

# MOE-Touch More Deformation: Shape-Based Soft Robotic Contact Estimation for Manipulation

Anonymous Author(s)

Affiliation

Address

email

1           **Abstract:** Contact-rich interaction with the world is crucial for many common  
2 manipulation tasks such as handling delicate objects or providing physical assistance.  
3 Unlike commonly used rigid manipulators, soft robotic manipulators can  
4 interact safely and robustly with large distributed contact with the world. How-  
5 ever, contact sensing for soft robots has been challenging because embedding sen-  
6 sors into soft bodies introduces rigidity, which undercuts the benefits of such com-  
7 pliant systems. In this paper, we present MOE-Touch, a method for reasoning  
8 about contact conditions on the soft robot by observing deformation. We intro-  
9 duce and test the idea that contact conditions and contact object geometry can be  
10 inferred by observing contact deformations in a compliant and soft robot manip-  
11 ulator. We propose Multi-finger Omnidirectional End-effector (MOE), a soft ma-  
12 nipulator capable of safely interacting with delicate surfaces. We use a mesh en-  
13 ergy optimization-based method to estimate the shape of MOE during interaction  
14 in occluded assistive dressing environments. We then use a Graph Neural Network  
15 (GNN)-based contact estimation module to predict distributed contact locations  
16 from deformation. MOE-Touch can accurately estimate contact with 3.03 mm  
17 Chamfer distance error, which is a 50.65 % improvement on the baseline. We then  
18 demonstrate an application of MOE-Touch shape estimation and contact localiza-  
19 tion modules for the reconstruction of an occluded surface modeled as Gaussian  
20 Process Implicit Surfaces (GPIS) with averaged errors of 3.62 mm, and showcase  
21 the application of using MOE-Touch for grasping a piece of paper on a flat sur-  
22 face with an unknown orientation.

23           **Keywords:** Soft Robotics, Contact Estimation, Manipulation

## 24 1 Introduction

25           We often make large distributed contact with objects in our daily lives. Such distributed contact-rich  
26 interaction with the world can serve two purposes. First, it allows us to perceive occluded surfaces  
27 and understand the underlying geometry. For example, a hairstylist can pat a customer’s head to  
28 estimate the contour of the scalp underneath the voluminous hair and select feasible hairstyles.  
29 Second, manipulating certain objects unavoidably results in large contact. We can consider the  
30 example of picking up a piece of paper from a flat table, which we often accomplish by laying finger  
31 pads on top of the paper and bending the paper into the hand. In either case, our ability to perceive  
32 and reason about contact with the world is crucial [1].

33           Common rigid robotic manipulators often cannot safely make large distributed contact without risk-  
34 ing damage to the fragile hardware or the environment. Given safety concerns, most prior work re-  
35 lies on using costly or specialized sensors to avoid applying unsafe contact forces [2], and explicitly  
36 avoiding direct contact during interaction [3]. Contact avoidance is especially common in assistive  
37 robotic works to ensure a user’s safety from rigid robots [4]. However, such constraints can produce  
38 overly conservative assistance that may be too slow and uncomfortable for human users [5].

39 Rather than avoiding contact, we target contact-rich manipulation scenarios for robots to embrace  
40 contacts safely to provide better assistance. To this end, soft robot manipulators offer unique advan-  
41 tages compared to rigid end effectors. The inherent compliance of soft robot manipulators [2] en-  
42 ables robust control and mechanically aids in safe real-world operation [6]. This is especially rele-  
43 vant for delicate manipulation [7] and human-robot interaction [8]. The ability to deform with con-  
44 tact also makes them safer than rigid manipulators, applying significantly less force on contacting  
45 objects during collision [9]. However, embedding contact and tactile sensors into such soft manip-  
46 ulators is an open challenge. Most previously proposed tactile sensors are either at least partially  
47 rigid [10] or limit strain [11], undermining soft robots’ advantages. The lack of effective and de-  
48 ployable contact estimation solutions for soft robots is a bottleneck to developing adaptable and in-  
49 telligent soft robotic manipulators [12].

50 Toward addressing contact sensing for soft robotic manipulators, we present MOE-Touch, a method  
51 for reconstructing a deformed soft robot shape and estimating its contact conditions for contact-rich  
52 soft robotic manipulation. MOE-Touch tracks the movement of keypoints on a soft robot manipula-  
53 tor and reconstructs watertight surface meshes of the deforming soft robot manipulator using a mesh  
54 energy-minimization method based on As-Rigid-As-Possible (ARAP) principles [13]. We show  
55 that this keypoint mesh optimization-based shape estimation method produces robust, high-fidelity  
56 shape reconstructions, providing more 3D shape structure compared to end-to-end learning-based  
57 approaches [14]. MOE-Touch then uses the observed deformations of the soft robot manipulator to  
58 predict points over the mesh that are in contact with other object surfaces. We evaluate MOE-Touch  
59 with two tasks. First, to reconstruct occluded surfaces during assistive-care manipulation tasks, we  
60 update a modified formulation of a Gaussian Process Implicit Surface (GPIS) [15, 16] model with  
61 the predicted contact conditions. We also demonstrate MOE-Touch in novel grasping tasks with 2D  
62 deformable objects such as paper on a flat surface, where we use MOE-Touch to predict the relative  
63 orientation of the surface to enable successful grasps.

64 With MOE-Touch, we introduce the idea of reasoning about contact conditions and contacting object  
65 geometry from observed deformations of a soft robotic manipulator, which is a unique advantage of  
66 soft robots. In summary, this paper makes three contributions. First, it proposes MOE, a multi-finger  
67 dexterous soft robotic manipulator. Second, we demonstrate how contact conditions can be inferred  
68 by perceiving deformations in MOE as it interacts with other object surfaces. Third, we demonstrate  
69 the use cases of MOE-Touch contact estimation methods for occluded surface reconstruction and  
70 flat deformable object grasping tasks.

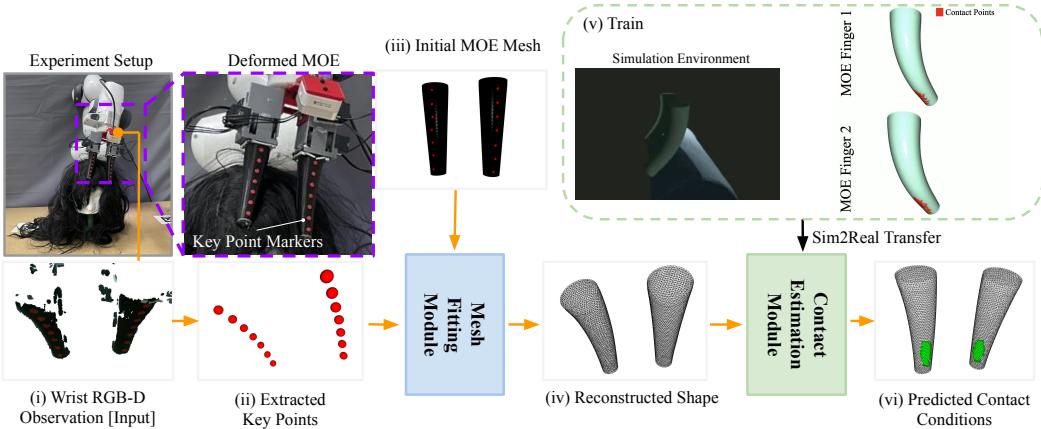
## 71 **2 Related Work**

### 72 **2.1 Soft Robotic Manipulators**

73 Soft robotic manipulators are typically characterized by their deformable and compliant constituent  
74 material [12]. They are becoming increasingly popular because of their ability to interact safely with  
75 delicate objects and environments [7]. A spectrum of soft robotic manipulators exists from partially  
76 rigid or functionally rigid-linked soft robotic manipulators [17, 18] to fully soft robotic manipulators  
77 that bend continuously [19]. Recent works have started to demonstrate the “mechanical intelligence”  
78 of fully soft robotic manipulators, where their continuous deformation behavior contributes to the  
79 robustness and dexterity [20]. We primarily focus on such fully soft robotic manipulators in this  
80 work and explore their unique advantages in the domain of perception for contact-rich manipulation.

### 81 **2.2 Soft Robotic Sensing**

82 The compliance and deformation of soft robot manipulators [2] pose a challenge for perception and  
83 sensing [12] to determine the manipulator’s proprioceptive state. Soft robot proprioceptive sensing  
84 and shape representations must capture complex deformation patterns of the soft robot [14, 21]. Con-  
85 ventionally, the shape of soft robots has been represented by parameterized 2-dimensional curves,  
86 which reduces the state estimation problem by modeling more tractable, low degrees-of-freedom



**Figure 1: Overview diagram of the MOE-Touch method to estimate contact conditions of soft robotic manipulators from deformation** (i) We capture RGB-D observations from the wrist camera during the interaction. (ii) We extract keypoints on each finger of MOE. (iii) Using the initial MOE mesh and extracted keypoints, we fit a mesh to the deforming state of MOE. (iv) We reconstruct MOE’s deforming surface geometry. (v) We collect a large dataset of MOE deformation and corresponding contact conditions for training a GNN over the MOE surface mesh from simulation. (vi) Using the predicted contact conditions for an interaction trajectory, we reconstruct the contacting surface under occlusions.

systems. The most compact state representation uses a single degree-of-freedom curve with constant curvature, defined by its bending radius [22, 23]. More expressive representations construct multiple geometric primitives such as piecewise constant curvature models [24], multiple rigid frames [25], or rigid links [26]. These primitive representations have been used for dynamic control of soft robot manipulators [27]. However, these representations fail to capture volumetric information, and more deformation behaviors such as distributed, contact-based deformations [14]. Some methods have been proposed to capture rich soft robot states using point clouds [28, 14], but they rely on learning a state estimation model to reconstruct shapes by training on large training datasets. Previous works have proposed both explicit representations such as meshes [21] and implicit representations such as neural Signed Distance Functions (SDFs) for soft bodies [29]. Explicit representations are particularly convenient for this work because we can directly leverage the reconstructed body’s nodes and their correspondences for downstream tasks such as contact surface reconstruction. Recent work grounds shape reconstruction with mechanics-based priors, which yields more data efficiency and stable proprioceptive state estimation [21]. In this paper, we show how these methods can be extended beyond manipulator-only proprioceptive state estimation to object interaction and robotic manipulation.

### 103 2.3 Robotic Tactile Sensing

104 Tactile sensors observe local deformations against a rigid reference surface to infer contact points, 105 maps, or forces [10, 30]. Specialized tactile sensors such as GelSight [31] and Digit [10] can be attached to infer contacts, as shown by Suresh et al. [32]. Although such sensors can provide high- 106 fidelity tactile and texture information about the contacting surface, they require contacts to occur 107 on the small sensorized contact region, which constrains sensor configuration when used in robot 108 manipulators [32]. Furthermore, tactile sensors tend to be difficult to embed into soft robots without 109 introducing undesired rigidity [33]. Prior work demonstrated that entirely compliant soft robot 110 manipulators deform significantly with contact [34, 21]. In this paper, we demonstrate that we can 111 reliably predict contact conditions from such observed deformations.

113 **3 Problem Statement**

114 In this work, we aim to estimate the deformed shape of a continuum soft robotic manipulator and its  
115 contact regions based on the estimated deformation. To this end, we can make assumptions that are  
116 afforded to us because of the unique features of fully soft robotic manipulators. We assume that the  
117 material property is largely homogeneous and known. We also assume that the soft robot's material  
118 is soft enough to deform with contact which we validated to be true in contact experiments.

119 The goal of the soft robot shape estimation in this work is to infer the overall mesh of the manipulator  
120 based on the sparse keypoint movements. We consider a soft robotic manipulator embodiment where  
121 the keypoints are tracked with visual markers attached to the soft fingers, although as with Yoo  
122 et al. [21], the keypoint movements could be indirectly tracked without external sensors or physical  
123 markers with a variety of sensors such as microphones. As such, the methods in this paper are  
124 relevant assuming the soft robot's sensors can lead to sufficiently reliable estimation of keypoints.  
125 We extend the previously proposed POE-M pipeline to multiple fingers [21]. The working principle  
126 of the POE-M pipeline relies on As-Rigid-As-Possible (ARAP) principles, which is framed as an  
127 energy minimization problem over mesh nodes [13]. Based on these high-fidelity soft robot mesh  
128 shape reconstructions, we seek to infer contact points on the mesh.

129 **4 Method**

130 **4.1 MOE Design**

131 We design a soft tendon-driven manipulator which we call  
132 Multi-finger Omnidirectional End-effector (MOE), building  
133 on a single-finger tendon-driven soft robot [21]. The  
134 design is largely modular, where each of the fingers is an  
135 independent subsystem that can be detached and assem-  
136 bled to get multiple-finger configurations. In this work,  
137 we present results for a MOE with two fingers, as shown  
138 in Figure 2, and three fingers for the paper grasping task.  
139 Each of MOE's soft fingers is molded from silicone with  
140 low hardness. Each finger has four embedded tendons,  
141 which are actuated by two servo motors. Each pair of ten-  
142 dons actuated by a single servo motor controls MOE finger's range of motion in a bending plane.  
143 We include an RGB-D camera on the wrist of MOE to provide egocentric-view depth as shown in  
144 Figure 1. Red markers are placed on the surfaces of the MOE fingers for the RGB-D camera to track  
145 MOE keypoints as the body deforms.

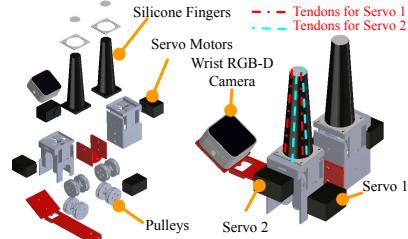


Figure 2: Design of MOE

146 **4.2 MOE Shape Estimation**

147 To guide the shape estimation of MOE, we track the 7 red keypoint markers placed on the surface of  
148 each MOE finger, as shown in Figure 1. We segment the markers using color thresholds, and apply  
149 DBSCAN [35] to cluster the 3D points, localizing marker centers based on the point densities. To  
150 account for the occlusion of intermittent points or missing points, we check that all seven markers  
151 are detected and discard frames with missing markers. In the initial frame, we find the nodes on  
152 the initial mesh closest to the keypoints and use them as handle points. From the initialization  
153 phase, we account for the movement of each of the keypoints frame-to-frame. In practice, some  
154 of the keypoints may become occluded, for instance due to hair getting in the way. To account for  
155 this, we remove the occluded keypoints from consideration in the ARAP mesh fitting phase. Visual  
156 occlusions are specific problems of the sensing modality that we chose in this work and as noted in  
157 Section 3, we may use other sensing modalities such as audio [21] to estimate keypoints without any  
158 reformulation of the presented methods.

159 To deform the mesh based on the keypoint movements, we define the source surface mesh  $S =$   
 160  $(E, V)$  and the deformed mesh  $S'$ , where a surface mesh is defined by edges  $e_{i,j} \in E$  composed  
 161 from vertices  $i, j \in V$ . As previously proposed [36, 21], we include a penalty on the rotations of the  
 162 neighboring edges  $e_l \in N(e_k)$  to produce mesh updates that are physically admissible. The energy  
 163 to minimize is

$$E_{\text{smoothed}}(S, S') = \min_{R_1, \dots, R_m} \sum_{k=1}^m \left( \sum_{i,j \in e_k} c_{ijk} \|e_{ij} - R_k e_{ij}\|^2 + \lambda \hat{A} \sum_{e_l \in N(e_k)} w_{kl} \|R_k - R_l\|^2 \right), \quad (1)$$

164 where  $c_{ijk}$  are the cotan weights [37],  $\lambda$  is the regularization weight,  $R_1, \dots, R_m \in SO(3)$  are the  
 165 local rotations for each of the edges  $e_k \in E$  where  $m = |E|$ ,  $\hat{A}$  is the triangle area and  $w_{kl}$  are  
 166 the scalar weight terms defined by the cotan weights of the dual mesh of  $e_{kl}$  [37]. We iteratively  
 167 minimize  $E_{\text{smoothed}}(S, S')$  with local-global optimizer as outlined in Levi and Gotsman [36].

168 To reconstruct the full mesh shape of MOE, we treat vertices corresponding to the keypoints  $p_{1,\dots,k}$   
 169 as being constrained to the new positions, based on the predicted keypoint positions. The rest of the  
 170 mesh vertex positions are moved to minimize  $E_{\text{smoothed}}(S, S')$ .

### 171 4.3 Contact Estimation

172 Contact points are difficult to obtain directly from the real world due to occlusions from the contacting  
 173 object. Previous works have demonstrated the capabilities of soft-body simulation to generate  
 174 training datasets of deformed shapes and contact information [38, 14]. We extend these works to  
 175 generate a dataset of fully simulated MOE deformations and contact points.

176 To sample from varying contact normals and surface orientations, we import objects from the YCB  
 177 Object and Model Set [39] into a SOFA simulation environment [40]. We also generate and import  
 178 tendon-actuated meshes of MOE. We randomize the selected contacting object's orientation and  
 179 position with respect to MOE's trajectory, to simulate various contact locations and orientations.  
 180 From these simulated trials, we generate a dataset of 31,722 meshes and corresponding contact  
 181 points, which were recorded as the indices of the MOE mesh nodes in contact with an object.

182 For MOE-Touch, we aim to use shape estimation of MOE's deforming state to infer its contact  
 183 conditions. Highly expressive graph neural network architectures have been shown to learn and  
 184 reason about complex physical interactions [41, 42]. We train a variant of Dynamic Graph CNN  
 185 (DGCNN) [41] on the simulated dataset, where the inputs are MOE point clouds labeled with con-  
 186 tact obtained from the simulation environment. To account for the imbalance in the dataset, where  
 187 there are noticeably more points not in contact than the ones in contact, we use a weighted softmax  
 188 cross entropy loss function. We deploy the trained contact estimation model directly on real-world  
 189 predictions of MOE mesh shapes to predict the contacting nodes as MOE deforms during an inter-  
 190 action trajectory, as shown in Figure 1.

### 191 4.4 Contacting Surface Reconstruction

192 MOE-Touch's shape estimation and contact estimation modules provide contact information. A  
 193 useful application of a soft robotic manipulator is safely interacting with an occluded surface, such  
 194 as the scalp under hair or arm under a hospital gown, and using the contact estimates to reconstruct  
 195 them. In such tasks, we have useful priors on the occluded body part's geometry. We use a task-  
 196 dependent prior mesh specific to the domain. For the initial task of reconstructing a human head,  
 197 we use an open-sourced canonical head 3D mesh [43] and trained a Gaussian Process (GP) to learn  
 198 a prior over the SDF of the mesh. Given a set of dense grid points  $\mathbf{X}$  and corresponding SDF values  
 199  $\mathbf{Y}$ , the GP model is

$$f(\mathbf{x}) \sim \mathcal{GP} \left( c, \sigma^2 \exp \left( -\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2l^2} \right) \right), \quad (2)$$

200 where  $\sigma^2$  is the variance and  $l$  is the length scale of the Radial Basis Function (RBF) kernel, with  
 201 the observation model

$$y = f(\mathbf{x}) + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma_n^2) \quad (3)$$

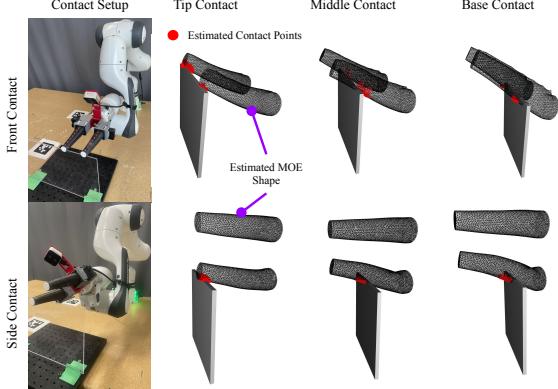


Figure 3: Controlled thin plate contact estimation experiment to demonstrate that MOE-Touch is sensitive in a large portion of the MOE soft robot.

Method	Contact Object	Avg. CD ↓
Class. [Baseline]	Plate (Tip)	3.41
MOE-Touch	Plate (Tip)	<b>3.03</b>
Class. [Baseline]	Plate (Middle)	13.48
MOE-Touch	Plate (Middle)	<b>7.08</b>
Class. [Baseline]	Plate (Base)	20.10
MOE-Touch	Plate (Base)	<b>9.92</b>
Class. [Baseline]	Head (Bald)	7.76
MOE-Touch	Head (Bald)	<b>6.58</b>
Class. [Baseline]	Head (Wig)	13.73
MOE-Touch	Head (Wig)	<b>12.22</b>
Class. [Baseline]	Arm (Gown)	6.88
MOE-Touch	Arm (Gown)	<b>6.24</b>

Table 1: Quantitative comparisons of MOE-Touch to a classification-based sparse contact estimation baseline [46] both for a controlled experiment setting and task-relevant settings.

202 where  $n$  is the number of training points. The training objective maximizes the marginal log-  
203 likelihood loss

$$\log p(\mathbf{Y}|\mathbf{X}) = -\frac{1}{2}\mathbf{Y}^\top(\mathbf{K} + \sigma_n^2\mathbf{I})^{-1}\mathbf{Y} - \frac{1}{2}\log|\mathbf{K} + \sigma_n^2\mathbf{I}| - \frac{n}{2}\log 2\pi, \quad (4)$$

204 where  $\mathbf{K}$  denotes the covariance matrix constructed using an RBF kernel over the training inputs in  
205  $\mathbf{X}$ . Once we have a trained prior, we fine-tune the GP with the contact-point information to obtain  
206 the posterior SDF. Finally, we reconstructed the head mesh using Poisson Surface Reconstruction  
207 (PSR) [44] on a point cloud obtained by running the Marching Cubes Algorithm (MCA) [45] over  
208 the zero-level set of the SDF.

## 209 5 Experiments

### 210 5.1 Contact Estimation

211 As a baseline for all contact estimation experiments, we use a method inspired by Zöller et al.  
212 [46] which addressed the most similar problem to the one outlined in this work. In their original  
213 formulation, they fit a k-nearest neighbor (KNN) model to classify which discrete region of the  
214 finger is in contact with the robot based on acoustic signals [46]. Inspired by their approach, we fit a  
215 KNN to a randomly selected subset of the simulation dataset to classify a contact as tip, middle, or  
216 base of MOE based on the keypoint positions. For evaluation, we use the centroid of the classified  
217 MOE finger segment as the single contact point from each contact condition. We use MOE-Touch’s  
218 shape estimation pipeline to map the contact segment classification to a point that can be evaluated.  
219 We also note that a significant difference with such approaches is that they produce only a single  
220 contact point, whereas MOE-Touch produces multiple contact points distributed over a small region  
221 on the finger.

222 We demonstrate that MOE-Touch can estimate contact conditions accurately in varying contacting  
223 conditions with controlled contact on a thin plate (see Figure 3). Table 1 (top rows) reports the  
224 quantitative contact estimation results with comparisons to the baseline. We note that the contact  
225 estimation is accurate with  $< 10$  mm unidirectional Chamfer Distance (CD) with notably higher error  
226 at the base. The performance is likely worse near the base of the MOE because the robot is less  
227 compliant and deforms less, making it more difficult for the model to disambiguate possible contact  
228 conditions. In all contact conditions, MOE-Touch performs better than the baseline, most notably at  
229 the base with 50.65 % reduction in CD error.

230 We also test the contact estimation module on accurate models of the head and arm. Both envi-  
231 ronments are motivated by common contact-rich assistive robotic settings, where visual occlusion

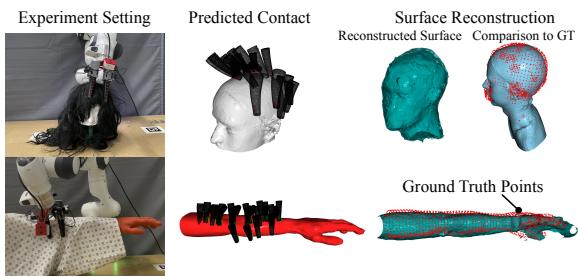


Figure 4: Surface reconstruction with MOE-Touch

Method	Contact Object	Avg. CD ↓
Non-Probabilistic	Head (Bald)	16.01
GP w/ Sphere	Head (Bald)	13.83
GP w/ Prior	Head (Bald)	<b>3.64</b>
Non-Probabilistic	Head (Wig)	16.05
GP w/ Sphere	Head (Wig)	14.41
GP w/ Prior	Head (Wig)	<b>3.62</b>
GP w/ Prior	Arm	<b>9.90 (4.82*)</b>

Table 2: Surface reconstruction results compared to the ground truth. \* denotes the result for the arm with the unsampled hand removed from evaluation.

232 may be common and unavoidable, requiring the robot to safely interact with the human subject.  
 233 For the head setup, we randomly selected a head mesh of an adult person from a craniofacial shape  
 234 dataset [43], 3D-print the meshed model, and test with and without a voluminous wig.

235 We then test 30 distinct MOE contact conditions on the head to evaluate shape and contact estimation  
 236 modules with and without a wig. We register the point clouds together from the wrist-mounted  
 237 RGB-D camera to show the contact coverage across the head in Figure 4. We then evaluate the shape  
 238 and contact estimation modules by registering the predicted contact points together, computing uni-  
 239 directional average CD from the contact points to the head ground-truth mesh nodes. We then per-  
 240 form a similar series of 15 contact trials on a model of an adult human arm occluded by a hospi-  
 241 tal gown (see Figure 4). Similar to the trials with the model head, we register the predicted contact  
 242 points, compared to the ground-truth mesh, and compute the CD metrics.

243 We observe that in all three settings, the MOE-Touch pipeline performs functionally well and im-  
 244 proves on the baseline method in all three cases with the lowest errors. The environments with the  
 245 bald head and arm both result in an average MOE-Touch CD error of around 6.5 mm. We can notice  
 246 a noticeably higher CD error of 12.22 mm in the environment with a head and a wig. A significant  
 247 portion of the error may come from the thickness that the wig’s inner hair net which is around 5mm  
 248 thick. Because we do not have a separate ground-truth mesh for the head with a wig, we still eval-  
 249 uate the metrics with the bald head mesh.

250 On a computer with a GeForce RTX 4090 GPU, the MOE-Touch shape estimation module outputs  
 251 a mesh with 2048 vertices from 50 iterations with a runtime of 49.55 ms, and the contact estimation  
 252 inference time runs on average 43.62 ms for each deformed shape. For comparison, a neural implicit  
 253 surface-based approach takes 2079 ms per scene to reconstruct the mesh and contact patch [29]. The  
 254 efficiency of MOE-Touch is largely a result of the methods that we develop around our domain-  
 255 specific assumptions for soft robotic perception, such as homogeneous material composition.

## 256 5.2 Contacting Surface Reconstruction

257 As outlined in Section 4.4, we implement a task-dependent prior as a GP for surface reconstruction  
 258 from real-world contact points. Results in table 2 show that our method reconstructs the mesh from  
 259 real-world contact points accurately with an average CD of 3.64 mm for the bald head, 3.62 mm for  
 260 the head with a wig, and 9.90 mm for an arm dressed in a hospital gown. The capability of the task-  
 261 dependent prior method to generate a watertight mesh after accommodating real-world data is shown  
 262 in Figure 4 for the arm and the head. For the evaluation of the arm, we reported errors for the entire  
 263 arm and for the arm with the hand removed since we did not sample from it during experiments.

264 We also present a baseline method based on some previous works that assume a primitive geometric  
 265 shape as initialization for interactive perception and mesh reconstruction [47, 16]. We use a spherical  
 266 prior as a naive method to obtain the posterior distribution over the real-world contact points. The  
 267 main point of failure in this method can be attributed to the fitted sphere mesh, with points that are  
 268 significantly out of distribution from an average human head.

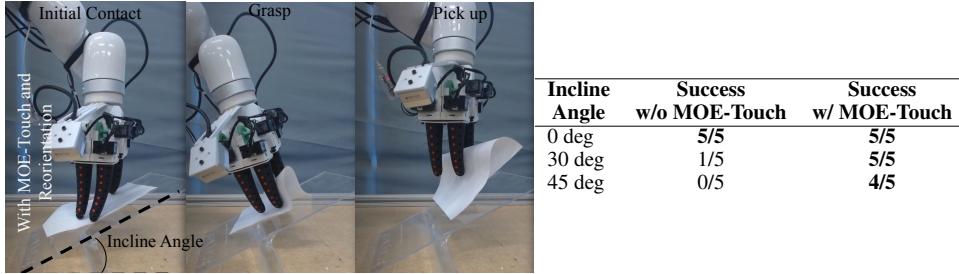


Figure 5: Application of MOE-Touch in flat deformable object grasping task on inclined surfaces.

269 We compare the GP-based method from Section 4.4 to using nearest neighbor mesh fitting with  
 270 smoothing, which results in a qualitatively worse formed surface compared to the GP method with  
 271 a spherical prior. This method performs the worst for surface reconstruction with average CD values  
 272 of 16.01 mm for bald head data, and 16.05 mm for head with a wig. The proposed task-dependent  
 273 prior-based surface reconstruction module performs better than the two baseline methods, resulting  
 274 in 73.68 % and 77.26 % reduced average CD metric error for the head without a wig compared to  
 275 using a spherical prior and the non-probabilistic method, respectively.

### 276 5.3 Paper Grasping

277 We evaluate MOE-Touch on the task of grasping paper on a surface with an initially unknown incline  
 278 angle. Additionally, to evaluate the modularity of MOE-Touch, we tested a variant of MOE with  
 279 three fingers. We prepared an inclined clear acrylic flat surface with a 190×130 mm common printer  
 280 paper on top. MOE made contact with the paper initially misaligned from the acrylic surface. We  
 281 used MOE-Touch to estimate contact points with the surface and fitted a plane to the points with  
 282 Random Sample Consensus (RANSAC). We then reoriented MOE to be normal to the surface and  
 283 grasped the paper (see Figure 5). We tested with 0, 30, and 45-degree incline of the surface. We  
 284 compared the success rates of the paper grasping task out of 5 trials for each setting against not  
 285 using MOE-Touch. Surprisingly, MOE could still grasp the paper at 30-degree incline once without  
 286 MOE-Touch, showing robustness of its compliance and mechanical intelligence. However, with 45-  
 287 degree inclines, MOE needed MOE-Touch to succeed in the task.

## 288 6 Conclusion

289 In this work, we introduce methods for contact estimation in contact-rich soft robotic manipulation.  
 290 We develop MOE, a modular Multi-finger Omnidirectional End-effector that can safely and robustly  
 291 interact with the world for contact-rich manipulation. We use a mesh energy optimization-based  
 292 method to estimate the shape of MOE in interaction with the environment. The proposed MOE-  
 293 Touch method takes an explicit mesh optimization-based approach to reconstruct the deformed shape  
 294 of the soft robot and reason about contact conditions with a GNN over the mesh. We show that  
 295 MOE-Touch can estimate occluded surface contact with an average distance error of 6.25 mm, im-  
 296 proving on the baseline by 17.53%. We show that the MOE-Touch can be deployed to reconstruct  
 297 an occluded surface with averaged errors of 3.62 mm. We then show the use case of MOE-Touch  
 298 for a manipulation task of grasping paper on arbitrarily inclined surfaces, where contact estimation  
 299 guides re-orientation of MOE to be normal to the contacting surface.

## 300 7 Limitations

301 One limitation of this work is that we trained our contact estimation module only with binary contact  
 302 labels. Extending MOE-Touch to estimate contact pressure may present advantages in downstream  
 303 manipulation tasks in the future. Additionally, just as human skins have four different mechanore-  
 304 ceptors responsible for different tactile stimuli [48], robotic tactile modalities offer different advan-  
 305 tages and multi-tactile modality sensor fusion may be a promising direction to augment MOE-Touch.

306 **References**

- 307 [1] Y. C. Nakamura, D. M. Troniak, A. Rodriguez, M. T. Mason, and N. S. Pollard. **The complexities of grasping in the wild**. In *2017 IEEE-RAS 17th International Conference on Humanoid Robotics (Humanoids)*, pages 233–240. IEEE, 2017.
- 310 [2] J. Hughes, U. Culha, F. Giardina, F. Guenther, A. Rosendo, and F. Iida. **Soft manipulators and grippers: A review**. *Frontiers in Robotics and AI*, 3:69, 2016.
- 312 [3] Y. Wang, Z. Sun, Z. Erickson, and D. Held. **One policy to dress them all: Learning to dress people with diverse poses and garments**. In *Robotics: Science and Systems XIX*, 2023.
- 314 [4] S. Li, N. Figueroa, A. Shah, and J. Shah. **Provably safe and efficient motion planning with uncertain human dynamics**. In *Robotics: Science and Systems XVII*, 2021.
- 316 [5] G. Canal, C. Torras, and G. Alenyà. **Are preferences useful for better assistance? a physically assistive robotics user study**. *ACM Transactions on Human-Robot Interaction (THRI)*, 10(4):1–19, 2021.
- 319 [6] N. Kuppuswamy, A. Alspach, A. Uttamchandani, S. Creasey, T. Ikeda, and R. Tedrake. **Soft-bubble grippers for robust and perceptive manipulation**. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 9917–9924. IEEE, 2020.
- 322 [7] N. R. Sinatra, C. B. Teeple, D. M. Vogt, K. K. Parker, D. F. Gruber, and R. J. Wood. **Ultragentle manipulation of delicate structures using a soft robotic gripper**. *Science Robotics*, 4(33):eaax5425, 2019.
- 325 [8] P. Polygerinos, N. Correll, S. A. Morin, B. Mosadegh, C. D. Onal, K. Petersen, M. Cianchetti, M. T. Tolley, and R. F. Shepherd. **Soft robotics: Review of fluid-driven intrinsically soft devices; manufacturing, sensing, control, and applications in human-robot interaction**. *Advanced Engineering Materials*, 19(12):1700016, 2017.
- 329 [9] U. Yoo, N. Dennler, M. Mataric, S. Nikolaidis, J. Oh, and J. Ichnowski. **MOE-Hair: Toward Soft and Compliant Contact-rich Hair Manipulation and Care**. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, pages 1163–1167, 2024.
- 332 [10] 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*, 5(3):3838–3845, 2020.
- 336 [11] R. Bhirangi, T. Hellebrekers, C. Majidi, and A. Gupta. **Reskin: versatile, replaceable, lasting tactile skins**. In *5th Annual Conference on Robot Learning*, 2021.
- 338 [12] H. Wang, M. Totaro, and L. Beccai. **Toward perceptive soft robots: Progress and challenges**. *Advanced Science*, 5(9):1800541, 2018.
- 340 [13] O. Sorkine and M. Alexa. **As-rigid-as-possible surface modeling**. In *Symposium on Geometry processing*, volume 4, pages 109–116. Citeseer, 2007.
- 342 [14] U. Yoo, H. Zhao, A. Altamirano, W. Yuan, and C. Feng. **Toward Zero-Shot Sim-to-Real Transfer Learning for Pneumatic Soft Robot 3D Proprioceptive Sensing**. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 544–551. IEEE, 2023.
- 345 [15] O. Williams and A. Fitzgibbon. **Gaussian process implicit surfaces**. In *Gaussian Processes in Practice*, 2006.
- 347 [16] S. Dragiev, M. Toussaint, and M. Gienger. **Gaussian process implicit surfaces for shape estimation and grasping**. In *2011 IEEE International Conference on Robotics and Automation*, pages 2845–2850. IEEE, 2011.

- 350 [17] W. Zhu, C. Lu, Q. Zheng, Z. Fang, H. Che, K. Tang, M. Zhu, S. Liu, and Z. Wang. A soft-  
 351 **rigid hybrid gripper with lateral compliance and dexterous in-hand manipulation.** *IEEE/ASME*  
 352 *Transactions on Mechatronics*, 28(1):104–115, 2022.
- 353 [18] P. Mannam, K. Shaw, D. Bauer, J. Oh, D. Pathak, and N. Pollard. **Designing anthropomorphic soft hands through interaction.** In *2023 IEEE-RAS 22nd International Conference on Humanoid Robots (Humanoids)*, pages 1–8. IEEE, 2023.
- 356 [19] S. Puhlmann, J. Harris, and O. Brock. **RBO hand 3: A platform for soft dexterous manipulation.** *IEEE Transactions on Robotics*, 38(6):3434–3449, 2022.
- 358 [20] A. Bhatt, A. Sieler, S. Puhlmann, and O. Brock. **Surprisingly robust in-hand manipulation: An empirical study.** In *Robotics: Science and Systems XVIII*, 2022.
- 360 [21] U. Yoo, Z. Lopez, J. Ichnowski, and J. Oh. **POE: Acoustic soft robotic proprioception for omnidirectional end-effectors.** *arXiv preprint arXiv:2401.09382*, 2024.
- 362 [22] C. Della Santina, R. K. Katzschmann, A. Biechi, and D. Rus. **Dynamic control of soft robots interacting with the environment.** In *2018 IEEE International Conference on Soft Robotics (RoboSoft)*, pages 46–53. IEEE, 2018.
- 365 [23] U. Yoo, Y. Liu, A. D. Deshpande, and F. Alamabeigi. **Analytical design of a pneumatic elastomer robot with deterministically adjusted stiffness.** *IEEE Robotics and Automation Letters*, 6(4):7773–7780, 2021.
- 368 [24] C. Della Santina, A. Bicchi, and D. Rus. **On an improved state parametrization for soft robots with piecewise constant curvature and its use in model based control.** *IEEE Robotics and Automation Letters*, 5(2):1001–1008, 2020.
- 371 [25] Y. Liu, U. Yoo, S. Ha, S. F. Atashzar, and F. Alambeigi. **Influence of antagonistic tensions on distributed friction forces of multisegment tendon-driven continuum manipulators with irregular geometry.** *IEEE/ASME Transactions on Mechatronics*, 27(5):2418–2428, 2021.
- 374 [26] Y. She, S. Q. Liu, P. Yu, and E. Adelson. **Exoskeleton-covered soft finger with vision-based proprioception and tactile sensing.** In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 10075–10081. IEEE, 2020.
- 377 [27] D. A. Haggerty, M. J. Banks, E. Kamenar, A. B. Cao, P. C. Curtis, I. Mezić, and E. W. Hawkes. **Control of soft robots with inertial dynamics.** *Science Robotics*, 8(81):eadd6864, 2023.
- 379 [28] R. Wang, S. Wang, S. Du, E. Xiao, W. Yuan, and C. Feng. **Real-time soft body 3d proprioception via deep vision-based sensing.** *IEEE Robotics and Automation Letters*, 5(2):3382–3389, 2020.
- 382 [29] M. J. Van der Merwe, Y. Wi, D. Berenson, and N. Fazeli. **Integrated object deformation and contact patch estimation from visuo-tactile feedback.** In *Robotics: Science and Systems XIX*, 2023.
- 385 [30] J. A. Collins, C. Houff, P. Grady, and C. C. Kemp. **Visual contact pressure estimation for grippers in the wild.** In *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 10947–10954. IEEE, 2023.
- 388 [31] W. Yuan, S. Dong, and E. H. Adelson. **Gelsight: High-resolution robot tactile sensors for estimating geometry and force.** *Sensors*, 17(12):2762, 2017.
- 390 [32] S. Suresh, H. Qi, T. Wu, T. Fan, L. Pineda, M. Lambeta, J. Malik, M. Kalakrishnan, R. Calan-  
 391 dra, M. Kaess, et al. **Neural feels with neural fields: Visuo-tactile perception for in-hand ma-  
 392 nipulation.** *arXiv preprint arXiv:2312.13469*, 2023.

- 393 [33] H. Kim, O. C. Kara, and F. Alambeigi. **A Soft and Inflatable Vision-Based Tactile Sensor for**  
 394 **Inspection of Constrained and Confined Spaces.** *IEEE Sensors Journal*, 2023.
- 395 [34] A. Sieler and O. Brock. **Dexterous soft hands linearize feedback-control for in-hand manipu-**  
 396 **lation.** In *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*,  
 397 pages 8757–8764. IEEE, 2023.
- 398 [35] E. Schubert, J. Sander, M. Ester, H. P. Kriegel, and X. Xu. **DBSCAN revisited, revisited: why**  
 399 **and how you should (still) use DBSCAN.** *ACM Transactions on Database Systems (TODS)*,  
 400 42(3):1–21, 2017.
- 401 [36] Z. Levi and C. Gotsman. **Smooth rotation enhanced as-rigid-as-possible mesh animation.** *IEEE*  
 402 *transactions on visualization and computer graphics*, 21(2):264–277, 2014.
- 403 [37] K. Crane, F. de Goes, M. Desbrun, and P. Schröder. **Digital geometry processing with discrete**  
 404 **exterior calculus.** In *ACM SIGGRAPH 2013 courses*, SIGGRAPH ’13, 2013.
- 405 [38] A. Sipos and N. Fazeli. **Simultaneous contact location and object pose estimation using pro-**  
 406 **prioception and tactile feedback.** In *2022 IEEE/RSJ International Conference on Intelligent*  
 407 *Robots and Systems (IROS)*, pages 3233–3240. IEEE, 2022.
- 408 [39] B. Calli, A. Walsman, A. Singh, S. Srinivasa, P. Abbeel, and A. M. Dollar. **Benchmarking in**  
 409 **manipulation research: Using the Yale-CMU-Berkeley object and model set.** *IEEE Robotics*  
 410 *& Automation Magazine*, 22(3):36–52, 2015.
- 411 [40] J. Allard, S. Cotin, F. Faure, P.-J. Bensoussan, F. Poyer, C. Duriez, H. Delingette, and  
 412 L. Grisoni. **Sofa—an open source framework for medical simulation.** In *MMVR 15-Medicine*  
 413 *Meets Virtual Reality*, volume 125, pages 13–18. IOP Press, 2007.
- 414 [41] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon. **Dynamic graph**  
 415 **cnn for learning on point clouds.** *ACM Transactions on Graphics*, 38(5):1–12, 2019.
- 416 [42] T. H. E. Tse, Z. Zhang, K. I. Kim, A. Leonardis, F. Zheng, and H. J. Chang. **S<sup>2</sup>Contact:**  
 417 **Graph-based network for 3d hand-object contact estimation with semi-supervised learning.** In  
 418 *European Conference on Computer Vision*, pages 568–584. Springer, 2022.
- 419 [43] H. Dai, N. Pears, W. Smith, and C. Duncan. **Statistical modeling of craniofacial shape and**  
 420 **texture.** *International Journal of Computer Vision*, 128(2):547–571, 2020.
- 421 [44] M. Kazhdan, M. Bolitho, and H. Hoppe. **Poisson surface reconstruction.** In *Proceedings of the*  
 422 *fourth Eurographics symposium on Geometry processing*, volume 7, 2006.
- 423 [45] W. E. Lorensen and H. E. Cline. **Marching cubes: A high resolution 3D surface construction**  
 424 **algorithm.** In *Seminal graphics: pioneering efforts that shaped the field*, pages 347–353. 1998.
- 425 [46] G. Zöller, V. Wall, and O. Brock. **Active acoustic contact sensing for soft pneumatic actuators.**  
 426 In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 7966–  
 427 7972. IEEE, 2020.
- 428 [47] S. Suresh, Z. Si, J. G. Mangelson, W. Yuan, and M. Kaess. **ShapeMap 3-D: Efficient shape**  
 429 **mapping through dense touch and vision.** In *2022 International Conference on Robotics and*  
 430 *Automation (ICRA)*, pages 7073–7080. IEEE, 2022.
- 431 [48] J. M. Loomis and S. J. Lederman. **Tactile perception.** *Handbook of perception and human*  
 432 *performances*, 2(2):2, 1986.
- 433 [49] D. C. Liu and J. Nocedal. **On the limited memory BFGS method for large scale optimization.**  
 434 *Mathematical programming*, 45(1):503–528, 1989.

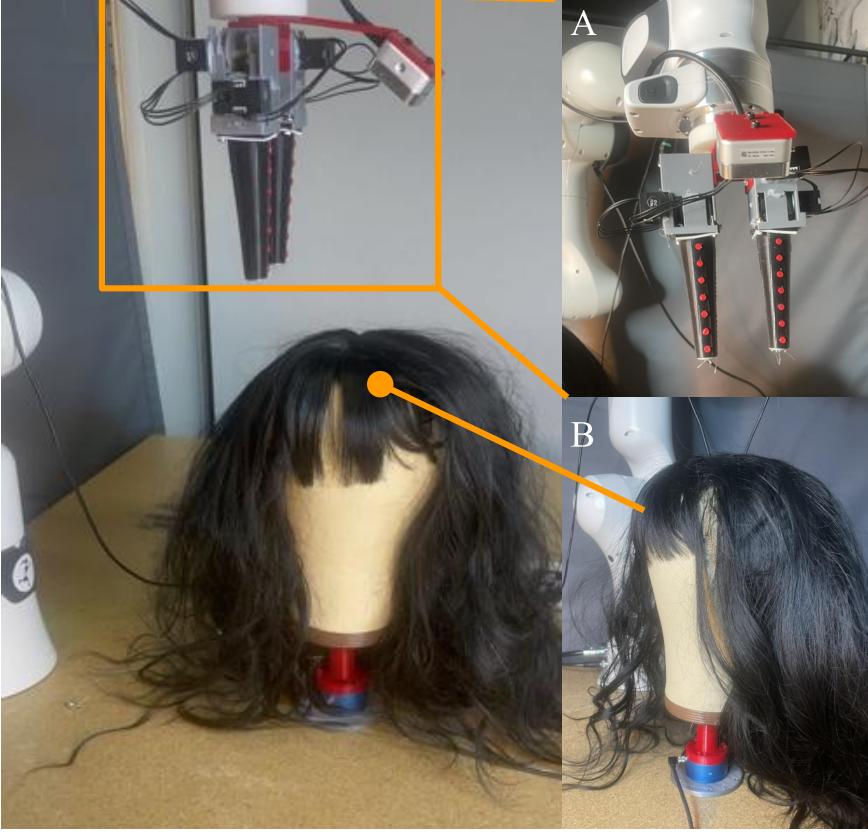


Figure 6: Experimental setup for evaluating the proposed MOE in interaction with a force-sensorized mannequin head. A: Two-fingered MOE soft manipulator with an RGBD camera. B: Mannequin head with a wig and 6-axis force sensor at its base.

Table 3: Hair Grasping Evaluation

End-effector	Depth [mm]	Performance Metrics	
		Max Force [N] ↓	Grasped Hair [mm] ↑
Rigid	2.0	1.11	4.0
	4.0	3.38	20.0
	6.0	7.67	25.0
MOE	2.0	<b>1.09</b>	<b>5.0</b>
	4.0	<b>1.38</b>	<b>18.7</b>
	6.0	<b>1.98</b>	<b>22.5</b>

## 435 A MOE Interaction Forces

436 We hypothesized that soft robotic manipulators would be safer and more comfortable for the human  
437 subject in hair manipulation and close-contact tasks.

438 Toward evaluating the hypothesis, we compared the forces experienced by the force-sensorized mannequin head with open-loop experiments, where a rigid parallel jaw gripper (FE Gripper, Franka  
439 Robotics) and the proposed MOE moved to a specified depth (2.0 mm, 4.0 mm, 6.0 mm) into the  
440 hair to grasp. The depths are measured with respect to the position where the robot is barely making  
441 contact with the hair to account for different lengths of the end-effectors. As the robot followed  
442 specified trajectories, we measured forces at the mannequin head base. After the grippers grasped  
443 the hair, the robot hand moved up to lift the grasped bundle of hair. We then measured the minimum  
444 packing perimeter of the bundle of hair. Figure 7 shows a sample result and the experimental proce-  
445 dure. Figure 9 shows the forces and torques experienced by the force-sensorized mannequin head.  
446

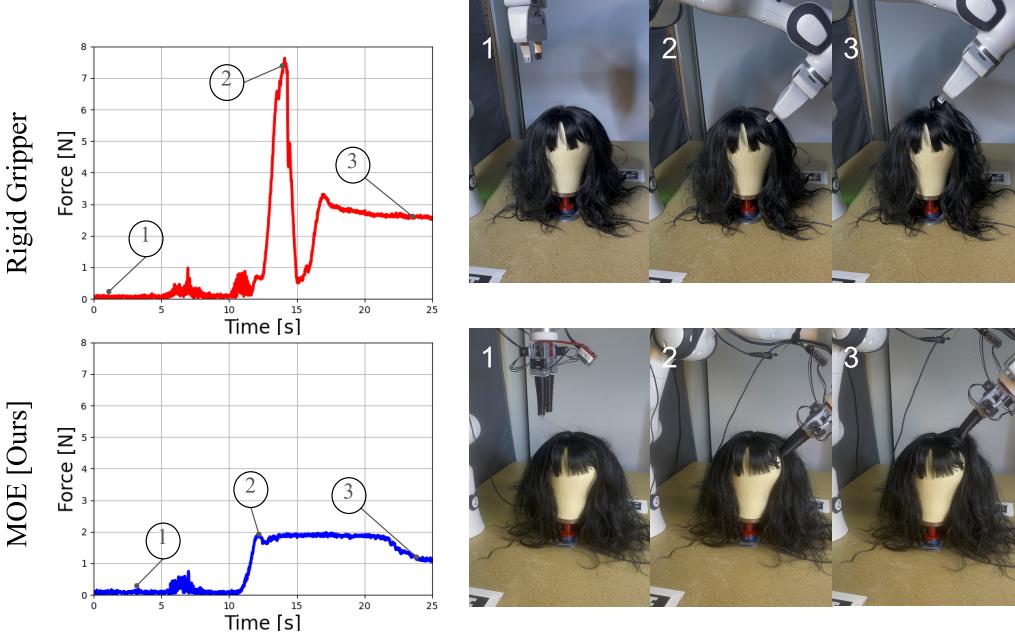


Figure 7: Hair grasping evaluation task experimental procedure and sample result at 6.0 mm depth. Top: experienced net forces and key frame images of the experiment with a baseline rigid gripper. Bottom: experienced net forces and key frame images of the experiment with the proposed MOE.

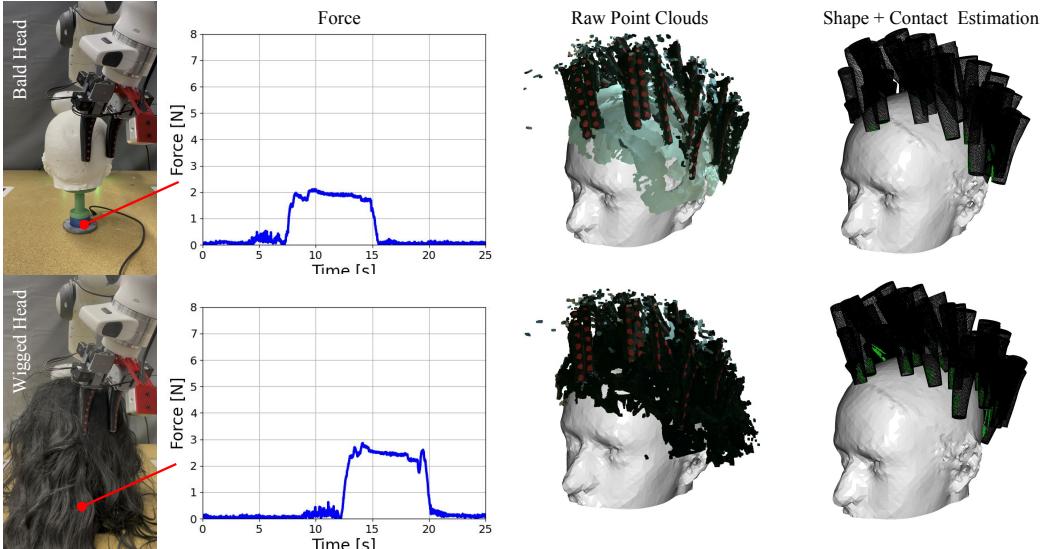


Figure 8: **Contact estimation experiment results.** The experimental setup for the head contact estimation experiments where a 3D-printed head is mounted on a force-torque sensor. The net force readings are plotted, showing the interaction forces experienced by the head. We visualize the registered wrist-mounted RGB-D camera point clouds from the 30 contact conditions, as well as the predicted MOE shape and contact points on the head. We show results for the head with and without a wig.

447 Lower forces and torques experienced by the mannequin head could indicate reduced discomfort if  
448 applied to a human subject. Concurrently, a hair-care robot will need to be able to grasp hair that  
449 may be close to the scalp, which will likely result in higher forces experienced by the mannequin  
450 head. Then, we note that an ideal hair-care robot must be able to grasp hair effectively while also

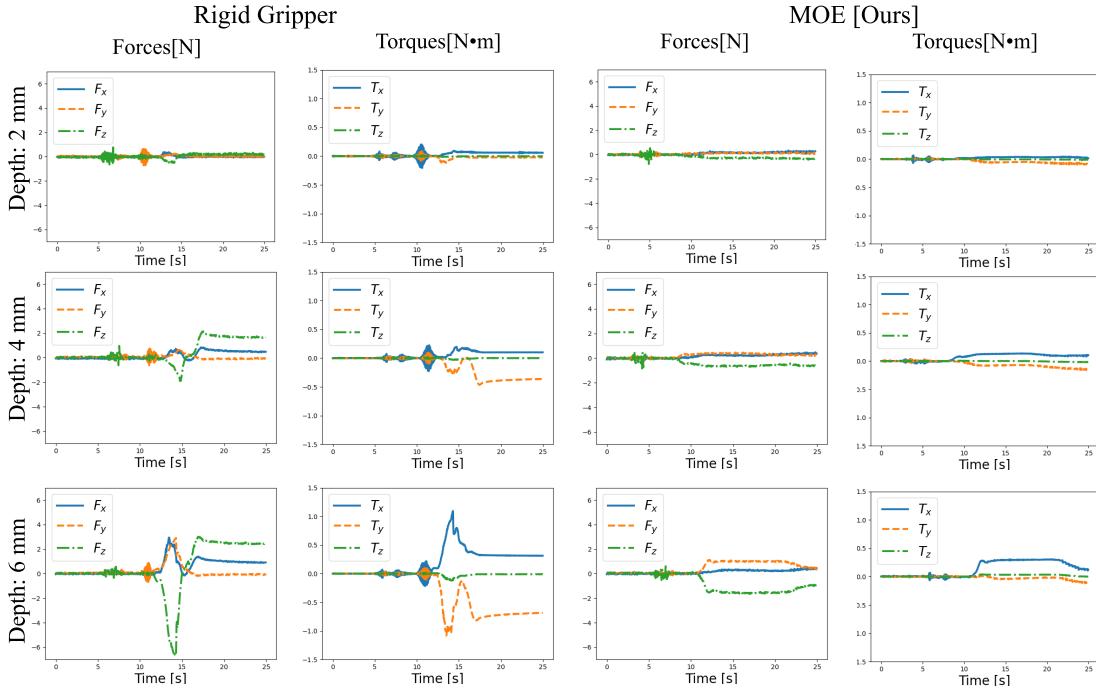


Figure 9: Sample set of YCB and a single headspace meshes simulated in contact with MOE. Relative poses were randomized to diversify the dataset.

451 applying minimal force on the head. Table 3 reports the maximum force experienced by the head at  
452 varying depths and the amount of hair grasped.

453 We note that at 6.0 mm depth, the rigid end-effector exerts 7.67 N of force on the mannequin head. At  
454 the same depth, MOE applied 1.98 N of force. This constitutes a 74.1 % reduction in the maximum  
455 force applied to the head. Meanwhile, on the grasped hair metric, MOE grasped approximately  
456 10 % less hair. A potential explanation of this marginal decrease in the amount of hair grasped is  
457 that the compliance of MOE allowed some of the grasped hair to be pried away as the end-effector  
458 moved away. This is partially supported by the fact that as the rigid gripper moved away from the  
459 head, the mannequin head experienced large changes in the forces applied, indicating possible hair-  
460 pulling by the end-effector. This change in forces as the robot hand moves away is not as evident in  
461 experiments with MOE.

## 462 B MOE Shape Estimation

463 As-Rigid-As-Possible (ARAP) involves minimizing the energy function  $E_{\text{ARAP}}$ , which is defined  
464 as the following:

$$E_{\text{ARAP}}(S, S') = \sum_{k=1}^{|E|} \min_{R \in SO(3)} \sum_{e_{i,j} \in E} w_{i,j} \|e'_{i,j} - Re_{i,j}\|.$$

465 We can then find the solution mesh that minimizes  $E_{\text{ARAP}}$  with an iterative local-global optimizer.  
466 Minimizing  $E_{\text{ARAP}}$  as is with sparse handle points on surface meshes can result in undesirable  
467 surface artifacts such as folds.

468 Minimizing the  $E_{\text{ARAP}}$  over a tetrahedral mesh can prevent these artifacts by implicitly applying  
469 soft volumetric constraints that prevent such artifacts from forming. However, operating over tetra-

Table 4: MOE Shape Estimation Evaluation.

Contact Condition	# of Keypoints	Performance Metrics [mm]	
		Avg. CD ↓	Max CD ↓
Side	7	1.16	3.19
Side	4	1.23	3.47
Back	7	1.17	3.18
Back	4	1.19	3.35

470 hedral meshes is more computationally expensive which is especially undesirable in the context of  
 471 real-time robotic tools.

472 Instead, a modification of ARAP to include a penalty on the rotations of the neighboring edges  
 473 produces more intuitively physically admissible results. The new energy to minimize is formulated  
 474 as

$$E_{\text{smoothed}}(S, S') = \min_{R_1, \dots, R_m} \sum_{k=1}^m \left( \sum_{i,j \in e_k} c_{ijk} \|e_{ij} - R_k e_{ij}\|^2 + \lambda \hat{A} \sum_{e_l \in N(e_k)} w_{kl} \|R_k - R_l\|^2 \right).$$

475 We note that minimization of  $E_{\text{smoothed}}$  with  $\lambda = 0$  results in the minimization of  $E_{\text{ARAP}}$ . We  
 476 consider the vertices corresponding to the keypoints  $p_1, \dots, |p_k|$  are constrained to the new positions  
 477 based on the predicted key-point positions, and the rest of the mesh vertex positions are moved to  
 478 minimize  $E_{\text{smoothed}}$ .

## 479 C Shape Estimation Evaluation

480 Prior work has shown that the rigidity and rotation regularization of the ARAP formulation as pre-  
 481 sented in Section 4.2 generally produces more physically admissible deformed soft bodies, com-  
 482 pared to end-to-end learning-based methods [21]. A key difference in our implementation of the  
 483 ARAP-based soft robot reconstruction is that the wrist-mounted RGB-D camera can only observe  
 484 one side of MOE’s soft surface. The underlying assumption with such an implementation choice  
 485 is that the observation of one side of MOE can directly inform us about the changes to the state of  
 486 the other side. As a consequence, we also assume that the cross-section of MOE’s fingers remains  
 487 largely the same, to allow us to infer the opposing surface’s transformation. This assumption is sup-  
 488 ported by previous works in mechanics-based modeling and validation of tendon-driven soft robotic  
 489 manipulators [25].

490 We validate shape fidelity and consistency on the side of MOE that is normally occluded from the  
 491 wrist-mounted RGB-D camera, as shown in Figure ???. We place a high-resolution RGB-D camera  
 492 (Zivid, One Plus) in a third-person view facing MOE, from either its side or back, to capture the  
 493 side that is normally unobserved in our pipeline. We also place a clear acrylic sheet facing the third-  
 494 person view RGB-D camera. This setup allows us to deform MOE against the clear sheet with a  
 495 large contact surface, while remaining fully observable to the third-person view RGB-D camera. We  
 496 present the average and maximum unidirectional Chamfer Distance (CD) results from third-person  
 497 RGB-D point cloud to the complete estimated shape, for both side and back contact conditions, in  
 498 Table 4. We can observe that the shape estimation average CD error is small at 1.16-1.17 mm for the  
 499 two contact conditions. Notably, the error is smaller than the 4.89 mm best average CD error reported  
 500 in [21]. Such results highlight a potential advantage of directly observing keypoint movements with  
 501 wrist-mounted cameras compared to indirectly inferring keypoint movements.

502 We also experiment with testing the robustness of the MOE shape estimation module by remov-  
 503 ing markers from being considered during ARAP mesh optimization. With 4 markers, we note a

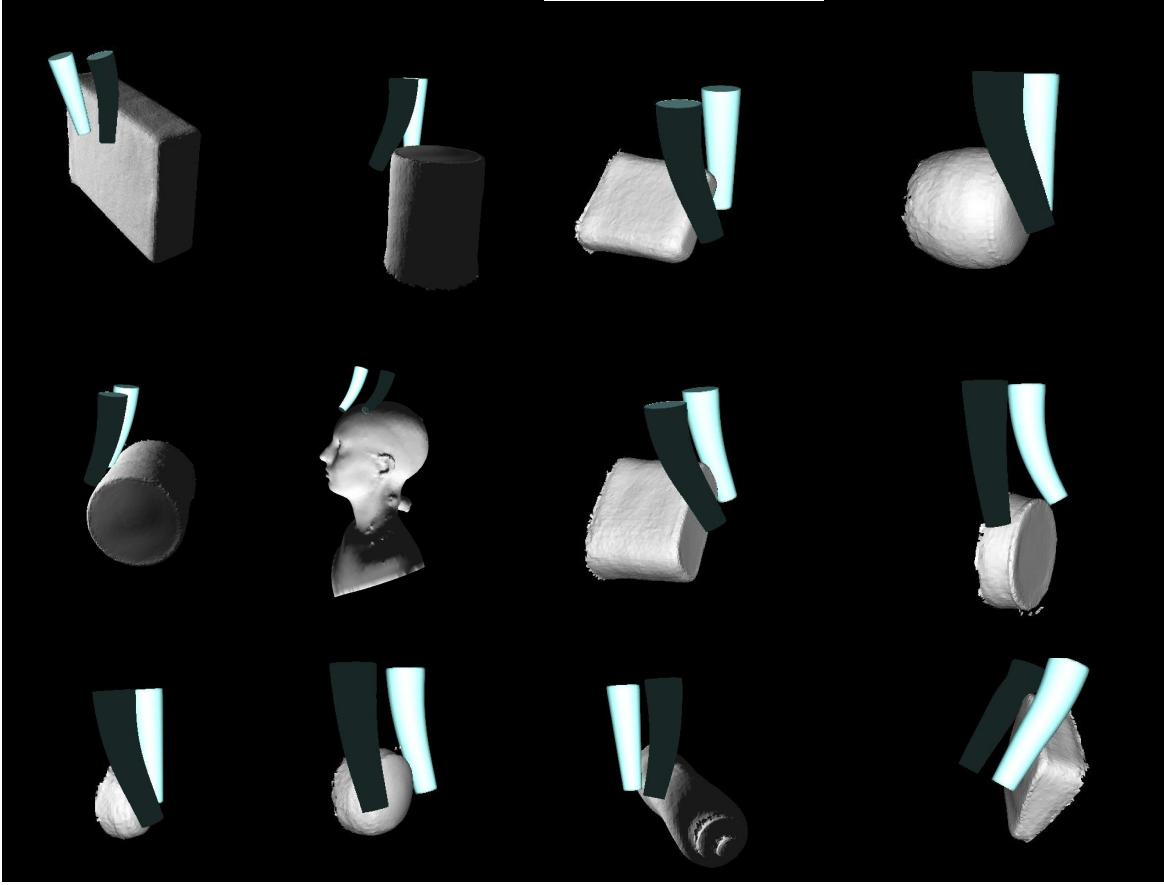


Figure 10: Sample of the simulated scenes for generating the training data for MOE contact estimation.

504 marginal increase in both average and maximum CD errors from when the shape estimation mod-  
 505 ule considered the full set of 7 markers for each finger. The relatively small change in performance  
 506 highlights the robustness of the shape estimation module, which can be partially attributed to the  
 507 well-tuned smoothening penalty to produce meshes that conform well to soft body mechanics.

## 508 D Simulation

509 We implemented the soft body simulator in SOFA Framework with its tools for solving Finite El-  
 510 ement Method (FEM) problems. We follow the previously recorded material properties for the sil-  
 511 cone body of MOE with the Poisson Ratio of 0.1 and Young’s modulus of 100 kPa. For the inte-  
 512 grator, we use the Rayleigh stiffness value of 0.1 and Rayleigh Mass of 0.1. We implemented cable  
 513 tensions for the tendons with displacement action input. The resulting simulator scenes are shown  
 514 in Figure 10.

## 515 E Surface Reconstruction

516 In this work, we used *GPyTorch* to train ExactGPs with Radial Basis Function (RBF) Kernels on  
 517 a single GPU with a sparse grid to fit within the memory of a single RTX 3090. We have shown  
 518 effective extrapolation capabilities of GPs by generating SDFs at twice the density during inference  
 519 using CPU.

520 The task-dependent GP-based surface reconstruction pipeline follows the following steps:

- 521 1. Pretrain a GP on a prior mesh that is dependent on the task to be done. The objective of  
 522 the GP is to take a grid of points ( $50 \times 50 \times 50$ ) and compute the SDF with respect to the  
 523 surface of the mesh. Since the GP is trained for 5000 epochs, this one-time process is slow  
 524 and takes about 30 mins on a single GPU. Due to sparse discretization, the reconstructed  
 525 prior is not watertight and results in holes in the mesh.
- 526 2. Next, given a set of real-world contact points, fine-tune the GP on the new points. This  
 527 process is much faster and takes about 100 epochs to train.
- 528 3. Finally, we create a dense grid and query the GP to obtain the SDF values of individual  
 529 points. Then we implement a Marching Cubes Algorithm to find the zero-level set of the  
 530 SDF. To reconstruct the final mesh, we use Poisson Surface Reconstruction from Open3D  
 531 and show the posterior reconstruction as wireframes overlayed on top of the prior recon-  
 532 struction in Figure 10 of the paper.

533 However, the grid formulation in (1.) is limiting for surfaces that are not uniformly distributed. This  
 534 is crucial because reconstructing the head is relatively easy due to the approximately 1:1:1 aspect  
 535 ratio. However, since the hand reconstruction grid is non-uniform with a 2:20:1 aspect ratio, a  
 536 uniformly distributed grid ( $50 \times 50 \times 50$ ) can not be used straightforwardly.

537 To address this issue, we sample an extremely dense grid of shape ( $200 \times 200 \times 200$ ), and randomly  
 538 sample 30,000 points and follow the same pipeline as above. This simple trick significantly im-  
 539 proves our reconstructions and allows us to leverage the expressivity of Gaussian Processes on non-  
 540 linear surfaces with a higher fidelity compared to uniformly generated dense grids.

#### 541 E.0.1 No Prior

542 A naive approach for surface reconstruction is a non-probabilistic method by fitting a sphere to the  
 543 contact points collected in the simulation. Given a set of points  $\{\mathbf{P}_i\}_{i=1}^N$ , where  $\mathbf{P}_i$  is the  $i^{th}$  point  
 544 in 3D space, the objective function for fitting a sphere to the points is defined as

$$f(\mathbf{c}, r) = \sum_{i=1}^N \left( \sqrt{\sum_{j=1}^3 (P_{ij} - c_j)^2} - r \right)^2, \quad (5)$$

545 where  $\mathbf{c} \in \mathcal{R}^3$  represents the center of the sphere, and  $r$  is the estimated sphere's radius. The initial  
 546 guess for the optimization is

$$\mathbf{c}_0 = \frac{1}{N} \sum_{i=1}^N \mathbf{P}_i, \quad (6)$$

$$r_0 = \frac{1}{N} \sum_{i=1}^N \sqrt{\sum_{j=1}^3 (P_{ij} - c_{0j})^2}. \quad (7)$$

547 We solve the optimization problem  $\min_{\mathbf{c}, r} f(\mathbf{c}, r)$  using L-BFGS [49] to find the  $\mathbf{c}$  and  $r$  that mini-  
 548 mize  $f(\mathbf{c}, r)$ .

549 We sample a point cloud for the sphere and implement a k-d tree-based nearest neighbor search to  
 550 average the residuals between the contact points and the spherical mesh. Finally, we smooth out the  
 551 abrupt changes to the mesh using a smoothing Laplacian filter.

#### 552 E.0.2 Spherical Prior

553 Gaussian Processes have been extensively studied for implicit surface reconstruction in the literature  
 554 [15, 16, 47]. We implement a modified version of GPIS that runs on GPU, to represent the signed  
 555 distance functions (SDFs) of the head without needing surface normals. Generally, active explo-  
 556 ration algorithms assume an initial condition of uniformly distributed points in a grid. Every mea-  
 557 surement reduces the uncertainty until the final shape of the object is represented by the GP mean.

558 We fit a spherical mesh to the these points, and use this sphere as a prior to train the GP over a dense  
 559 3D array of grid points encompassing the mesh. The SDF values for each point  $P_i$  are computed as

$$SDF(P_i) = r - \|P_i - \mathbf{c}\|. \quad (8)$$

560 Given a set of dense grid points  $\mathbf{X}$  and corresponding SDF values  $\mathbf{Y}$ , the GP model is defined as

$$f(\mathbf{x}) \sim \mathcal{GP} \left( c, \sigma^2 \exp \left( -\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2l^2} \right) \right), \quad (9)$$

561 where  $\sigma^2$  is the variance  $l$  is the length scale of the Radial Basis Function (RBF) kernel, with the  
 562 observation model

$$y = f(\mathbf{x}) + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma_n^2) \quad (10)$$

563 where  $n$  is the number of training points. The training objective is to maximize the marginal log-  
 564 likelihood loss

$$\begin{aligned} \log p(\mathbf{Y}|\mathbf{X}) &= -\frac{1}{2} \mathbf{Y}^\top (\mathbf{K} + \sigma_n^2 \mathbf{I})^{-1} \mathbf{Y} \\ &\quad - \frac{1}{2} \log |\mathbf{K} + \sigma_n^2 \mathbf{I}| - \frac{n}{2} \log 2\pi, \end{aligned} \quad (11)$$

565 where  $\mathbf{K}$  denotes the covariance matrix constructed using an RBF kernel over the training inputs in  
 566  $\mathbf{X}$ . Once we have a spherical prior, the GP is updated with the contact point information to obtain  
 567 the posterior SDF. Finally, the head mesh is reconstructed using Poisson Surface Reconstruction  
 568 (PSR) on a point cloud obtained by running the Marching Cubes Algorithm (MCA) over the zero-  
 569 level set of the SDF.