

Estimating High-Resolution Neural Stiffness Fields using Visuotactile Sensors

Jiaheng Han*

Shaoxiong Yao*

Kris Hauser

Abstract—High-resolution visuotactile sensors provide detailed contact information that is promising to infer the physical properties of objects in contact. This paper introduces a novel technique for high-resolution stiffness estimation of heterogeneous deformable objects using the Punyo bubble sensor. We developed an observation model for dense contact forces to estimate object stiffness using a visuotactile sensor and a dense force estimator. Additionally, we propose a neural Volumetric Stiffness Field (VSF) formulation that represents stiffness as a continuous function, which allows dynamic point sampling at visuotactile sensor observation resolution. The neural VSF significantly reduces artifacts commonly associated with traditional point-based methods, particularly in stiff inclusion estimation and heterogeneous stiffness estimation. We further apply our method in a blind localization task, where objects within opaque bags are accurately modeled and localized, demonstrating the superior performance of neural VSF compared to existing techniques. Project page: <https://jhj371.github.io/Neural-VSF/>.

I. INTRODUCTION

Tactile perception of physical properties, such as stiffness, is crucial for tasks including manipulation in clutter [1] and deformable object packaging [2]. Recent advances in vision-based tactile sensors capture high-resolution tactile images by observing the deformation of media like gels [3] or bubbles [4] using embedded cameras. These high-resolution tactile observations provide detailed contact information and are promising for fine-grained texture recognition and material parameter estimation. However, the vast majority of prior work in tactile sensing considers contact with rigid objects. Deformable objects undergo deformation in a manner that is coupled to the sensor medium, and material parameter estimation requires simultaneous understanding of both object and sensor geometry, as well as their respective deformations. Moreover, many objects of interest for high-resolution tactile sensing are heterogeneous, with a mix of hard and soft parts, such as bagged objects or localization of hard tissues underneath soft tissues in palpation. Existing identification techniques based on analytical models, e.g., finite element models (FEM), do not scale easily to large numbers of material parameters and handling deformable deformable contact can be challenging.

This work extends a recent line of research into dense representations of heterogeneous deformable objects called the volumetric stiffness field (VSF) [5]. VSF approximates a

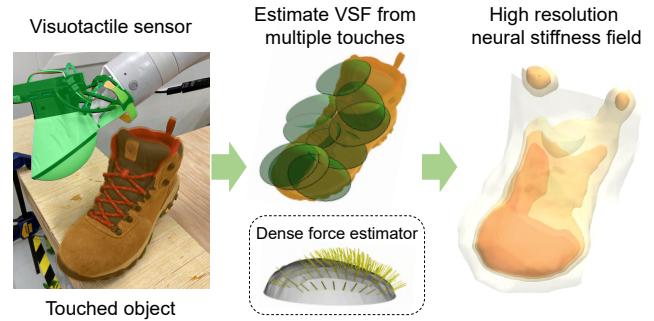


Fig. 1. Our method estimates a high-resolution neural stiffness field of a deformable object using a corresponding visuotactile sensor and a dense force estimator. The toe region of the boot is correctly estimated with higher stiffness compared to the softer tongue region.

volume containing the object as a set of particles connected to rest positions by springs, and the spring stiffness estimation problem becomes parallelizable, enabling real-time estimation across tens of thousands of particles. However, past work has relied only on low-dimensional sensors with nearly-rigid mountings, such as robot joint torque sensors or force-torque sensors. This paper proposes novel techniques for dense stiffness estimation using high-resolution visuotactile sensors that deform along with the object.

Our contributions are twofold. First, we formulate a dense force observation model that considers coupled deformation between the sensor and the object in the estimation process. We model the Punyo bubble sensor as a thin-shell FEM along with internal air. The observation from an embedded camera estimates a deformation of the bubble in response to a robot movement, and we use an existing FEM-based dense force estimator and the principle of virtual work to estimate the object stiffnesses across the entire volume.

Moreover, we propose a novel *neural VSF* formulation that represents the stiffness field as a continuous function, estimated as a neural network. Rather than pre-discretizing the object volume into a set of points, which can lead to discretization artifacts, we use the high-resolution sensor observations and robot motion to dynamically determine *sample locations* at which the observed sensor deformation is related to the underlying continuous stiffness field. We use the principle of virtual work to establish an observation function and estimate the stiffness through gradient descent of the neural VSF parameters. This approach lets us efficiently solve the material parameter estimation problem in approximately 1 minute using around 60 touch sequences, containing over 1000 frames (Fig. 1). Moreover, the neural representation provides a form of spatial regularization that

*These authors contributed equally to this work.

J. Han, S. Yao, and K. Hauser are with the Siebel School of Computing and Data Science, University of Illinois at Urbana-Champaign, IL, USA. {jhj371, syao16, kkhauser}@illinois.edu. This project was partially supported by USDA/NIFA Grant #2020-67021-32799 and an equipment loan from Toyota Research Institute.

leads to fewer artifacts than the point-based representation.

To validate our method, we address a *blind tactile localization* problem in which an object is placed in a deformable, opaque, plastic bag and the robot builds a model of the object by pressing through the bag. Our method is able to build deformation models through the bag through touch alone, and localize objects to a high degree of accuracy.

In summary, this paper has three main contributions: 1) We formulate an observation model for dense force estimation using visuotactile sensors combined with the VSF tactile model; 2) We propose a novel neural VSF tactile model capable of adaptively instantiating particles based on visuotactile sensor resolution; 3) We demonstrate that the proposed VSF model significantly enhances the accuracy of blind tactile object localization compared to baseline methods.

II. RELATED WORK

Tactile perception using visuotactile sensors. Tactile perception provides contact information that characterizes the physical properties of touched objects [6]–[8]. Visuotactile sensors that can provide high-resolution contact observations have gained increasing interest in recent years, including GelSight [3], Punyo [4] and TacTip [9]. By tracking the deformation of touched media from images, these sensors can accurately estimate contact regions [10] and dense contact forces [11], [12]. The high-resolution contact information enables accurate rigid geometry estimation since sensor deformation is the indentation of object geometry [13]. However, when touching deformable objects, the observed sensor deformation is coupled with object deformation, and estimating object material parameters from sensor observation is non-trivial. Analytical relations can be derived but require a known shape of touched object [14]. Data-driven approaches can learn a material classification model independent of object geometry [15]–[17]. However, these works require collecting annotated data and can only output discrete classification labels from single-touch observations. In contrast, our approach estimates a NeRF-style neural stiffness model directly from tactile data obtained through touching the object. By integrating multiple touch observations, our method can effectively recognize the heterogeneous stiffness properties of the object.

Blind tactile localization of rigid objects has been explored in previous works, such as [18], which utilizes a force/torque sensor, and [19], which generates diverse plausible shapes. However, to the best of our knowledge, our work is the first to incorporate object stiffness for the accurate localization of deformable objects.

Deformable object model estimation. Deformable objects are ubiquitous in daily life and various deformable object models have been used for robotic manipulation [20]. Existing physics-based models, such as the Mass-Spring System [21] and FEM [22], as well as learning-based models like graph neural networks [23], [24], often assume homogeneity of the object to simplify estimation. However, this assumption limits the models' ability to capture spatial variations in material properties.

The Volumetric Stiffness Field (VSF) model addresses material heterogeneity by estimating the stiffness of dense independent Hookean springs [5]. However, the resolution of VSF is constrained by the point density, which may not align with the high resolution of visuotactile sensors. Inspired by the NeRF-style neural field, which provides a continuous representation of the light field for novel view synthesis [25], we extend our point-based approach by adopting a NeRF-style neural field to define a continuous stiffness distribution. Previously, NeRF-style networks have been used by VIRDO [26], NDCF [27] and ACID [28] to model deformation fields of objects, but our approach is the first to apply them to deformable material parameter estimation. More recently, NeRF has been used to generate geometry maps paired with generative models to synthesize tactile images [29], [30]. In contrast, our neural VSF addresses the material parameter estimation problem and works in the “blind” case when no vision and geometry data are available.

III. METHOD

The input to our systems includes a volume containing the object, a set of trajectories from a high-resolution tactile sensor, and the corresponding sequences of sensor readings. We assume that the visuotactile sensor readings are displacements of the contact surface, and we have access to a calibrated FEM model of the tactile sensor that relates contact forces to these displacements. The object is assumed to deform in a manner that is dominated by elastic effects.

Our high-resolution VSF tactile estimation system, using a visuotactile sensor, has four key components: a dense contact forces estimator, a VSF representation, a dense contact forces observation model for the VSF, and a VSF estimation algorithm. We first review the existing dense force estimator and earlier point-based VSF. Then, we present point-based VSF estimation using dense contact forces and introduce a novel neural VSF representation that reduces artifacts and improves estimation time compared to a point-based VSF.

A. Dense force estimation

We use an FEM-based force estimator to determine the contact force distribution on the Punyo bubble sensor [31]. The sensor provides RGBD images from an embedded camera and internal pressure readings from a pressure sensor. The sensor bubble surface is modeled as a triangle mesh M , defined as $M = (V, \Delta)$, where V represents the set of vertices and Δ denotes the set of triangular faces given by $\Delta = \{\Delta_j = (j_1, j_2, j_3) \mid j_1, j_2, j_3 \in \{1, 2, \dots, n\}\}$. We denote the world coordinates of n vertices $V = \{v_i\}_{i=1}^n$ and $v_i \in \mathbb{R}^3$. At time step t , the local coordinates of vertices are tracked in the sensor frame using optical flow, projected into 3D space with depth imaging, and transformed into world coordinates v_i^t using robot forward kinematics.

A dense force estimator estimates external contact forces at each vertex on the triangle mesh at current time t , denoted as $f_v^t = (f_{v,1}^t, f_{v,2}^t, \dots, f_{v,n}^t)$, where $f_{v,i}^t \in \mathbb{R}^3$. The external contact forces can be computed as the residual forces from internal air pressure forces f_{pressure} and elastic forces f_{elastic}

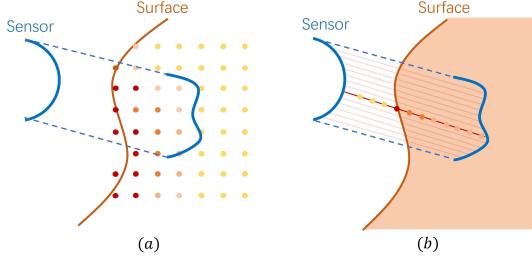


Fig. 2. (a). Point-based VSF. The object space is discretized as a uniform grid. Each VSF point has an independent spring with its own stiffness. (b). Neural VSF. The stiffness field is continuously defined by a NeRF-like network. The contact force is evaluated by sampling points along the vertex trajectories on the sensor surface, illustrated by the solid and blurred lines.

using force equilibrium conditions. The pressure forces are computed from the internal pressure reading multiplied by the triangle area. A calibrated linear FEM model relates the vertices' displacements u_v to the internal elastic forces at each vertex through a stiffness matrix, $f_{\text{elastic}} = Ku_v$. To mitigate noise in the tracked deformation, a Quadratic Programming (QP) formulation is used to regularize the estimated forces.

B. Point-based VSF Model

The original point-based VSF represents the object as a discrete set of points $\{p_i\}_{i=1..m}$ within the volume $\Omega \subset \mathbb{R}^3$. When the visuotactile sensor interacts with Ω , at time t , a subset of points $S^t \subset \{1, \dots, m\}$ will undergo deformation u_i^t , which is simulated using a stick-slip simulation model. In this paper, due to high friction on the bubble surface, we simulate VSF points sticking to the sensor surface, detaching only when reactive forces point outward. Each point is modeled as a Hookean spring with stiffness coefficient $K_i \geq 0$. Upon deformation by u_i^t , the point generates a reactive force $f_i^t = -K_i \cdot u_i^t$. Given a low-dimensional linear force/torque observation model, a diagonal Kalman filter can be used to enable real-time estimation.

C. Estimating point-based VSF from dense contact forces

To estimate VSF stiffness, we need to define an observation model of contact forces in terms of VSF stiffness. For high-resolution visuotactile sensors, we can see the point density of VSF may not match the sensor resolution as in Fig. 2. Therefore, we use the principle of virtual work to derive this relation.

Dense force observation model for point-based VSF. When the visuotactile sensor touches the object, we can detect VSF points in contact with the triangle mesh surface. For VSF point i in contact with the sensor surface on triangle $\Delta_j = (j_1, j_2, j_3)$, its current location is a linear combination of triangle vertices in contact. The current point location is $p_i^t = p_i^0 + u_i^t = \lambda_1 v_{j_1}^t + \lambda_2 v_{j_2}^t + \lambda_3 v_{j_3}^t$ where $\lambda_1, \lambda_2, \lambda_3$ are the barycentric coordinates of p_i^t in Δ_j . We can succinctly write this linear relation in a matrix format $u_i^t = A_i^t V^t - p_i^0$, where A_i^t has size $\mathbb{R}^{3 \times 3n}$ and V^t is the concatenation of sensor vertices at time t .

Therefore when the bubble vertices changes by δV , point deformation in VSF changes by $\delta u_i^t = A_i^t \delta V^t$ and the VSF

elastic energy changes by:

$$\delta E_{\text{vsf}} = \delta \left(\sum_{i \in S^t} \frac{1}{2} K_i \|u_i^t\|_2^2 \right) = \left(\sum_{i \in S^t} K_i u_i^t \top A_i^t \right) \delta V^t \quad (1)$$

Using the principle of virtual work, at the equilibrium state with a deformation δV^t on the visuotactile sensor, the change in VSF energy plus the external work should equal zero:

$$\delta E_{\text{vsf}} + f_v^t \top \delta V^t = \left(\sum_{i \in S^t} K_i u_i^t \top A_i^t + f_v^t \top \right) \delta V^t = 0. \quad (2)$$

Therefore, the expected contact forces on the sensor are:

$$\hat{f}_v^t = - \sum_{i \in S^t} A_i^t \top K_i u_i^t. \quad (3)$$

In this way, the dense contact forces \hat{f}_v^t on the visuotactile sensor are a linear model of VSF stiffness K_i . The observation matrix is formed by concatenating $A_i^t \top K_i$ over contact points in S^t . Here by assuming a Gaussian noise on the observation vector \hat{f}_v^t , we can use the same diagonal Kalman filter to estimate VSF stiffness K_i .

D. Continuous VSF Model with Neural Representation

The point-based VSF model provides dense stiffness estimates but relies on choosing a particle resolution, leading to artifacts when the resolution is too low or too high. In contrast, our continuous neural VSF allows point sampling at sensor resolution and simplifies the dense forces observation model evaluation.

Continuous VSF model. Our key innovation is to treat the stiffness as a continuous field $K(\cdot) : \Omega \rightarrow \mathbb{R}_+$, where we can dynamically sample points at visuotactile sensor force estimation resolution. For a tiny volume dV centered around $p \in \Omega$, there exists a Hookean spring with stiffness $K(p)dV$. When the robot pushes through this volume, the object deforms according to a time-dependent continuous deformation field $u(p, t) : \Omega \times \mathbb{R} \rightarrow \mathbb{R}^3$.

Dense force observation model for continuous VSF. The continuous stiffness field enables direct evaluation of dense contact forces, achieved by integrating the reactive forces of springs within the swept volume. For a vertex k , moving along its trajectory $v_k(t) : \mathbb{R} \rightarrow \mathbb{R}^3$, the observed contact force $\hat{f}_{v,k}(t)$ is the integration of all Hookean springs within its swept volume Ω_k . We assume high friction on the sensor surface and we only consider stick motion of VSF point. For VSF point in contact with vertex k at time τ , its deformation at $t \geq \tau$ is $u(v_k(\tau), t) = v_k(t) - v_k(\tau)$. To compute the swept volume, the vertex is approximated with an oriented surface s_k in the direction of its vertex normal, and its size $|s_k|$ is defined as the sum of 1/3 all triangle areas incident to it. When the vertex moves by an infinitesimal displacement dv , the differential swept volume is given by $dV = s_k(t) \cdot dv_k(t)$. Thus, the observed contact force $\hat{f}_{v,k}(t)$ is an integration over the trajectory up to time t , expressed as

$$\begin{aligned} \hat{f}_{v,k}(t) &= \int_{\Omega_k} -K(p) u(p, t) dV \\ &= \int_{v_k(0)}^{v_k(t)} -K(v_k(\tau)) u(v_k(\tau), t) s_k(\tau) \cdot dv_k(\tau) \end{aligned} \quad (4)$$

where $K(v_k(\tau))$ represents the local stiffness at the vertex, and $u(v_k(\tau), t)$ denotes the deformation at the vertex position $v_k(\tau)$ over time.

This line integration for $\hat{f}_{v,k}(t)$ is approximated by sampling points along vertex trajectories from 0 to time t (by default $N = 100$ samples). The line integration in Eq. 4 can be approximated as a summation over line segments between each time step. For a touch sequence with a total N samples, the line segment between time step τ_i and τ_{i-1} is $\delta v_k(\tau_i) = v_k(\tau_i) - v_k(\tau_{i-1})$, and the summation is

$$\hat{f}_{v,k}(t) = \sum_{i=1}^N -K(v_k(\tau_i))u(v_k(\tau_i), t)s_k(\tau_i) \cdot \delta v_k(\tau_i). \quad (5)$$

In this way, the dense contact forces $\hat{f}_{v,k}(t)$ at each vertex are evaluated in parallel.

NeRF-like neural VSF. We instantiate the continuous VSF using a NeRF-like neural network as $K(p) = g_\psi(p) : \mathbb{R}^3 \rightarrow \mathbb{R}_+$, where the input p is a 3D point coordinate, the output is a nonnegative scalar stiffness value, and ψ denotes the neural network parameters. The network g_ψ has an 8-layer architecture with sinusoidal position encoding functions, similar to NeRF [25], to capture high-frequency details of the objects.

Training neural VSF. We optimize ψ by minimizing the following loss function on dense contact forces:

$$\ell(\psi) = \sum_{i=1}^B \|f_{v,k_i}^{t_i} - \hat{f}_{v,k_i}(t_i)\|_2 + \lambda \frac{1}{n_{\text{reg}}} \sum_{i=1}^{n_{\text{reg}}} K(p_i) \quad (6)$$

Neural VSF is trained over multiple touch sequences that touch the object at different locations. During each training iteration, a batch of $B = 1000$ force observations are randomly selected from different sequences to compute VSF force loss. Here, the second term is a free space regularization term that encourages the neural network zero stiffness outputs in untouched regions. This regularization is implemented by randomly sampling n_{reg} points within the bounding box and applying a small penalty to their stiffness values. The Adam optimizer [32] is used with a learning rate of 2×10^{-4} , and the total number of training iterations ranges from 1,000 to 3,000, depending on the specific task.

E. Blind tactile object localization using VSF

In this section, we demonstrate how the estimated VSF can be utilized for tactile-only object localization. We consider an object wrapped in a flexible material with lower stiffness compared to the object itself. Assuming we have a known triangle mesh M_{obj} of the object, and our goal is to determine its 6-DOF pose through tactile sensing.

Our VSF-based localization aligns both geometry and material parameters distribution. We first extract the VSF points with stiffness higher than K_{thres} and form a point cloud P_{vsf} . Here K_{thres} is chosen to be greater than the cover stiffness. The localization algorithm iteratively generates possible poses of the object. In each iteration, we register the object model, M_{obj} , to the point cloud P_{vsf} using RANSAC for global registration, followed by ICP for fine-tuning, resulting

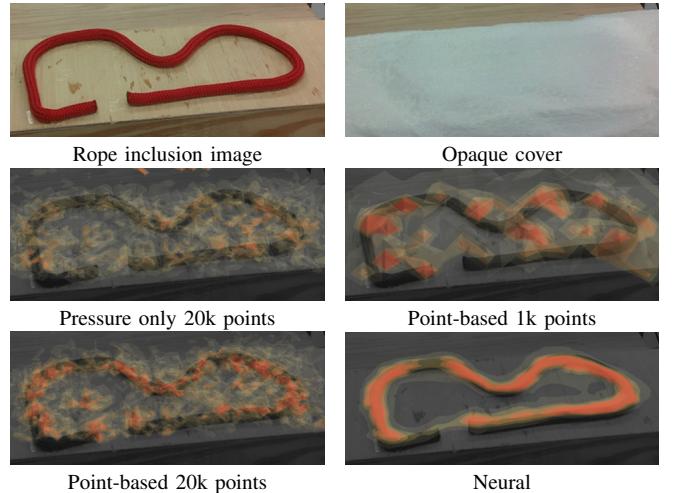


Fig. 3. Blind dense tactile mapping of a deformable rope. VSFs are shown as transparent heat-maps overlaid on a desaturated image of the object. The point-based methods generally identify stiffer parts of the rope but exhibit significant artifacts. Our proposed neural VSF leads to a smoother and more accurate localization.

in a rigidly transformed models M'_{obj} . We then penalize low stiffness values inside the transformed mesh using the following loss function:

$$\mathcal{L}(M'_{\text{obj}}) = \sum_{p_i \in M'_{\text{obj}}} \max(K_{\text{thres}} - K(p_i), 0), \quad (7)$$

where p_i denotes the touched VSF points in M'_{obj} and $K(p_i)$ is the stiffness at point p_i . We select the registered pose with the smallest loss, $\mathcal{L}(M'_{\text{obj}})$, across a fixed number of iterations. This approach allows us to localize the object pose that best matches both the geometric structure and the stiffness distribution of the touched object.

IV. EXPERIMENTS AND RESULTS

Our experiments evaluate the high-resolution VSF estimation on a set of household objects. Our system uses a Kinova robot arm with Punyo tactile sensor mounted at the end-effector. First, in a stiff inclusion estimation setup, we demonstrate that the neural VSF achieves significantly better shape estimation accuracy than the point-based VSF. Second, through the estimation of heterogeneous stiffness in shoes, we further highlight neural VSF can estimate smooth and more plausible stiffness distribution. Finally, we show that the estimated VSF can be used for blind tactile localization.

A. Stiff inclusion estimation

Setup and Baselines. We first consider a setting inspired by palpation where the robot needs to identify stiff inclusions beneath an opaque cover. For this experiment, we use a nylon rope as the stiffness inclusion and a plastic bag as the cover, as shown in Fig. 3. Here we sample 40-70 touching sequences to uniformly cover the occluded region. We first compare a pressure-only baseline using point-based VSF that considers Punyo total force as only observation. This baseline uses the scalar pressure reading from the Punyo sensor as the observed total force and the predicted total force is calculated by summing the forces from all VSF points. We consider

TABLE I
STIFF INCLUSION ESTIMATION EVALUATION.

	Recall	Precision	F1-score
Pressure only 20k points	0.346	0.290	0.314
Point-based 1k points	0.025	0.251	0.044
Point-based 20k points	0.338	0.335	0.329
Neural	0.712	0.382	0.493

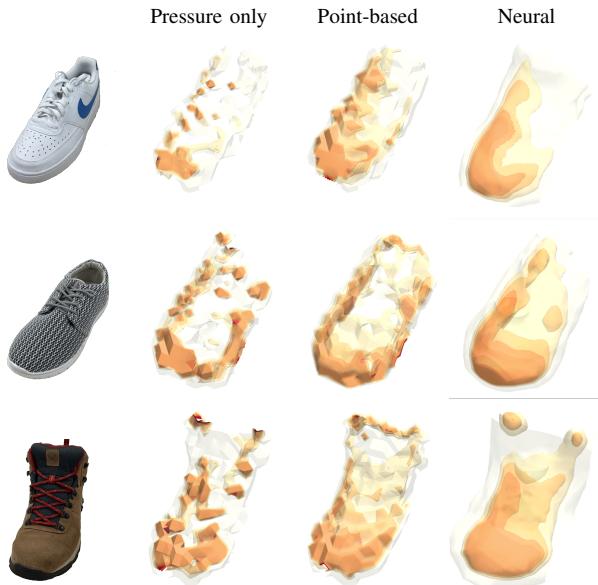


Fig. 4. Qualitative stiffness estimation comparison using Pressure only vs. Point-based VSF estimation vs. Neural VSF estimation using dense force estimation.

point-based VSF models at different resolutions, specifically with 1,000 points and 20,000 points. The VSF estimation procedure follows details from Sec. III-C and Sec. III-D.

Results. As shown in Fig. 3, all point-based VSFs fail to accurately identify the underlying shape of the rope, even when the number of points is increased. In contrast, the neural VSF model effectively recognizes the rope shape due to implicit regularization from the NeRF-style network structure, resulting in more accurate estimations. To evaluate the shape estimation accuracy quantitatively, we compare the high-stiffness region with the ground truth rope mask across three rope configurations in Tab. I. The ground truth rope mask is manually annotated in the image space and the high-stiffness VSF points are projected in the image space as rope mask prediction. The stiffness threshold is chosen individually for each method. Comparing these two binary masks, we can evaluate the precision, recall, and F1-score in rope shape prediction. We can see the neural VSF has much higher recall and the predicted rope shape is more complete. However, the neural VSF has a smaller improvement in precision, likely due to the overestimated high stiffness around the rope edges.

B. Estimate Objects with Heterogeneous Stiffness

Setup and Baselines. We evaluate VSF variants on heterogeneous stiffness estimation with household objects. We select three shoes, each with higher stiffness in the toe region and lower stiffness in the tongue region. We compare three different VSF formulations, including the pressure-only

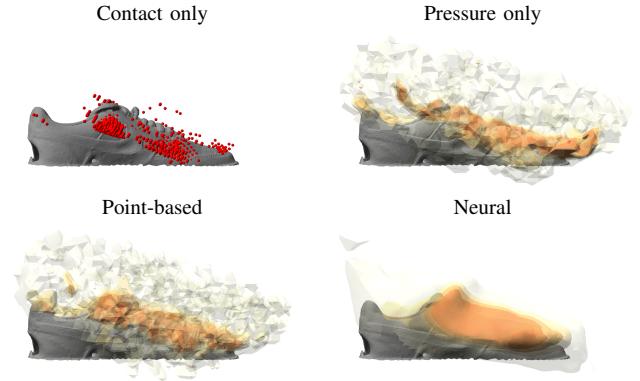


Fig. 5. Blind tactile localization of a shoe hidden inside the plastic bag of Fig. 6, with tactile models overlaid over the ground truth pose. The contact-only approach detects a discrete number of contact points in red balls, while the other three approaches estimate the shoe’s stiffness as VSF. Registration accuracies are shown in Fig. 7.

method mentioned above, point-based VSF and neural VSF. For both the pressure-only model and the point-based model, we generate 20,000 VSF points from RGBD image of the object obtained with an external L515 camera using ray marching. For neural VSF, the bounding box of the object is used to define object space.

Results. The stiffness estimation results are illustrated in Fig. 4. The pressure-only model provides a noisy estimation, whereas both the point-based and neural models, which utilize the dense force estimation from the Punyo sensor, offer more accurate stiffness estimates. These models generally produce a reasonable stiffness distribution. When comparing the point-based VSF to the neural VSF, it is evident that the point-based method exhibits more artifacts.

C. Blind tactile object localization evaluation

In this section, we evaluate the effectiveness of the estimated VSF model for blind tactile object localization.

Setup. We select three objects with varying shapes and stiffness levels as test subjects, evaluating each in four different poses on the table. For each experiment, the objects were enclosed in a low-stiffness bag. Between 25 and 75 trajectories were generated by pressing uniformly random locations on the cover to collect contact force data. The ground truth configuration of each object was captured using an RGBD image of the object without cover from a calibrated external L515 camera.

Baselines. We consider three baseline methods in localizing the objects based on this contact force information, including the contact point model, pressure-only model, and point-based VSF. A contact-only baseline that approximates the rigid object surface geometry by detecting contact points higher than a force threshold. This baseline uses RANSAC + ICP point cloud registration without VSF loss. Two point-based VSF baselines using pressure reading or dense contact forces that estimate stiffness using method the same as Sec. IV-A. The same stiffness loss in Eq. 7 is used for point-based VSF registration.

Results. As shown in Figure 5, the variability in object stiffness and the non-zero stiffness of the covering bag result

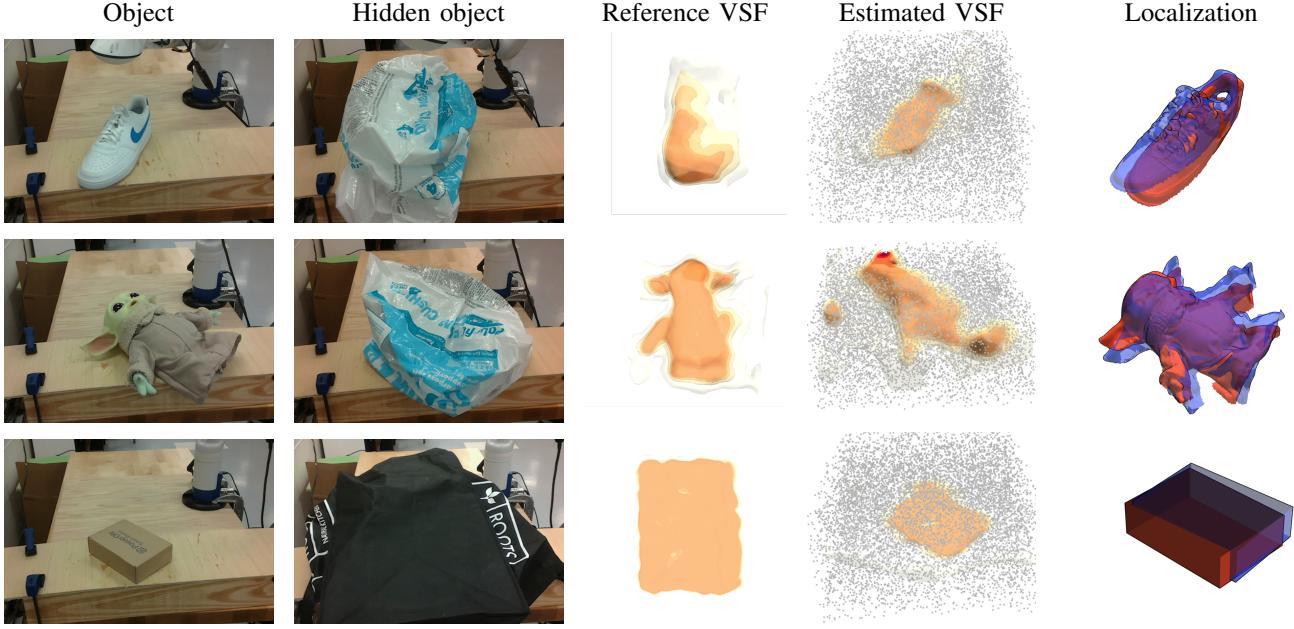


Fig. 6. Blind localization of hidden objects. Columns from left to right indicate: 1) the object, 2) the deformable container hiding the object, 3) reference object VSF, 4) blindly estimated VSF of the object within the container, 5) blind localization of object pose (blue) compared to ground truth (red).

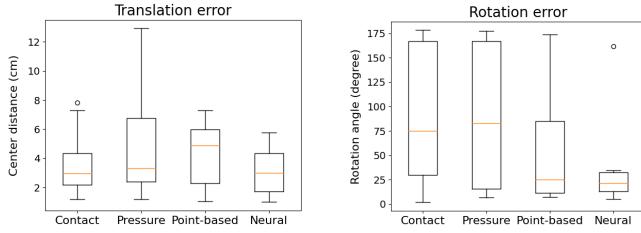


Fig. 7. Box plot of blind localization error in translation and rotation.

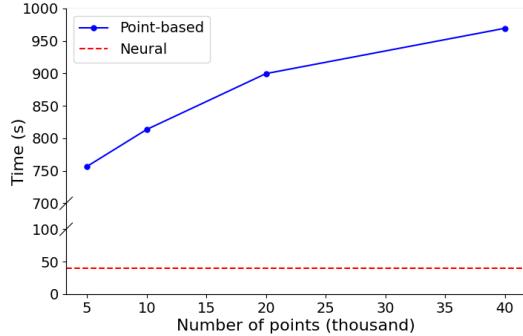


Fig. 8. Total estimation time for a shoe object with 66 touch sequences using point-based or neural VSF representation. For point-based VSF, training time increases as the VSF resolution increases.

in noisy contact point estimations when a constant force threshold is employed. Among the stiffness-based methods, both the point-based VSF and the neural VSF provide reasonable results; however, the point-based method tends to produce noisier estimates. We report the translation and rotation errors between the ground truth configuration and the predictions, as illustrated in Fig. 7. While all four methods achieve reasonable object localization, as indicated by translation error, the contact point-based method and the pressure-only model perform poorly in orientation estimation. Among them, the neural approach has the highest accuracy.

D. Computation time and storage analysis

We evaluate the computation time of different VSF variants through shoe VSF estimation using 66 touching sequences and 1010 frames of observations shown in Fig. 8. Although the point-based VSF utilizes a diagonal Kalman filter to speedup the estimation, the neural VSF significantly reduces computation time by 15-20 times due to its simpler observation model, which allows for batched evaluation over an entire touch sequence on GPU.

The neural VSF also offers substantial memory savings. A single neural VSF requires only \sim 100 KB compared to \sim 500 KB for the point-based VSF with 20k points, and the gap increases as the resolution of point-based VSF becomes finer or the object volume grows larger.

V. CONCLUSION AND FUTURE WORK

We proposed a novel high-resolution stiffness estimation technique for heterogeneous deformable objects using high-resolution tactile sensors. Our experiments demonstrated that the neural VSF representation significantly reduces estimation artifacts compared to earlier point-based methods in both stiff inclusion and heterogeneous stiffness estimation. The estimated neural VSF is also effective in blind tactile object localization, outperforming baseline methods by accurately modeling and localizing objects within opaque bags. This demonstrates that neural VSF is a promising representation to enable downstream deformable object manipulation tasks, such as object retrieval under heavy occlusion and dense packaging. Looking ahead, we plan to extend this high-resolution stiffness estimation system to larger scenes involving multiple objects, where the neural representation has a clear computational advantage. By expanding the neural VSF dataset, we aim to apply meta-learning techniques to enable neural VSF estimation with few touches [33].

REFERENCES

- [1] A. Jain, M. D. Killpack, A. Edsinger, and C. C. Kemp, "Reaching in clutter with whole-arm tactile sensing," *The International Journal of Robotics Research*, vol. 32, no. 4, pp. 458–482, 2013.
- [2] B. Ai, S. Tian, H. Shi, Y. Wang, C. Tan, Y. Li, and J. Wu, "RoboPack: Learning Tactile-Informed Dynamics Models for Dense Packing," in *Proceedings of Robotics: Science and Systems*, Delft, Netherlands, July 2024.
- [3] W. Yuan, S. Dong, and E. Adelson, "GelSight: High-Resolution Robot Tactile Sensors for Estimating Geometry and Force," *Sensors*, vol. 17, no. 12, p. 2762, Nov. 2017. [Online]. Available: <http://www.mdpi.com/1424-8220/17/12/2762>
- [4] A. Alspach, K. Hashimoto, N. S. Kuppuswamy, and R. Tedrake, "Soft-bubble: A highly compliant dense geometry tactile sensor for robot manipulation," *2019 2nd IEEE International Conference on Soft Robotics (RoboSoft)*, pp. 597–604, 2019. [Online]. Available: <https://api.semanticscholar.org/CorpusID:102336499>
- [5] S. Yao and K. Hauser, "Estimating tactile models of heterogeneous deformable objects in real time," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 12 583–12 589.
- [6] Y. Zhu, K. Lu, and K. Hauser, "Semi-empirical simulation of learned force response models for heterogeneous elastic objects," in *IEEE Int. Conf. Robotics and Automation (ICRA)*, 2020, pp. 1646–1652.
- [7] S. Luo, J. Bimbo, R. Dahya, and H. Liu, "Robotic tactile perception of object properties: A review," *Mechatronics*, vol. 48, pp. 54–67, 2017. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0957415817301575>
- [8] Q. Li, O. Kroemer, Z. Su, F. F. Veiga, M. Kaboli, and H. J. Ritter, "A review of tactile information: Perception and action through touch," *IEEE Transactions on Robotics*, vol. 36, no. 6, pp. 1619–1634, 2020.
- [9] B. Ward-Cherrier, N. Pestell, L. Cramphorn, B. Winstone, M. E. Giannaccini, J. Rossiter, and N. F. Lepora, "The tactip family: Soft optical tactile sensors with 3d-printed biomimetic morphologies," *Soft Robotics*, vol. 5, no. 2, pp. 216–227, 2018, PMID: 29297773. [Online]. Available: <https://doi.org/10.1089/soro.2017.0052>
- [10] R. Li, R. Platt, W. Yuan, A. ten Pas, N. Roscup, M. A. Srinivasan, and E. Adelson, "Localization and manipulation of small parts using gelsight tactile sensing," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2014, pp. 3988–3993.
- [11] D. Ma, E. Donlon, S. Dong, and A. Rodriguez, "Dense tactile force estimation using gelslim and inverse fem," in *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 5418–5424.
- [12] C. Zhao, J. Liu, and D. Ma, "ifem2.0: Dense 3-d contact force field reconstruction and assessment for vision-based tactile sensors," *IEEE Transactions on Robotics*, vol. 41, pp. 289–305, 2025.
- [13] M. Bauzá, O. Canal, and A. Rodriguez, "Tactile mapping and localization from high-resolution tactile imprints," *2019 International Conference on Robotics and Automation (ICRA)*, pp. 3811–3817, 2019. [Online]. Available: <https://api.semanticscholar.org/CorpusID:129945708>
- [14] 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)*, 2016, pp. 208–215.
- [15] W. Yuan, S. Wang, S. Dong, and E. Adelson, "Connecting look and feel: Associating the visual and tactile properties of physical materials," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.
- [16] J. Lin, Q. Hu, J. Xia, L. Zhao, X. Du, S. Li, Y. Chen, and X. Wang, "Non-destructive fruit firmness evaluation using a soft gripper and vision-based tactile sensing," *Computers and Electronics in Agriculture*, vol. 214, p. 108256, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0168169923006440>
- [17] S. Nam, T. Jack, L. Y. Lee, and N. F. Lepora, "Softness prediction with a soft biomimetic optical tactile sensor," *2024 IEEE 7th International Conference on Soft Robotics (RoboSoft)*, pp. 121–126, 2024. [Online]. Available: <https://api.semanticscholar.org/CorpusID:269775700>
- [18] A. Sipos and N. Fazeli, "MultiSCOPE: Disambiguating In-Hand Object Poses with Proprioception and Tactile Feedback," in *Proceedings of Robotics: Science and Systems*, Daegu, Republic of Korea, July 2023.
- [19] S. Zhong, D. Berenson, and N. Fazeli, "CHSEL: Producing Diverse Plausible Pose Estimates from Contact and Free Space Data," in *Proceedings of Robotics: Science and Systems*, Daegu, Republic of Korea, July 2023.
- [20] V. E. Arriola-Rios, P. Guler, F. Ficuciello, D. Kragic, B. Siciliano, and J. L. Wyatt, "Modeling of deformable objects for robotic manipulation: A tutorial and review," *Frontiers in Robotics and AI*, vol. 7, 2020.
- [21] V. E. Arriola-Rios and J. L. Wyatt, "A multimodal model of object deformation under robotic pushing," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 9, no. 2, pp. 153–169, 2017.
- [22] B. Frank, C. Stachniss, R. Schmedding, M. Teschner, and W. Burgard, "Learning object deformation models for robot motion planning," *Robotics and Autonomous Systems*, vol. 62, no. 8, pp. 1153–1174, 2014.
- [23] H. Shi, H. Xu, Z. Huang, Y. Li, and J. Wu, "Robocraft: Learning to see, simulate, and shape elasto-plastic objects with graph networks," *Robotics: Science and Systems (RSS)*, 2022.
- [24] K. Zhang, B. Li, K. Hauser, and Y. Li, "Adaptigraph: Material-adaptive graph-based neural dynamics for robotic manipulation," in *Proceedings of Robotics: Science and Systems (RSS)*, 2024.
- [25] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "Nerf: Representing scenes as neural radiance fields for view synthesis," *Communications of the ACM*, vol. 65, no. 1, pp. 99–106, 2021.
- [26] Y. Wi, P. Florence, A. Zeng, and N. Fazeli, "Virdo: Visuo-tactile implicit representations of deformable objects," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 3583–3590.
- [27] M. J. V. der Merwe, Y. Wi, D. Berenson, and N. Fazeli, "Integrated Object Deformation and Contact Patch Estimation from Visuo-Tactile Feedback," in *Proceedings of Robotics: Science and Systems*, Daegu, Republic of Korea, July 2023.
- [28] B. Shen, Z. Jiang, C. Choy, L. J. Guibas, S. Savarese, A. Anandkumar, and Y. Zhu, "Acid: Action-conditional implicit visual dynamics for deformable object manipulation," *Robotics: Science and Systems (RSS)*, 2022.
- [29] S. Zhong, A. Albini, O. P. Jones, P. Maiolino, and I. Posner, "Touching a nerf: Leveraging neural radiance fields for tactile sensory data generation," in *6th Annual Conference on Robot Learning*, 2022. [Online]. Available: <https://openreview.net/forum?id=No3mbanRIZJ>
- [30] Y. Dou, F. Yang, Y. Liu, A. Loquercio, and A. Owens, "Tactile-augmented radiance fields," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 26 529–26 539.
- [31] J.-C. Peng, S. Yao, and K. Hauser, "3d force and contact estimation for a soft-bubble visuotactile sensor using fem," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 5666–5672.
- [32] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Y. Bengio and Y. LeCun, Eds., 2015. [Online]. Available: <http://arxiv.org/abs/1412.6980>
- [33] S. Yao, Y. Zhu, and K. Hauser, "Structured bayesian meta-learning for data-efficient visual-tactile model estimation," in *8th Annual Conference on Robot Learning*, 2024.