The Teenager's Problem: Efficient Garment Decluttering With Grasp Optimization

Aviv Adler[1]*, Ayah Ahmad[1]*, Shengyin Wang[2], Wisdom C. Agboh[1,2], Edith Llontop[1]
Tianshuang Qiu[1], Jeffrey Ichnowski[3], Mehmet Dogar[2], Thomas Kollar[4], Richard Cheng[4], Ken Goldberg[1]

***Abstract*— This paper addresses the "Teenager's Problem": efficiently removing scattered garments from a planar surface. As grasping and transporting individual garments is highly inefficient, we propose analytical policies to select grasp locations for multiple garments using an overhead camera. Two classes of methods are considered: *depth-based*, which use overhead depth data to find efficient grasps, and *segment-based*, which use segmentation on the RGB overhead image (without requiring any depth data); grasp efficiency is measured by *Objects per Transport (OpT)*, which denotes the average number of objects removed per trip to the basket. Experiments suggest that both depth- and segment-based methods reduce OpT by 20%; furthermore, these approaches complement each other, with combined *hybrid* methods yielding improvements of 34%. Finally, a method employing *consolidation* (with segmentation) is considered, which manipulates the garments on the work surface to increase OpT; this yields an improvement of 67% over the baseline.**

I. Introduction

We introduce the "Teenager's Problem": removing a large number of scattered garments from a surface (e.g. the floor of a teenager's room, or a work surface) in the shortest time. This problem has applications in hotels, retail dressing rooms, garment manufacturing, and other domains where scattered garments must be grasped efficiently.

We formalize the Teenager's Problem and then consider several methods to solve it. Consider Fig. 1 with multiple garments on a work surface. Given an overhead RGB or RGBD image, what robot pick-and-place actions would minimize the total time to remove all of the garments? Removing individual garments, one at a time, would be inefficient. We consider how the robot can use the deformable nature of garments to pick multiple garments at once.

Given a scene like the one in Fig. 1, one approach is to ignore the separation between individual garments and to treat the whole scene as a homogeneous volume to be removed. This motivates *depth-based* methods, i.e., methods that use the depth image to infer grasp points that would remove as much volume as possible. We consider two depth-based methods in this paper. The first method uses *height* and grasps at the highest point of the scene. The second method estimates *volume*, by integrating the depth data within a grasp radius, and grasps at the garment point in the scene that gives the largest estimated volume.

Fig. 1: An instance of the *Teenager's Problem* with a UR5 robot, Robotiq gripper, and RGBD overhead camera not shown. The scale automatically records weight in the bin.

The depth-based methods are able to identify large heaps and grasp multiple garments at a time, even when some of these garments are completely buried under others and not individually visible to the camera. However, the depth-based methods also miss some good grasps. Particularly, since they do not detect individual garment positions and boundaries, they miss grasp points that have lower height/volume but would still pick multiple garments simultaneously, e.g., points where multiple garment boundaries meet.

A second approach to solving the Teenager's Problem is to distinguish between individual garments and optimize grasps to pick as many garments as possible. This motivates *segment-based* methods, i.e., methods that use the RGB image to segment the individual garments. We use the Segment

*Equal Contribution
[1]The AUTOLab at University of California, Berkeley, USA.
[2]University of Leeds, UK.
[3]Carnegie Mellon University, USA.
[4]Toyota Research Institute, USA.

Anything Model (SAM) [1] to distinguish the individual garments. Then, given a set of garments and a candidate grasp point, it predicts the probability that those garments will be picked by that grasp. The segment-based method uses these predictions to optimize the grasp to pick the largest number of garments.

While the segment-based method is able to identify grasps that would pick multiple garments simultaneously, it can also miss some good grasps. Particularly, since only the top surface of the garment pile is visible to the camera, garments that are under others are ignored by the segment-based method.

We also consider a *hybrid* method that combines the complementary strengths of the two approaches. The hybrid method combines the depth-based method and segment-based method, depending on the maximum height available.

We also consider a method that makes use of *consolidation* actions, which are movements within the workspace to gather the garments into heaps, before removing them. This improves the efficiency of grasps to transport the objects to the bin at the cost of the time used in consolidation, which can improve efficiency in cases where the removal bin is located far from the work surface.

Experiments suggest that depth- and segment-based methods significantly improve OpT by 20%; furthermore, these approaches complement each other, with combined *hybrid* methods yielding improvements of 34%; finally, OpT can be further by 67% above the baseline by taking additional *consolidation* actions within the workspace set up extremely efficient transport actions.

We make the following contributions:

- A formalization of the Teenager's Problem.
- Five methods (two depth-based, one segment-based, and two hybrid) to generate effective multi-garment grasps.
- A method that uses heap consolidation along with the segment-based grasp generation method to efficiently solve the Teenager's Problem.
- Physical experiments and data from grasping 2000 garments, that compare the performance of the various methods above, as well as a random baseline.

II. Related work

Our work is related to two lines of work: *manipulation of deformable objects* and *multi-object grasping*.

A. Manipulation of Deformable Objects

Prior work on deformable object manipulation includes folding [2, 3], fabric smoothing [4, 5, 6], bed-making [7], untangling ropes [8], and singulating clothes from a heap [9, 10]. Several works aimed to detect specific features, such as the corners and edges of fabrics, and to identify optimal grasp points [11, 12, 13]. Other techniques employ deep learning to identify successful grasps [14, 15]. Some studies have focused on determining optimal grasp points by considering not only the depth of the cloth but also targeting wrinkles as highly graspable regions [16, 17, 18]. These prior works focus on manipulating a single deformable object at a time or singulating a deformable object from among others. Our work, on the other hand, concentrates on grasping multiple garments simultaneously.

B. Multi-object Grasping

Multi-object grasping can improve decluttering efficiency [19]. It has been studied, with analytic methods [20], learning-based methods [21, 22], and with special gripper designs [23]. The focus, however, has remained on rigid objects. Instead, we consider the problem of grasping multiple deformable objects at a time, and using such grasps to efficiently clear a surface.

Multi-object grasping scenarios can encompass cluttered [24] environments, which can include both deformable and rigid objects [25], however, the goal is to singulate the objects to grasp them individually. Prior work on manipulating multiple rigid objects used methods such as pushing, stacking, and destacking [26, 19, 27]. In cluttered scenes with multiple rigid objects, one method for determining how, or where, to grasp is by using image segmentation [28], detecting and isolating individual objects in the scene. We also use a segmentation approach but for deformable objects.

III. The Teenager's Problem

We formulate the *Teenager's Problem* as follows: deformable objects rest on a planar work surface. A fixed target bin is provided, and the goal is to transfer all the garments efficiently from the workspace to the bin with the smallest number of grasps (which naturally maximizes OpT).

A. Problem Statement

The Teenager's Problem models garments as sets in $\mathbb{R}^2$ and grasps as tuples $(x, y, \theta) \in \mathbb{R}^2 \times [-\pi/2, \pi/2]$; there is a *predictor* function p which, given a set $S \subset \mathbb{R}^2$ (representing a garment) and a grasp (x, y, θ), returns an estimate of the probability that the given grasp will pick up the garment:

$$\mathbb{P}[\text{grasp } (x, y, \theta) \text{ gets } S] = p(S, (x, y, \theta))$$

Finally, there is an accuracy target denoting the (minimum) desired probability of removing each garment.

The problem is then as follows: let $S_1, \ldots, S_n$ be sets in $\mathbb{R}^2$, each representing a garment, let p be a predictor function and let $\delta > 0$ be the accuracy target. The probability that a given sequence of grasps $(x_1, y_1, \theta_1), \ldots, (x_m, y_m, \theta_m)$ gets garment S is then given by:

$$\mathbb{P}[\{(x_i, y_i, \theta_i)\}_{i=1}^m \text{ gets } S] = 1 - \prod_{i=1}^{m} (1 - p(S, (x_i, y_i, \theta_i)))$$

Then, the objective is to find a set of grasps $(x_1, y_1, \theta_1), \ldots, (x_m, y_m, \theta_m)$ which minimizes m such that for all $j \in \{1, 2, \ldots, n\}$,

$$\mathbb{P}[\{(x_i, y_i, \theta_i)\}_{i=1}^m \text{ gets } S_j] \geq 1 - \delta\,.$$

The Teenager's Problem can thus be seen as a probabilistic variant of the classic Set Cover problem, since each grasp corresponds to the (weighted) set of garments it could potentially grasp, and the goal is to find a set of grasps whose

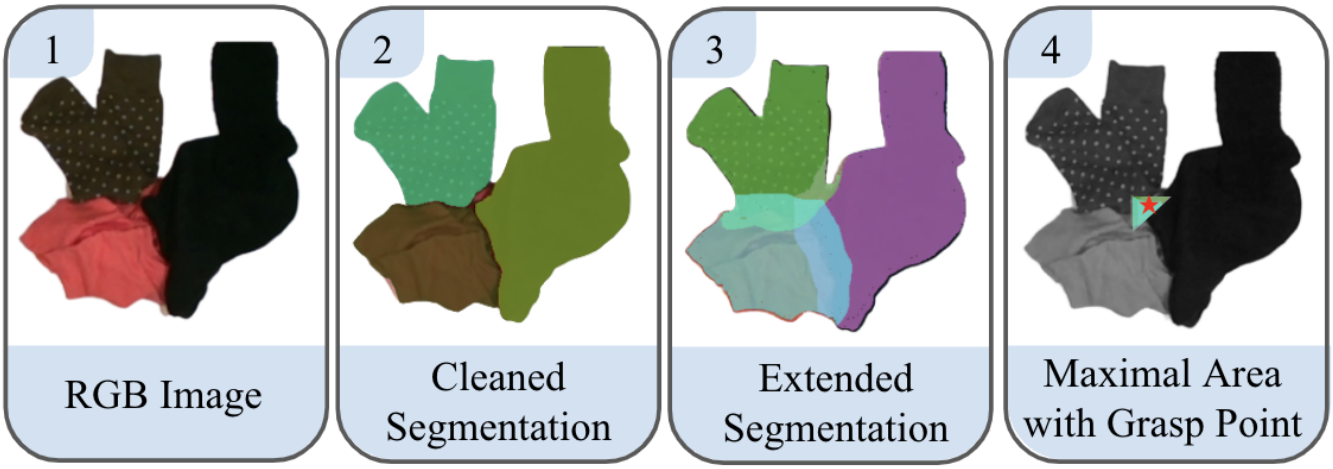

Fig. 2: An example of the segment-based grasp point selection algorithm (before orientations are chosen). **From left to right:** *(1)* The original overhead RGB image. *(2)* The cleaned segmentation $\mathcal{M}$. *(3)* The segments expanded by gripper radius to include 'nearby' garment points (all grasps must be done on garment points as a sanity check); overlapping regions thus correspond to points that are near multiple segments. *(4)* The maximal area (points near a maximal set of segments) with the chosen grasp point shown in red.

'union' encompasses all the garments. This also means that the benefit of a grasp depends on the set of other grasps that will also be carried out - even if it is likely to get several garments, it may not be useful if those garments were already likely to be removed by other grasps in the set.

B. The Analytic Predictor

The Teenager's Problem keeps the predictor function p general in order to accommodate a variety of different approaches, both analytic and learned, to estimating the effects of a grasp. This work studies in particular an analytic predictor which models the area under the gripper as an ellipse and the probability of a successful grasp as depending on the total area of the garment within the ellipse.

Specifically, any (x, y, θ) let $E(x, y, \theta)$ denote the ellipse centered at (x, y) whose major axis is oriented at angle θ with major axis length d_1 and minor axis length d_2; furthermore, let $b > 0$ be a normalization constant. Then the analytic predictor estimates the probability of successfully grasping a garment S with grasp (x, y, θ) as

$$p(S, (x, y, \theta)) = \frac{\text{area}(E(x, y, \theta) \cap S)}{\text{area}(E(x, y, \theta) \cap S) + b}.$$

The axis lengths are derived from the gripper dimensions, while b is chosen to keep the relevant values reasonable.

One property of this predictor is that, while it is important to choose grasps which get a large total area the segments within the ellipse, having several different segments with nonzero area under the ellipse is generally preferable to having just one (even if the total area within the ellipse is the same), because the benefit of added area for one segment decreases as the area already captured increases. Thus, the best grasps will generally occur at or near the boundary between multiple segments.

C. Metrics

The primary metric for evaluating the methods is *Objects per Transport* (OpT), which denotes the average number of objects taken during each transport and measures the general effectiveness of the performed grasps. We use OpT (as opposed to Picks Per Hour (PPH)) as our metric as OpT directly measures grasp quality and PPH depends heavily on implementation details (particularly concerning computation time) which reduces its reliability as a metric in this setting.

IV. Teenager's Problem Methods

This study explores various strategies for efficiently grasping multiple garments concurrently, categorized broadly into two types: *depth-based* and *segment-based* approaches.

All the methods described below use a pre-processing of the RGB pixels to determine the *garment points*, denoted $\mathcal{X}_g$. Since we assume the system knows the color and/or pattern of the background, this is achieved with color thresholding.

A. Depth-Based Methods

Depth-based methods use the depth output of the RGBD overhead camera to select the next grasp. To solve the Teenager's Problem, these methods are used repetitively until the workspace is clear: at each step, we capture a new depth image of the scene, use one of the methods below to generate a new grasp, and then execute that grasp. We examine two variations:

1) Height: This method selects the highest point in the scene and chooses the orientation to be the orientation of the major axis of a local principal component analysis (PCA) around the grasp point.

2) Volume: This method considers the total volume of garments in a disc of radius R around a candidate grasp point (x, y), which is estimated by summing the heights of all the pixels within that radius, and then selects the point with the largest total volume. As in the Height method, orientation is selected using a local PCA.

B. Segment-Based Method

The segment-based method divides the task into *cycles*, each corresponding to one instance of the Teenager's Problem; a segmentation is generated at the start of each cycle,

generating the input sets $S_1, \ldots, S_n$, and a sequence of grasps is carried out which is predicted to remove many sets (however, an overhead RGB image is still taken between each grasp). Each cycle relies on three subroutines:

(1) Segmentation and cleanup; (2) grasp selection; (3) execution. Once all the planned grasps have been carried out the next cycle begins. The steps of a cycle are detailed below.

1) Segmentation and cleanup: The method uses Meta's Segment Anything Model (SAM) [1] with the *vit-b* weights, and prompts the image as a whole. However, the initial segmentation often contains multiple overlapping segments, gaps, and regions corresponding to the work surface itself. The method then performs redundant segment removal, gap filling, and color thresholding to eliminate segments representing the work surface.

This produces a set of segments $S_1, S_2, \ldots, S_n$, which is used as an instance of the Teenager's Problem.

2) Grasp selection: Given the segments $S_1, \ldots, S_n$, the method then solves the Teenager's Problem (with the analytic predictor described in Section III-B) via a greedy heuristic. As the number of potential grasps is infinite, to simplify the problem, the method first selects a set of grasp points (excluding orientations) and then selects an orientation for each point.

The grasp point selection algorithm does the following (given a radius $r > 0$ in pixels, and segments $\mathcal{S} = S_1, S_2, \ldots, S_n$):

- For each garment pixel $(x, y) \in \mathcal{X}_g$, determine the set $A(x, y) \subseteq \mathcal{S}$ of segments which are within distance r from (x, y).
- Construct the set $\mathcal{A}$ of *maximal sets* $A(x, y)$, i.e. $\mathcal{A}$ consists of every $A(x, y)$ which is not a proper subset of some other $A(x', y')$.
- For each maximal set $A_i \in \mathcal{A}$, choose a (uniformly) random (x_i, y_i) such that $A(x_i, y_i) = A_i$.
- Return $\{(x_i, y_i)\}_{i=1}^{m}$, where m is the number of maximal sets.

The rationale behind this algorithm is that, as suggested by the Teenager's Problem with the analytic predictor, a grasp point should be near as many segments as possible but not too similar to other grasp points and that for any grasp point that is not in a maximal set, there is another that is near strictly more segments than it.

Then the algorithm chooses an orientation for each grasp point using a greedy heuristic. First, it enumerates a list of ℓ equally-spaced orientations in $[-\pi/2, \pi/2]$ and runs the analytic predictor p for each orientation to estimate the probability of getting each segment; the orientation which is predicted to remove the largest number of segments (i.e. the sum of the predicted probability of removal over all the segments) is then chosen for each grasp. For a balance between efficiency and thoroughness, we used $\ell = 6$.

3) Grasp execution: The algorithm then attempts all grasps in sequence, in increasing order of distance to the bin; this is to prevent, as far as possible, dragging garments from disturbing the positions of the garments that remain (which may cause garments to fall off the work surface).

An RGB image is also captured after each transfer to the bin (when the arm is out of frame), although a new segmentation is *not* generated (until the next cycle). Instead, for each planned grasp remaining, the difference between its current state and the state at the beginning of the cycle (when the segmentation was generated) is estimated using the squared difference between the pixel values within a small square neighborhood around the grasp point; if the difference is too large, the grasp is deemed to be in a different state from when it was planned and is not performed. This ensures that grasps are not performed unless the system knows that the local configuration of the garments is approximately the same as when it was planned.

C. Hybrid methods

One observation borne out by the experiments was that depth-based methods and segment-based methods have different strengths – in particular depth-based methods excel at picking occluded garments (which generally result in taller piles) while segment-based methods excel at simultaneously picking adjacent garments. This motivated the idea of considering *hybrid* methods which make use of both depth data and segmentation data.

To take advantage of how these methods complement each other, hybrid methods do the following: given a height threshold, if the tallest pile is taller than the threshold, execute a (single) grasp as given by the depth-based method; if all piles are below the threshold, execute one cycle of the segment-based method.

D. Segment-based method with consolidation

Another avenue to improving OpT is to first consolidate the garments into large piles for transport to the bin; this can improve overall efficiency in cases where the bin is located at some distance from the work surface, making transports costly relative to manipulations within the workspace. An efficient primitive for consolidation is the *grasp sequence* where each pick-and-place movement picks up where the last one placed; this both saves on robot movement time (no travel distance to the next pick point) and, ideally, allows the robot to accumulate more garments as it goes before depositing them in the bin.

An extension of the segment-based method above to include consolidations is the following:

1) generate the grasp points as in the no-rearrangement segment-based method;
2) starting from the furthest grasp point, estimate the *expected area* of grasped garments, and execute the next available grasp which does not exceed a given expected grasped area threshold;
3) if no such grasp exists, transport the currently-held garments to the bin.

Going from the furthest grasp point to the closest follows the intuition that a method using consolidation should consolidate towards the bin since this will always shorten the

Fig. 3: The test set of 10 garments, representing a variety of different sizes, weights, textures, colors, patterns, flexibility, and garment classes.

distance between the grasped garments and the bin, even if some are dropped along the way – and this will also tend to compress them into a smaller space, facilitating later multi-object grasps.

The grasp area threshold corresponds to the intuition that the gripper has a limit to the amount of fabric it can hold and thus trying to accumulate more than that limit in one grasp is counterproductive.

E. Baseline

Finally, as a baseline, we use the *random* method, which uniformly randomly selects a garment point $(x, y) \in \mathcal{X}_g$ with a uniformly random orientation $\theta \in [-\pi/2, \pi/2]$, accounting for the gripper's symmetry.

V. Experiments

We tested all algorithms on the test set of 10 garments (see Fig. 3) with 25 sample runs. Each sample run begins with a randomized scene containing all 10 test set garments, and ends when the workspace is cleared of garments.

A. Data collection pipeline

To run the experiments, we used a semi-autonomous data collection pipeline, in which experimental scene reset, randomization, and data recording are done automatically, with the experimenter only needing to correct problems when they arise (for instance, if a garment falls off the work surface, the experimenter must return it for the next sample). The system uses the recorded weight data to automatically notify the experimenter when such a problem occurs, to minimize the amount of human attention necessary for data collection.

The scene is automatically reset in the following way:

1) The robot grasps the bin and empties it over the work surface to deposit the garments, then places the bin back to its original position.
2) The robot executes a sequence of random pick-and-place actions on the surface to shuffle the garments; in our experiments, 10 such moves were performed for every scene reset.

Then the experiment is performed with the selected method, recording at each step the overhead RGBD output, grasp location and orientation, and weight of the garments in the bin. The experiment is paused for 3 seconds after every transport to allow the scale's output to settle.

For each algorithm tested, OpT was evaluated on all 25 sample runs, which were then averaged to yield the final result and 95% confidence bounds.

VI. Results

The results of the experiments are given in Table I, show that both depth-based and segment-based methods yield clear improvements (approximately 20% additional OpT) over the baseline. Furthermore, these approaches complement each other: hybridizing them yielded 26% and 34% more OpT (for volume/segment and height/segment respectively) as compared to the baseline. However, it is important to note that only the segment-based method achieves this without the use of depth information, which may not be available on all systems.

It should be noted that while grasp quality is the focus and OpT is the most meaningful metric for this work, the ultimate goal remains improving pick efficiency as measured by PPH. The depth-based methods, which do not perform significant computations to find grasps, improve PPH from 477 for the baseline to 526 and 525 for max-volume and max-height grasps respectively. While the segmentation method registers a slight decrease in PPH (to 453), optimizing the methods' speed may increase the PPH up to a comparable 523 (determined by subtracting computation time in the experiments) with no need for depth data.

Finally, at the cost of both computational overhead and additional physical actions, the segmentation with consolidation method improved OpT by 67% over the baseline.

A. Comparison of Methods

- Depth-based methods require both RGB and depth images to compute grasps. In contrast, segment-based methods only need inexpensive RGB images.
- Segment-based methods often demand more computational resources, as they involve neural network-driven segmentation and subsequent cleanup. To ensure efficient computation without compromising speed, it may be necessary to deploy GPUs.
- A notable challenge faced by segment-based methods is their limited ability to detect grasps that remove occluded garments. In contrast, depth-based methods more often grasp over occluded garments due to their utilization of depth information, which provides an enhanced perception of garment depth.
- Conversely, segment-based methods explicitly choose grasps to simultaneously capture garments situated closely together, whereas depth-based methods cannot

TABLE I: OpT (Objects per Transport) by grasping method

Random	Depth-Based Methods		Segment	Hybrid Methods		Segment with Consolidation
	Volume	Height		Volume/Segment	Height/Segment	
1.48 ± 0.11	1.76 ± 0.13	1.77 ± 0.20	1.77 ± 0.13	1.87 ± 0.13	1.98 ± 0.24	$\mathbf{2.47 \pm 0.2}$

determine which points are in proximity to multiple visible garments.

VII. Conclusion

We formalize the Teenager's Problem and develop both depth- and segment-based methods to solve it. We used recent advances in image segmentation [1] to explore an approach that uses it to distinguish garments in the image and find grasps that are likely to capture as many as possible.

A. Limitations and Future Work

This work has certain limitations and leaves a number of areas open for improvement:

- All the proposed methods rely on accurately separating the garments from the work surface using the RGB image, which is done here via color thresholding.
- All the methods considered here grasp at a fixed height above the work surface with a vertical gripper, the most efficient grasp may not share those characteristics. Additional improvements might be obtained by optimizing the grasp height or angle.
- While the 67% OpT increase from the segmentation with consolidation method is large, an average of roughly 4.0 rearrangement actions were performed per transport saved over the baseline. Nevertheless, such methods may increase efficiency in cases where transports are relatively costly, e.g. when the target bin is far from the workspace.

Extensions such as sorting of clothes (for instance, separating clothing by type or color) are a natural fit for the techniques discussed here, especially the segment-based approach. Additionally, although we present only the analytic grasp predictor, the segment-based method described in Section IV-B is designed to be compatible with any predictor. Another direction for future work is to improve the predictor. Finally, the Teenager's Problem formulation does not consider the possibility of rearrangement actions or of depth data, and may be generalized to include these aspects of the problem in future work.

References

[1] A. Kirillov *et al.*, "Segment anything," *arXiv:2304.02643*, 2023.

[2] Y. Avigal, L. Berscheid, T. Asfour, T. Kröger, and K. Goldberg, "Speedfolding: Learning efficient bimanual folding of garments," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022, pp. 1–8.

[3] R. Hoque *et al.*, "Learning to fold real garments with one arm: A case study in cloud-based robotics research," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022, pp. 251–257.

[4] D. Seita *et al.*, "Deep imitation learning of sequential fabric smoothing from an algorithmic supervisor," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 9651–9658.

[5] A. Ganapathi *et al.*, "Learning dense visual correspondences in simulation to smooth and fold real fabrics," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 11 515–11 522.

[6] S. Sharma *et al.*, "Learning switching criteria for sim2real transfer of robotic fabric manipulation policies," in *2022 IEEE 18th International Conference on Automation Science and Engineering (CASE)*, 2022, pp. 1116–1123.

[7] D. Seita *et al.*, "Deep transfer learning of pick points on fabric for robot bed-making," in *International Symposium of Robotics Research*, 2018.

[8] P. Sundaresan *et al.*, "Untangling Dense Non-Planar Knots by Learning Manipulation Features and Recovery Policies," in *Proceedings of Robotics: Science and Systems*, Virtual, Jul. 2021.

[9] B. Willimon, S. Birchfield, and I. Walker, "Classification of clothing using interactive perception," in *Int. S. Robotics Research (ISRR)*, 2011, pp. 1–7.

[10] S. Tirumala, T. Weng, D. Seita, O. Kroemer, Z. Temel, and D. Held, "Learning to singulate layers of cloth using tactile feedback," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022, pp. 7773–7780.

[11] J. Qian, T. Weng, L. Zhang, B. Okorn, and D. Held, "Cloth region segmentation for robust grasp selection," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 9553–9560.

[12] J. Maitin-Shepard, M. Cusumano-Towner, J. Lei, and P. Abbeel, "Cloth grasp point detection basedon multiple-view geometric cues with application to robotic towel folding," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2010.

[13] Y. Deng, C. Xia, X. Wang, and L. Chen, "Graph-transporter: A graph-based learning method for goal-conditioned deformable object rearranging task," in *2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2022, pp. 1910–1916.

[14] I. Lenz, H. Lee, and A. Saxena, "Deep learning for detecting robotic grasps," in *Robotics: Science and Systems (RSS)*, 2013.

[15] F.-J. Chu, R. Xu, and P. A. Vela, "Real-world multi-object, multi-grasp detection," in *IEEE Robotics and Automation Letters*, 2018.

[16] A. Ramisa, G. Alenyà, F. Moreno-Noguer, and C. Torras, "Determining where to grasp cloth using depth information," in *International Conference of the Catalan Association for Artificial Intelligence*, 2011.

[17] A. Ramisa, G. Aleny`, F. Moreno-Noguer, and C. Torras, "Using depth and appearance features for informed robot grasping of highly wrinkled clothes," in *IEEE International Conference on Robotics and Automation*, 2012, pp. 1–6.

[18] X. Wang, X. Jiang, J. Zhao, S. Wang, and Y.-H. Liu, "Picking towels in point clouds," vol. 8, 2020, pp. 129 338–129 346.

[19] W. C. Agboh, J. Ichnowski, K. Goldberg, and M. R. Dogar, "Multi-object grasping in the plane," in *International Symposium on Robotics Research (ISRR)*, 2022.

[20] T. Yamada and H. Yamamoto, "Static grasp stability analysis of multiple spatial objects," *Journal of Control Science and Engineering*, vol. 3, pp. 118–139, 2015.

[21] W. C. Agboh *et al.*, *Learning to efficiently plan robust frictional multi-object grasps*, 2022.

[22] S. Chen and Y. Zhu, "Grasping objects in clutter with deep learning and grasping affordance prediction," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2021.

[23] P. V. Nguyen, P. N. Nguyen, T. Nguyen, and T. L. Le, "Hybrid robot hand for stably manipulating one group objects," *Archive of Mechanical Engineering*, vol. vol. 69, no. No 3, pp. 375–391, 2022.

[24] H. Kasaei, M. Kasaei, G. Tziafas, S. Luo, and R. Sasso, "Simultaneous multi-view object recognition and grasping in open-ended domains," 2021, pp. 8295–8300.

[25] X. Wang, X. Jiang, J. Zhao, S. Wang, and Y.-H. Liu, "Grasping objects mixed with towels," *IEEE Access*, vol. 8, pp. 129 338–129 346, 2020.

[26] H. Huang *et al.*, "Mechanical search on shelves with efficient stacking and destacking of objects," in *Int. S. Robotics Research (ISRR)*, 2022, pp. 1–16.

[27] T. Sakamoto, W. Wan, T. Nishi, and K. Harada, "Efficient picking by considering simultaneous two-object grasping," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021.

[28] K. M. Varadarajan and M. Vincze, "Object part segmentation and classification in range images for grasping," in *2011 15th International Conference on Advanced Robotics (ICAR)*, 2011, pp. 21–27.