] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Springer International Publishing, 2015, pp. 234–241.

[70] S. Jégou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio, "The one hundred layers tiramisu: Fully convolutional DenseNets for semantic segmentation," Nov. 2016.

[71] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," *IEEE Trans. Pattern Anal. Mach. Intell.*, June 2018.

[72] N. Dvornik, K. Shmelkov, J. Mairal, and C. Schmid, "BlitzNet: A Real-Time deep network for scene understanding," Aug. 2017.

[73] "Jaco robotic arm," https://www.kinovarobotics.com/en/products/robotic-arm-series,[Online; Retrieved on 27th August, 2018].

[74] "Rovi wheelchair," http://www.rovimobility.com/,[Online; Retrieved on 27th August, 2018].

[75] "Force-torque sensor," https://www.ati-ia.com/products/ft/ft_models.aspx?id=Nano25,[Online; Retrieved on 27th August, 2018].

[76] "Ros kinetic," http://wiki.ros.org/kinetic,[Online; Retrieved on 27th August, 2018].

[77] P. R. Lab, "AIKIDO," https://github.com/personalrobotics/aikido.

[78] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," 2017.

[79] "Dlib library," http://dlib.net/,[Online; Retrieved on 27th August, 2018].

[80] Tzutalin, "LabelImg," 2015.

[81] W. Qiu and A. Yuille, "UnrealCV: Connecting computer vision to unreal engine," Sept. 2016.

[82] D. Gallenberger, T. Bhattacharjee, Y. Kim, and S. Srinivasa, "Bite acquisition sampling application," 2018. [Online]. Available: https://github.com/personalrobotics/acquisition_sampler

[83] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *Journal of Machine learning research*, vol. 7, no. Jan, pp. 1–30, 2006.