Inter-finger Small Object Manipulation with DenseTact Optical Tactile Sensor

Won Kyung Do[1], Bianca Aumann[1], Camille Chungyoun[1], and Monroe Kennedy[1]

***Abstract*—The ability to grasp and manipulate small objects in cluttered environments remains a significant challenge. This paper introduces a novel approach that utilizes a tactile sensor-equipped gripper with eight degrees of freedom to overcome these limitations. We employ DenseTact 2.0 for the gripper, enabling precise control and improved grasp success rates, particularly for small objects ranging from 5mm to 25mm. Our integrated strategy incorporates the robot arm, gripper, and sensor to manipulate and orient small objects for subsequent classification effectively. We contribute a specialized dataset designed for classifying these objects based on tactile sensor output and a new control algorithm for in-hand orientation tasks. Our system demonstrates 88% of successful grasp and successfully classified small objects in cluttered scenarios.**

***Index Terms*—Dexterous Manipulation, In-Hand Manipulation, Grasping**

I. Introduction

GRASPING objects commonly found in daily environments is essential for human-robot collaboration tasks. Nevertheless, in-hand manipulation and grasping in cluttered settings continue to pose significant challenges in robotics. Recent research has increasingly focused on incorporating tactile feedback as a vital element in control systems to manage contact kinematics and manipulation tasks more effectively.

Despite this, the specific issue of grasping small objects in cluttered environments remains largely unresolved. When a robot interacts with an object, the situation changes, requiring a revised approach. This adaptability is common in human interactions but challenging for robots. The solution involves enabling robots to manipulate or identify small objects in cluttered scenarios.

Tactile sensors are instrumental in overcoming these issues. When grasping objects in cluttered spaces, traditional external vision systems often prove insufficient. Visuotactile sensors, however, offer a remedy by providing high-resolution data in localized areas. Additionally, hemispherical tactile sensors like DenseTact offer enhanced sensing capabilities and greater adaptability in terms of deformation, which is advantageous for compliance control.

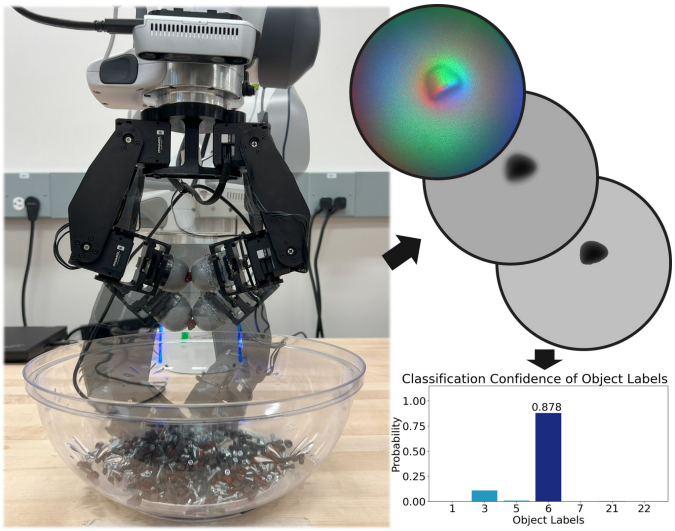

Fig. 1: **Overview of the grasping and classifying of small objects in cluttered environments.** The left image shows the process of grasp and control to classify the object, the right top shows the result of images from the sensor, and the right bottom shows the result of classification confidence of object labels.

In this study, we use tactile sensing and extra degrees of freedom on the gripper to tackle grasping, manipulating, and classifying small objects in cluttered environments. The transient dynamics of small objects, simulation challenges, and inadequacy of traditional controls post-grasp complicate the problem.

The main contributions of this paper shown in Fig. 1 are:

1) Development of a novel gripper with DenseTact 2.0, featuring 8 degrees of freedom for rolling manipulation.
2) Establishment of an integrated strategy involving the robot arm, gripper, and sensor for the manipulation and orientation of small objects for classification.
3) Creation of a dataset for classifying small objects based on tactile sensor outputs.
4) Successful classification and manipulation of objects smaller than the sensor and gripper sizes.
5) Design of a new control algorithm for in-hand orientation tasks involving 'unknown' small objects.

The paper is structured as follows: Section II reviews related works; Section III outlines the problems addressed; Section IV discusses the methodologies for gripper development, perception, object grasping, manipulation, and classification; Section

Manuscript received: August 30, 2023; Revised October 18, 2023; Accepted November 10, 2023.

This paper was recommended for publication by Editor Hong Liu upon evaluation of the Associate Editor and Reviewers' comments. This work was supported by the National Science Foundation under Grants 2142773 and 2220867. The first author is supported by a fellowship from the Kwanjeong Educational Foundation. *(Corresponding author: Won Kyung Do.)* Project website with videos are available here: https://sites.google.com/view/inter-finger-manipulation

[1]All Authors are with the School of Engineering, Mechanical Engineering Department, Stanford University, USA. {wkdo, biancalj, camillec, monroek}@stanford.edu

Digital Object Identifier (DOI): see top of this page.

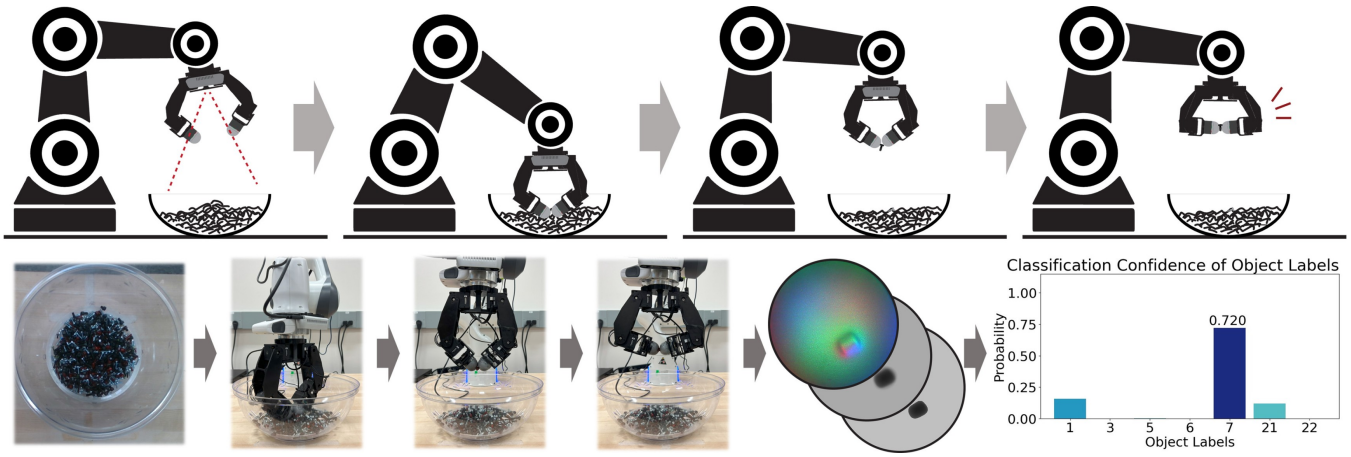

Fig. 2: The pipeline includes small-object grasping, pinching, classifying, and sorting. The first row depicts the overall process, while the second row displays process images and label classification confidence.

V presents the results and demonstrations, and conclusions and future work are discussed in Section VI.

II. Related Works

Grasping and manipulating small objects through tactile sensor input is a complex endeavor. A plethora of research initiatives have been aimed at various facets of this task. Specifically, in-hand manipulation has emerged as an active research domain in recent years. Works such as those cited in [1], [2] have proficiently tackled challenges associated with continuous contact or variations in friction during object interactions. In-hand manipulation employing external vision is discussed in [3], [4] and the use of Adaptive RL (reinforcement learning) policy derived from simulation torque input in robotic hands [5]. However, despite solutions to the sim-to-real problem, grasping objects in cluttered environments complicates policy training. Both model-based and model-free RL approaches often struggle in such dynamically altering environments. Moreover, relying solely on external vision for object orientation may become unfeasible during gripping, as the object becomes partially or fully occluded.

Tactile sensors, particularly visual tactile sensors, play a crucial role in in-hand manipulation and classification tasks. Placing a tactile sensor at the tip of the gripper enables intricate activities such as cable manipulation [6], [7], box packing [8], 3D pinching between fingers [9], [10], and grasping of both soft and rigid objects [11]. However, these tasks primarily focus on manipulating larger objects or involve specialized object manipulation, thus limiting their generalizability for handling small objects.

Tactile sensors are also effective in object or environmental classification. They can detect the hardness of objects, whether the sensor is vision-based or electrical transduction-based [12]–[15]. Classification of objects can be accomplished using multiple tactile sensors in a single grasp [16] or with vision-based tactile sensors [17]. However, the majority of these sensors are designed for classifying larger or deformable objects and may not be appropriate for small object classification in cluttered environments due to issues such as sensing resolution and gripper size. To address these challenges, we have developed a new gripper equipped with a sensor designed to both manipulate and classify small objects from a single grasp.

III. Problem Statement

This paper addresses the integrated tasks of grasping, re-orienting, and classifying small objects ($5mm \sim 25mm$) using optical tactile sensor input, all in a quasi-static state. The objects are smaller than the sensor size (30mm diameter). The components of the problem statement in this paper are defined as follows:

- **Grasping in Cluttered environment**: The primary challenge is grasping a small object from a cluttered bowl. We assume that the gripper interacts only with the objects, not the bowl itself. The task is solved using a robotic arm equipped with soft tactile sensors.
- **Object Reorientation**: After it is grasped, the object must be reoriented within the gripper for stable holding. This is achieved using a multi-degree of freedom (DOF) gripper.
- **Object Classification**: Finally, classification is performed using the tactile sensor on the gripper. Vision-based methods are unsuitable due to occlusion when the object is grasped.

IV. Method

A. Hardware setup

1) Gripper for inter-finger manipulation: Numerous gripper designs have been proposed for various tasks [18]. Among these, grippers capable of grasping small objects in cluttered environments often focus on specific usage or are limited to two parallel grippers. Even simple grippers typically require grasp detection and the prediction of the grasp pose to handle unknown objects using external vision [19]. However, small objects are challenging to grasp in cluttered environments, and the environment constantly changes as the gripper interacts with it. To address this challenge, we developed a gripper that can both grasp and manipulate small objects while the object is between the fingers.

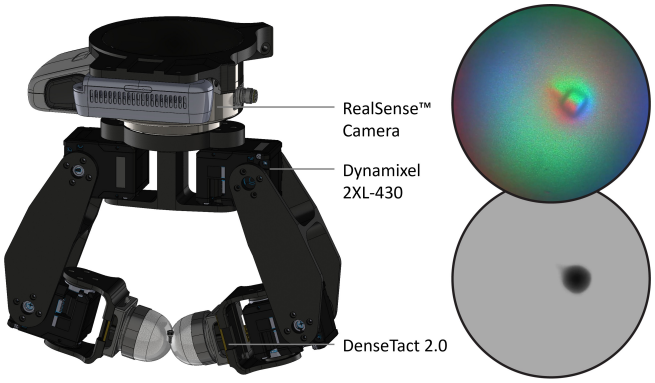

Fig. 3: 8-DOF gripper realistic model. The gripper can be attached directly to the Franka robot arm.

Given the nature of cluttered environments, where stable grasps are not guaranteed, and implementing re-grasping strategies or object manipulation through interaction with the environment can be difficult, we opted for a gripper with an additional DOF and compliant fingertips. This choice enhances our ability to achieve and maintain stable grasps.

We implemented four degrees of freedom for each finger to maximize inter-finger manipulation during grasping, enhancing the manipulation range during successful grasps. Effective inter-finger manipulation demands a maximal contact area between the fingertips of each finger. Assuming a deformable, hemispherical fingertip shape—beneficial for the unpredictability of cluttered environments—the contact workspace can be maximized if we rotate the fingertip in multiple directions while maintaining contact, ensuring the object remains securely held. To meet these requirements, we introduce a gripper design featuring two fingers with 4 DOF revolute joints for each finger. The left side of Fig. 3 presents this design, accompanied by a camera. As the fingers make contact and initiate the grasp, in-hand manipulation can be achieved through anti-symmetric fingertip motion by controlling the rotation along the x and z axes and the translation in the y direction in the gripper's frame. A gripper with 4 degrees of freedom offers enhanced control when grasping small objects, as demonstrated in Fig. 6. The gripper's working range during finger contact is illustrated in Fig. 4. The gripper uses four 2XL430-W250-T Dynamixel motors, boasting a total of eight degrees of freedom. This provides the gripper with a more expansive workspace and allows for dexterity beyond that of a conventional two-finger gripper. The gripper's arms are 3D-printed, ensuring it remains lightweight and reduces load.

2) Dimensions: The gripper is installable to the Franka robot arm by replacing the end-effector. The size of the arm between each joint is $(p_{12}, p_{23}, p_{34}, p_{4ft}) = (24mm,\ 95.52mm,\ 24mm,\ 55mm)$, where p_{ij} is the length between i-th and j-th joint, where ft refers to the fingertip. The diameter of the hemispherical part of the fingertip is 31mm, which is suitable for grasping a small object with a size of $5mm \sim 25mm$. STL and URDF files, complete with accurate mass and inertia values, are available on the project webpage.

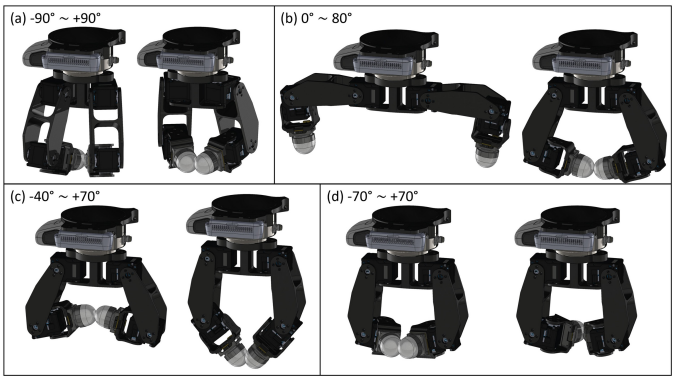

Fig. 4: Joint limit of the gripper for each joint.

3) Tactile compliance design: Soft fingertips in grippers have been shown to facilitate in-hand manipulation [20], [21]. We chose the DenseTact 2.0 sensor as the gripper's fingertip due to its compliant gel component [22], [23], which offers advantages over flat-surface sensors like Gelsight and Digit [24], [25]. The soft, rounded gel enhances the contact area and friction, leading to secure grasps, especially for small objects.The compliant nature of the DenseTact 2.0 gel allows the gripper to adapt to uncertainties and distribute force more evenly during grasping. This adaptability is particularly useful for handling objects of varying shapes and poses. In contrast to sensors like SoftBubble [26], DenseTact 2.0's hemispherical design offers a larger sensing area per volume, contributing to more precise in-hand manipulation.

When integrated with our multi-DOF gripper, the compliant features of the DenseTact 2.0 silicone enhance the gripper's versatility for manipulating a diverse range of objects. The fingertip can deform up to 20mm, facilitating a secure grasp and minimizing object damage. Additionally, DenseTact 2.0's deformation feedback aids in precise control, crucial for tasks like object orientation and dense-environment grasping.

B. Perception from Tactile Sensor

We used a tactile sensor on the gripper's end-effector to determine the object's pose and position. Opting for a patternless DenseTact 2.0 sensor, we focused on sensor deformation instead of force estimation, given the object's negligible mass and our quasi-static manipulation assumption. The sensor was calibrated using the method in [23], enabling depth image-based point cloud generation.

For experiments, we isolated relevant points from the point cloud by setting a 3mm threshold against the undeformed state. We then segmented the deformed points using DBSCAN [27], with specific distance and sample count parameters. A random 4% sample from the undeformed section was added to improve clustering. DBSCAN was chosen for its real-time applicability over alternatives like HDBSCAN [28]. During real-time control, the point cloud was truncated to 5000 points, allowing a frame rate of $10 \sim 13$Hz on an Intel Core i7-11800H CPU. Fig. 5 shows the segmented point cloud.

After segmenting, up to four labels were extracted, allowing the sensor to recognize a maximum of four objects per perception step. The label with the most points was prioritized

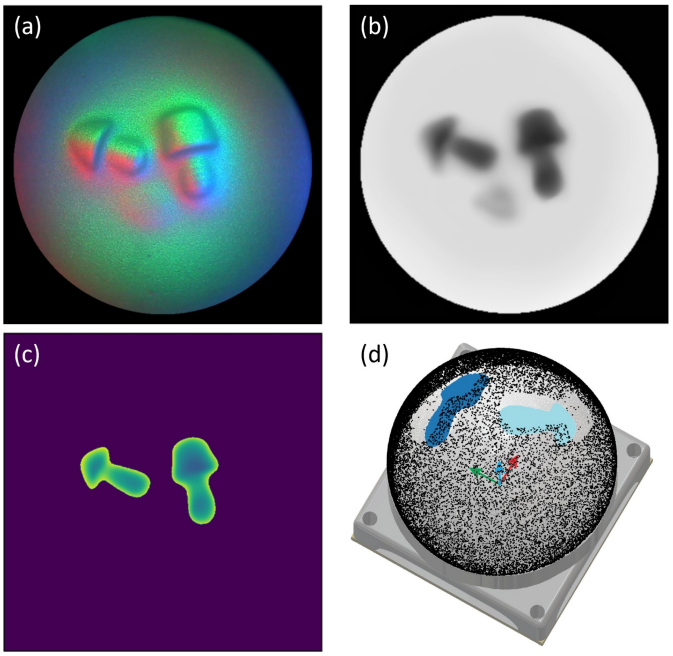

Fig. 5: The tactile sensor measurement results are as follows: (a) displays the sensor's captured RGB input, while (b) presents the estimated depth output. (c) features the filtered and labeled depth output, and (d) illustrates the clustered point cloud using DBSCAN, overlaid on the tactile sensor.

during control. For classification, all labels contributed to the training dataset. Thus, the labeled point cloud is effectively transformed into inputs for either control or classification tasks.

C. Grasping small object in cluttered environment

Grasping a small object in a cluttered, ever-changing environment is challenging. We tackled this issue by using depth data from the robot arm and the adaptive capabilities of soft fingertips. The use of soft fingertips enhances adaptive grasping capabilities, especially in cluttered environments characterized by high uncertainty. Thus, if the gripper can position itself to the desired point in the cluttered environment, a simple closing motion with the soft fingertip can easily achieve grasping, even in a highly uncertain cluttered environment.

However, objects need to be in a specified region of interest for successful grasping. The challenge amplifies in a cluttered space with diverse small objects, like a bowl filled with assorted items, as the pile's profile changes during grasping attempts. To counter this, we use depth data from a RealSense camera on the Franka arm to identify the highest points in our target area, shown in Fig. 3.

Our goal is to fine-tune the gripper's pose to maximize grasping success. We determined the optimal position and orientation within a square region, 24mm on each side, centered on the target point. From the camera's depth information, we extracted the top 800 elevation points in our target region. We then calculated the average position of these points in the world frame, represented as ${}^{W}\mathbf{p}_{\text{mean}} = [x_{\text{mean}}, y_{\text{mean}}, z_{\text{mean}}]^T$. A vector, $\mathbf{v}$, is defined as the difference between this mean position and the center of the opening rim of the bowl (fixed point), ${}^{W}\mathbf{p}_{\text{cen}}$. Additionally, the orientation angle, θ, the angle for the 7th joint of the franka arm, is derived from the horizontal components of this vector:

$$ {}^{W}\mathbf{v} = {}^{W}\mathbf{p}_{\text{mean}} - {}^{W}\mathbf{p}_{\text{cen}}, \quad \theta = \tan^{-1}\left(\frac{v_x}{v_y}\right) \tag{1} $$

During the Detection and Grasping phase, the gripper first moves to the position 60cm above the desk. As shown in the first bottom left image of Fig. 2, the depth camera detects the pile and returns the position to grasp. During this stage, the gripper remains open. Next, the gripper moves to the center position, adjusting the orientation of the last joint by the computed rotation angle, θ. Following this adjustment, the gripper advances guided by the vector $\mathbf{v}$, ensuring its trajectory towards the pile is both optimal in angle and position, thereby maximizing the success rate of the grasp. Finally, the gripper grasps the object by closing the gripper, and we move the gripper's position 2 seconds after the gripper grasps the object.

D. In-hand Orientation of Small Objects

After the gripper grasps an object and detects it within the fingertip via DenseTact, the small objects that have been grasped often deviate from the center of the gripper's fingertip. This deviation necessitates additional inter-finger manipulation for a stable grasp and proper classification. To address this challenge, we introduce a control strategy for securely grasping unknown small objects utilizing tactile feedback.

Even though the initial state of the gripper remains consistent, the objects it grasps are unpredictable and unfamiliar. Consequently, the controller's primary goal is to align the fingertip's position with the detected object while ensuring consistent pressure between the two fingertips of the gripper. Therefore, the objectives of our controller are to: 1) to maintain a specific distance between the fingertips, ensuring a stable grasp, 2) to maneuver the gripper within its joint limits; and 3) to center the fingertip's origin with the grasped object.

We select the state of our controller as

$$ \mathbf{x} = \{{}^{G}y, {}^{G}\theta_x, {}^{G}\theta_z\} \in \mathbb{R}^3 \tag{2} $$

where the ^{G} refers to the gripper frame, ${}^{G}y$ is the y-coordinate position of the fingertip in the gripper frame, ${}^{G}\theta_x, {}^{G}\theta_z$ are the angles of the fingertip coordinate frame in x and z axis of the gripper frame respectively, as defined in the left image of the Fig. 6. According to the figure, the Jacobian of one finger can be defined as the following:

$$ \dot{\mathbf{x}}_{all} = J_{all}\,\dot{\mathbf{q}}, \quad J_{all} = \begin{pmatrix} J_v & J_w \end{pmatrix}^T, \quad J_{all} \in \mathbb{R}^{6\times 4} \tag{3} $$

Where $\mathbf{x}_{all} \in \mathbb{R}^6$ refers to the position and angular position of the fingertip, and $\mathbf{q} = \begin{pmatrix} q_1 & q_2 & q_3 & q_4 \end{pmatrix} \in \mathbb{R}^4$ is the joint value of the one finger of the gripper. From the Jacobian of the joint, we can extract the corresponding differential value of each state by building the new Jacobian. Furthermore, since we have additional DOF for the new Jacobian, we can control the gripper to move within the joint limit through null space:

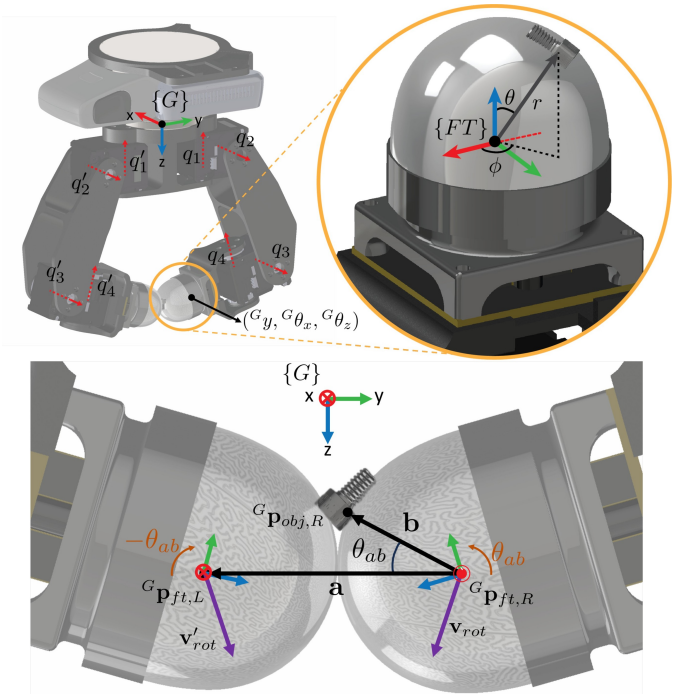

Fig. 6: Axis of the gripper and controller input specified in the fingertip coordinate frame. The image below shows the magnified view of the gripper while grasping an object.

$$J = \begin{pmatrix} J_{v,2} \\ J_{w,1} \\ J_{w,3} \end{pmatrix} \in \mathbb{R}^{3\times 4}, \quad J^+ = (J^T J + \lambda^2 I)^{-1} J^T \tag{4}$$

Where $J_{v,i}$ or $J_{w,i}$ is the i-th row of the velocity Jacobian or angular Jacobian respectively. Then, the desired joint position can be computed by integrating the desired velocity through a geometric controller. The desired joint velocity can be computed as:

$$\dot{\mathbf{q}} = J^+ \mathbf{v}_{\text{des}} + (I - J^+ J) f_{pen}(\mathbf{q}) \tag{5}$$

Where $f_{pen}(\mathbf{q}) = -C(\mathbf{q}_{curr} - \mathbf{q}_{mid})$. f_{pen} refers to the penalty term to ensure the joint inside of the range, C is constant for the penalty term, $\mathbf{q}_{mid}$ is the median of joint value in the joint range. The v_{des} is the desired fingertip velocity. From the desired fingertip velocity, we can get the desired joint position command.

Due to the absence of prior information about the object, and given our objective is its classification, the controller's goal needs dynamic adjustments. Based on tactile sensor input, we compute the controller's goal as **minimizing the θ_{ab} while maintaining the grasp of object**, where θ_{ab} is:

$$\theta_{ab} = \cos^{-1} \frac{\mathbf{a}\cdot\mathbf{b}}{|a||b|}, \quad \hat{\mathbf{v}}_{rot} = \frac{\mathbf{b}\times\mathbf{a}}{|a||b|} \tag{6}$$

$$\mathbf{a} = {}^G p_{ft,L} - {}^G p_{ft,R}, \quad \mathbf{b} = {}^G p_{obj,R} - {}^G p_{ft,R} \tag{7}$$

Where ${}^G p_{ft,L}$ and ${}^G p_{ft,R}$ are the position of the left and right fingertip (center of the hemispherical tactile sensor, the origin of the fingertip coordinate frame), and ${}^G p_{obj,R}$ is the position of the detected object in the right fingertip in the gripper frame. Those values are also shown in the bottom image of Fig. 6. We derive ${}^G p_{obj,R}$ by averaging the bottom 30% of point clouds with the lowest deformation values. This means we leverage the point clouds that occupy the top 30% in terms of the **r** value, as illustrated in the top right image of Fig. 6. This strategy lets the gripper determine the subsequent movement point without settling on the currently detected state. From the above value, we can get the desired velocity:

$$v_{des} = K_p \begin{pmatrix} \dot{\mathbf{p}}_{ft,R} \\ \omega_x \\ \omega_z \end{pmatrix} = \frac{K_p}{\Delta t} \begin{pmatrix} C_y - \lambda \\ \hat{\mathbf{v}}_{rot,x}\theta_{ab} \\ \hat{\mathbf{v}}_{rot,z}\theta_{ab} \end{pmatrix} \tag{8}$$

Where $K_p \in \mathbb{R}^{3\times 3}$, Δt is a constant value, C_y is a constant value which refers to the offset of the fingertip from the contact, and λ is the deformed radius value of the detected object in the DenseTact. From the first row, the gripper can maintain constant pressure while keeping contact between the fingertip and the object. Since the gripper detects the object before the controller starts, the object will always exist while the controller is executed. $\hat{\mathbf{v}}_{rot,x}, \hat{\mathbf{v}}_{rot,z}$ are the x and z components of the rotation axis, respectively. The process finishes when the fingertip and the object's center align. The controller is operated for one finger of the gripper, and the other finger gets the same value to achieve the anti-symmetric movement for a stable grasp rolling without slipping.

Given the small, lightweight nature of the target objects, inertial and force inputs are less relevant and unpredictable. We thus use a position controller that integrates the commanded velocity (Eqn. 5) in a quasi-static state, and this controller choice is driven by motor control limitations as well as the pipeline's real-time processing speed (10Hz $\sim$ 13Hz).

While the controller could be modeled through optimization or RL, these options present challenges. The complex gel deformation we're tackling is best represented by hyperelastic material models like the Ogden hyperelastic model, which require computationally heavy FEM programs [29]. Additionally, RL or dynamic learning approaches often need extensive simulated data, making them less practical for our task. Other issues involve sim-to-real gaps and errors in dynamic modeling.

E. Small Object Segmentation

1) Dataset Collection: Tactile sensors are crucial for object identification in grasping, particularly with soft fingertips that significantly occlude the object. External vision proves insufficient for object verification in such cases. Leveraging the high-resolution ($640\times 640\times 3$) input from tactile sensors, we curated a dataset of objects grasped between two such sensors. Each object was positioned on one sensor and encapsulated by pressing the other sensor onto it. Live RGB and depth images were captured once the object became discernible. For each object type, 50 RGB and 50 corresponding depth images were collected during a single press.

The dataset also accommodates scenarios of grasping one or two small objects simultaneously. We focused on select combinations due to the exponential increase in potential

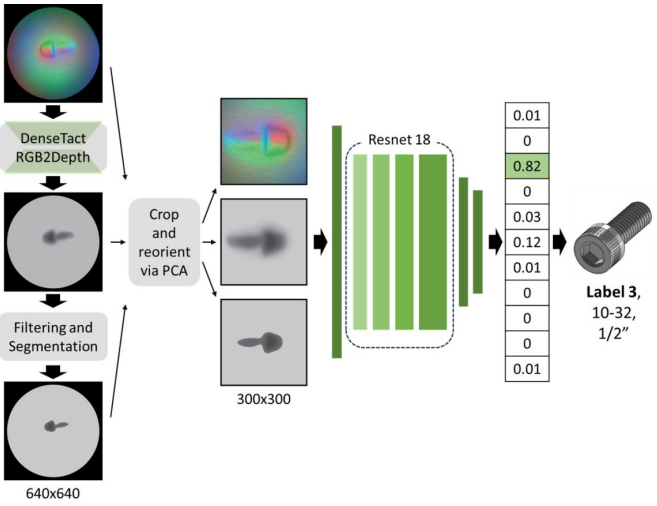

Fig. 7: Pipeline architecture for small object classification.

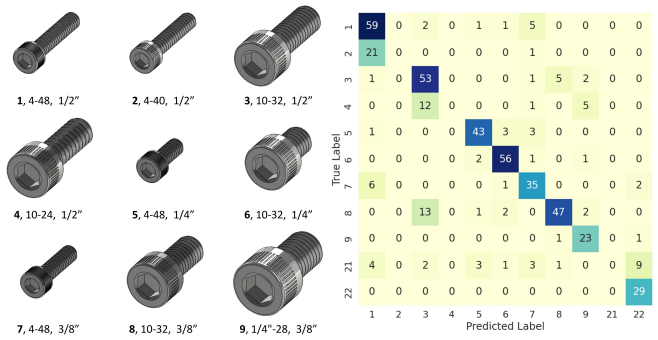

Fig. 8: Classification result of screw objects and classified labels for 466 touches.

object pairings, denoted by $\frac{n(n-1)}{2}$ for n different objects. Our dataset comprises 20 types of small objects, including 9 varieties of screws and 11 daily objects.

The screw dataset was designed such that the screws vary by length, head diameter, and thread size. Different combinations of variables are changed across the dataset classes. Each screw is either 1/2”, 3/8”, or 1/4” in length. Furthermore, each screw has a head diameter and thread spacing combination of either 4-40, 4-48, 10-24, 10-32, or 1/4”-28.

2) Preprocessing step of the Image Input: Given the relatively small size of the dataset when compared to the variety of object types (20 distinct objects), directly utilizing the raw input from the tactile sensor becomes impractical. Additionally, there’s a potential for classification errors when the gripper unintentionally captures two small objects simultaneously. This challenge can be addressed by integrating an additional input layer and conducting suitable image preprocessing.

Initially, we incorporated input from the labeled image derived through DBSCAN, along with the RGB and depth images generated by the tactile sensor. Following this, the labeled and deformed pointcloud was extracted and projected onto the depth image. We then employed PCA analysis to ascertain the orientation and center of the deformed point. Given the prior labeling of the deformed pointcloud, PCA analysis was conducted for each labeled pointcloud. As indicated in section IV-B, PCA can handle up to four labels in a single tactile input.

Relying on the central values and angles obtained from the PCA, the images were cropped to a size of 300×300, and rotated according to the identified angle. By integrating RGB, depth, and labeled images, the resulting input dimension became $300 \times 300 \times 5$. This preprocessing approach enhances the efficiency of network training, even with a limited dataset size. Furthermore, the labeling step allows the localization of the classified objects and completes the segmentation of the multiple objects detected from the raw sensor image.

3) Model for Classification: The network architecture chosen for classification is grounded in the ResNet18 framework, a decision driven by the compact size of our dataset, as shown in Fig. 7. Rather than maintaining the model in its static form and solely training the concluding MLP layer, we opted to activate the final fourth layer block for training while keeping the other layer blocks of ResNet18 static [30]. Preceding the initial layer block of ResNet18, a 2D convolutional layer accepts the input, which is subsequently processed through batch normalization, ReLU, and a max-pooling layer. After the fourth layer block of ResNet18, two fully connected layers are employed, utilizing a hidden channel size 256. Ultimately, a softmax function is invoked to classify the object type. The classification confidence can be visualized as a one-hot vector in Fig. 7, or as a bar chart in Fig. 1, or in Fig. 2.

Training was conducted on a composite dataset, incorporating single-object and multi-object datasets. This amalgamation inherently led to a disparity in the dataset count for individual object types. To counterbalance this, 12% of the total dataset was randomly collected as a testing set for every object category. The number of datasets per class was recorded while splitting training and testing datasets. This count was then employed as a weight in the cross-entropy loss calculation throughout the training phase. Utilizing the Adam optimizer, we set a learning rate of 2×10^{-5} and a weight decay of 1×10^{-4} over a span of 400 epochs and with a batch size of 8. The training duration was approximately an hour, executed on four NVIDIA A4000 GPUs.

V. Experiment

A. Classification

Before physically demonstrating the complete procedure, the classification results were evaluated using the test dataset. The right image in Fig. 8 displays the confusion matrix for the classified screws. In the left image, the classified label, thread size, and length of each screw are indicated at the bottom of their respective images. Label 21 is designated for instances involving two screws, while Label 22 signifies that the sensor either detected a plane or failed to detect the screw. Given that Labels 1 and 2 share identical lengths and head sizes but differ in thread type, it’s understandable that Label 2 is occasionally classified as Label 1. A similar misclassification occurs between Labels 3 and 4. Due to the combinations required for a two-screw dataset exceeding 45 distinct cases, achieving uniform dataset size via human input proved challenging. Nonetheless, by accumulating a

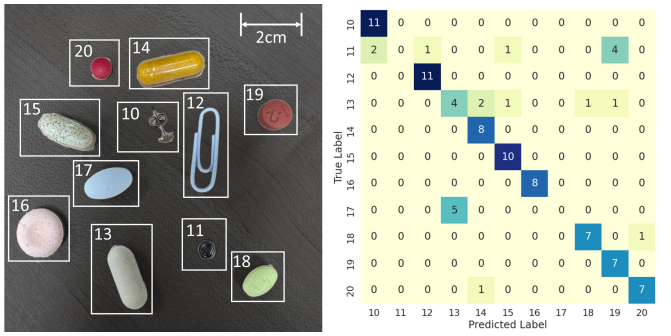

Fig. 9: Classification result of small random objects and classified labels for 93 test touches.

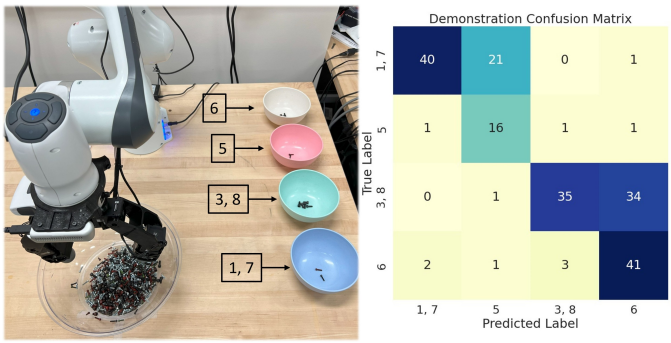

Fig. 10: Experimental setup and demonstrated result of the whole process for 198 grasps.

broader combination of datasets or gathering additional data during demonstrations, we believe classification errors can be reduced.

Fig. 9 presents the confusion matrix for the classification of other objects. The left image annotates the corresponding labels for each object. We opted for a diverse set, including various types of pills, earrings, and paperclips. Despite the dataset's limited size, the classifier has demonstrated proficiency in correctly identifying most objects.

B. Full pipeline

Building on the methods described and as shown in Fig. 2, we structured the grasping, reorienting, and classification sequence into a finite state machine with several key states: 'initial,' 'detect,' 'ready for grasp,' 'grasp,' 'control,' and 'classification.' In the 'detect' phase, the system cycles back to detection if the depth camera provides inadequate sensor values. Sensor feedback determines grasp success after the 'grasp' phase; failure redirects the process back to the 'detect' phase. When the gripper grasps multiple objects at the same time, the controller can drop one during rolling, classifying just one screw if the screws are spaced widely. However, screws sometimes interlock or touch in ways that necessitate finger sliding rather than the rolling method outlined in the paper. Consequently, during 'classification,' if the sensor detects 'two screws' (interlocking or touching) or 'plane' (detects only the other finger), the system reverts to the 'initial' state for a new cycle.

Given the small object sizes and the sensor's high deformability, one tactile sensor usually suffices for single object detection and manipulation. Experiments were conducted using output from a single sensor while the other finger moved in tandem. As the objects classified are symmetrical, consistent sensor readings are assured for both fingers. Classification of asymmetric objects, though feasible, would require data from both object facets.

C. Demonstration result

Utilizing the established pipeline and integrating all processes, we executed the object sorting task autonomously, eliminating the need for human intervention. The left side of Fig. 10 illustrates the experimental setup during the demonstration, whereas the right side depicts the confusion matrix derived from the demonstration results. The cluttered environment is represented in a transparent bowl and only the depth camera was employed to detect the highest point of the pile. For this experiment, all objects in the bowl are screws. The pile comprises 50 objects with Labels 3, 6, and 8. These objects are considerably larger in size compared to others. Additionally, 100 objects with Labels 1, 5, and 7, recognized by their smaller head size, are present within the environment. The variation in object numbers stems from the gripper's inherent tendency to seize larger objects, due to its grasping characteristics. Importantly, the entire grasping process remained devoid of human influence (for instance, altering the pile profile during the demonstration or manual re-grasping). The gripper still exhibited a marked preference for grasping larger-headed screws.

The process of grasping the objects proved largely successful. Out of 225 attempts, there were 198 successful grasps. In contrast, there were 12 instances of unsuccessful grasping and 12 trials where the results were classified under Labels 21 or 22, indicating scenarios where two screws were grasped simultaneously or when a plane was detected. Consequently, **88%** of the trials resulted in successful object extraction from the cluttered environment and subsequent object classification.

The results presented in Fig. 10 highlight a recurring misclassification. Specifically, objects with Labels 1 or 7 were frequently mistaken for Label 5, while objects Labeled 3 and 8 were often misclassified under Label 6. This trend can be attributed to the gripper's occasional tendency to grasp the head of the screw first and hold the grasp. When this happens, even after the finger position is changed, there is a possibility that the sensor only observes the head part of the screw. Both Labels 1 and 7 possess long screws that share head sizes with Label 5, while Labels 3 and 8 have similar characteristics with Label 6. One of the results of the example can be observed in the sensing results displayed in Fig. 2. Given that objects under Labels 1, 5, and 7 have identical screw heads, misclassifications amongst them are plausible. However, Labels 5 and 6 are not mistakenly classified under other Labels, mainly due to the shorter lengths of these objects, which increased the likelihood of head detection during dataset collection. There were instances of failed trials where the gripper occasionally grasped two objects simultaneously, rendering the secondary object invisible to the sensor. Such challenges could be potentially addressed by leveraging sensor

feedback from both fingertips.

VI. Conclusions

In this study, we present a novel approach for manipulating and classifying small objects in cluttered settings using optical tactile sensors. Our key innovation is a gripper fitted with DenseTact 2.0, designed to enhance both grasping success and post-grasp manipulation, thanks to its highly deformable soft fingertip. A unique manipulation strategy using a newly devised Jacobian combination ensures stable grasps and precise classification.

Our network model efficiently classifies objects, even with a limited dataset, demonstrating broad applicability to general small objects. The end-to-end pipeline operates autonomously, underscoring the potential for human-free small object classification and manipulation. This work not only advances current grasping strategies and object pose estimation techniques but also lays the groundwork for more versatile robotic grasping solutions.

This control strategy relies on highly curved and soft deformable surface sensors, limiting its application to such sensors. Future research could focus on utilizing tactile sensors on both fingertips to improve grasping stability and classification, addressing the concurrent grasping of multiple objects, and extending the gripper's functionality in human-robot collaborative settings. Future work could involve integrating extra perception stages to grasp specific objects in cluttered environments using our proposed approach.

References

[1] C. B. Teeple, B. Aktaş, M. C. Yuen, G. R. Kim, R. D. Howe, and R. J. Wood, "Controlling palm-object interactions via friction for enhanced in-hand manipulation," IEEE Robotics and Automation Letters, vol. 7, no. 2, pp. 2258–2265, 2022.

[2] N. Chavan-Dafle, R. Holladay, and A. Rodriguez, "Planar in-hand manipulation via motion cones," The International Journal of Robotics Research, vol. 39, no. 2-3, pp. 163–182, 2020.

[3] A. Handa, A. Allshire, V. Makoviychuk, A. Petrenko, R. Singh, J. Liu, D. Makoviichuk, K. Van Wyk, A. Zhurkevich, B. Sundaralingam, et al., "Dextreme: Transfer of agile in-hand manipulation from simulation to reality," arXiv preprint arXiv:2210.13702, 2022.

[4] A. S. Morgan, K. Hang, B. Wen, K. Bekris, and A. M. Dollar, "Complex in-hand manipulation via compliance-enabled finger gaiting and multimodal planning," IEEE Robotics and Automation Letters, vol. 7, no. 2, pp. 4821–4828, 2022.

[5] H. Qi, A. Kumar, R. Calandra, Y. Ma, and J. Malik, "In-Hand Object Rotation via Rapid Motor Adaptation," in Conference on Robot Learning (CoRL), 2022.

[6] Y. She, S. Wang, S. Dong, N. Sunil, A. Rodriguez, and E. Adelson, "Cable manipulation with a tactile-reactive gripper," The International Journal of Robotics Research, vol. 40, no. 12-14, pp. 1385–1401, 2021.

[7] A. Wilson, H. Jiang, W. Lian, and W. Yuan, "Cable routing and assembly using tactile-driven motion primitives," in 2023 IEEE International Conference on Robotics and Automation (ICRA), 2023, pp. 10 408–10 414.

[8] S. Dong and A. Rodriguez, "Tactile-based insertion for dense box-packing," in 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2019, pp. 7953–7960.

[9] E. Psomopoulou, N. Pestell, F. Papadopoulos, J. Lloyd, Z. Doulgeri, and N. F. Lepora, "A robust controller for stable 3d pinching using tactile sensing," IEEE Robotics and Automation Letters, vol. 6, no. 4, pp. 8150–8157, 2021.

[10] X. Mao, Y. Xu, R. Wen, M. Kasaei, W. Yu, E. Psomopoulou, N. F. Lepora, and Z. Li, "Learning fine pinch-grasp skills using tactile sensing from real demonstration data," CoRR, vol. abs/2307.04619, 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2307.04619

[11] M. C. Welle, M. Lippi, H. Lu, J. Lundell, A. Gasparri, and D. Kragic, "Enabling robot manipulation of soft and rigid objects with vision-based tactile sensors," arXiv preprint arXiv:2306.05791, 2023.

[12] J. A. Solano-Castellanos, W. K. Do, and M. Kennedy III, "Embedded object detection and mapping in soft materials using optical tactile sensing," arXiv preprint arXiv:2308.11087, 2023.

[13] W. Yuan, M. A. Srinivasan, and E. H. Adelson, "Estimating object hardness with a gelsight touch sensor," in 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2016, pp. 208–215.

[14] I. Andrussow, H. Sun, K. J. Kuchenbecker, and G. Martius, "Minsight: A fingertip-sized vision-based tactile sensor for robotic manipulation," Advanced Intelligent Systems, vol. 5, no. 8, p. 2300042, 2023. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/aisy.202300042

[15] A. Drimus, G. Kootstra, A. Bilberg, and D. Kragic, "Design of a flexible tactile sensor for classification of rigid and deformable objects," Robotics and Autonomous Systems, vol. 62, no. 1, pp. 3–15, 2014, new Boundaries of Robotics. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S092188901200125X

[16] L. Deng, Y. Shen, G. Fan, X. He, Z. Li, and Y. Yuan, "Design of a soft gripper with improved microfluidic tactile sensors for classification of deformable objects," IEEE Robotics and Automation Letters, vol. 7, no. 2, pp. 5607–5614, 2022.

[17] N. Venkatayogi, O. C. Kara, J. Bonyun, N. Ikoma, and F. Alambeigi, "Classification of colorectal cancer polyps via transfer learning and vision-based tactile sensing," in 2022 IEEE Sensors, 2022, pp. 1–4.

[18] B. Zhang, Y. Xie, J. Zhou, K. Wang, and Z. Zhang, "State-of-the-art robotic grippers, grasping and control strategies, as well as their applications in agricultural robots: A review," Computers and Electronics in Agriculture, vol. 177, p. 105694, 2020.

[19] H. Tian, K. Song, S. Li, S. Ma, J. Xu, and Y. Yan, "Data-driven robotic visual grasping detection for unknown objects: A problem-oriented review," Expert Systems with Applications, p. 118624, 2022.

[20] M. Ciocarlie, C. Lackner, and P. Allen, "Soft finger model with adaptive contact geometry for grasping and manipulation tasks," in second joint eurohaptics conference and symposium on haptic interfaces for virtual environment and teleoperator systems (WHC'07). IEEE, 2007, pp. 219–224.

[21] Q. Lu and N. Rojas, "On soft fingertips for in-hand manipulation: Modeling and implications for robot hand design," IEEE Robotics and Automation Letters, vol. 4, no. 3, pp. 2471–2478, 2019.

[22] W. K. Do and M. Kennedy, "Densetact: Optical tactile sensor for dense shape reconstruction," in 2022 International Conference on Robotics and Automation (ICRA). IEEE, 2022, pp. 6188–6194.

[23] W. K. Do, B. Jurewicz, and M. Kennedy, "Densetact 2.0: Optical tactile sensor for shape and force reconstruction," in 2023 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2023, pp. 12 549–12 555.

[24] W. Yuan, S. Dong, and E. H. Adelson, "Gelsight: High-resolution robot tactile sensors for estimating geometry and force," Sensors, vol. 17, no. 12, p. 2762, 2017.

[25] M. Lambeta, P.-W. Chou, S. Tian, B. Yang, B. Maloon, V. R. Most, D. Stroud, R. Santos, A. Byagowi, G. Kammerer, et al., "Digit: A novel design for a low-cost compact high-resolution tactile sensor with application to in-hand manipulation," IEEE Robotics and Automation Letters, vol. 5, no. 3, pp. 3838–3845, 2020.

[26] A. Alspach, K. Hashimoto, N. Kuppuswamy, and R. Tedrake, "Soft-bubble: A highly compliant dense geometry tactile sensor for robot manipulation," 2019.

[27] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, et al., "A density-based algorithm for discovering clusters in large spatial databases with noise." in kdd, vol. 96, no. 34, 1996, pp. 226–231.

[28] L. McInnes, J. Healy, and S. Astels, "hdbscan: Hierarchical density based clustering." J. Open Source Softw., vol. 2, no. 11, p. 205, 2017.

[29] R. W. Ogden, "Large deformation isotropic elasticity–on the correlation of theory and experiment for incompressible rubberlike solids," Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences, vol. 326, no. 1567, pp. 565–584, 1972.

[30] 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, 2016, pp. 770–778.