MIT Open Access Articles

Shape-independent hardness estimation using deep learning and a GelSight tactile sensor

Massachusetts Institute of Technology

Shape-independent Hardness Estimation Using Deep Learning and a GelSight Tactile Sensor

Wenzhen Yuan[1], Chenzhuo Zhu[2], Andrew Owens[3], Mandayam A. Srinivasan[4] and Edward H. Adelson[5]

Abstract— Hardness is among the most important attributes of an object that humans learn about through touch. However, approaches for robots to estimate hardness are limited, due to the lack of information provided by current tactile sensors. In this work, we address these limitations by introducing a novel method for hardness estimation, based on the GelSight tactile sensor, and the method does not require accurate control of contact conditions or the shape of objects. A GelSight has a soft contact interface, and provides high resolution tactile images of contact geometry, as well as contact force and slip conditions. In this paper, we try to use the sensor to measure hardness of objects with multiple shapes, under a loosely controlled contact condition. The contact is made manually or by a robot hand, while the force and trajectory are unknown and uneven. We analyze the data using a deep constitutional (and recurrent) neural network. Experiments show that the neural net model can estimate the hardness of objects with different shapes and hardness ranging from 8 to 87 in Shore 00 scale.

I. INTRODUCTION

How can we tell the difference between a stone and sponge? How can we learn if a tomato is ripe? We touch them casually, and quickly learn their hardness. Humans learn a significant amount of information about the objects around them through touch [1], including hardness. Hardness is defined as the resistant force of a solid matter when a compressive force is applied, or in other words, the ratio between the displacement created by an indentation and the contact force. And indeed, hardness-measuring devices such as durometers generally work by measuring the indentation produced by a known force. Humans, however, seem to estimate hardness via a different procedure – one that involves observing the deformation of a fingertip passively [2]. In other words, they make heavy use of cutaneous sensing, rather than kinesthetic information.

In contrast, techniques for robots to estimate object hardness are limited. As the majority of tactile sensors measure force, a typical trial for a robot to measure hardness is by

[1]Department of Mechanical Engineering, and Computer Science and Artificial Intelligence Laboratory (CSAIL), MIT, Cambridge, MA 02139, USA yuan_wz@csail.mit.edu

[2] CSAIL, MIT, Cambridge, MA, USA and Department of Electronical Engineering, Tsinghua University, Beijing, 100084, China zhucz13@mails.tsinghua.edu.cn

[3]Electrical Engineering and Computer Science Department, U.C. Berkeley and CSAIL, MIT, Cambridge, MA 02139, USA andrewo@mit.edu

[4]Laboratory for Human and Machine Haptics (MIT TouchLab), Research Laboratory of Electronics and Department of Mechanical Engineering, MIT, Cambridge, MA 02139, USA and UCL TouchLab, Computer Science Department, UCL, London, UK srini@mit.edu

[5]Department of Brain and Cognitive Sciences and CSAIL, MIT, Cambridge, MA 02139, USA adelson@csail.mit.edu
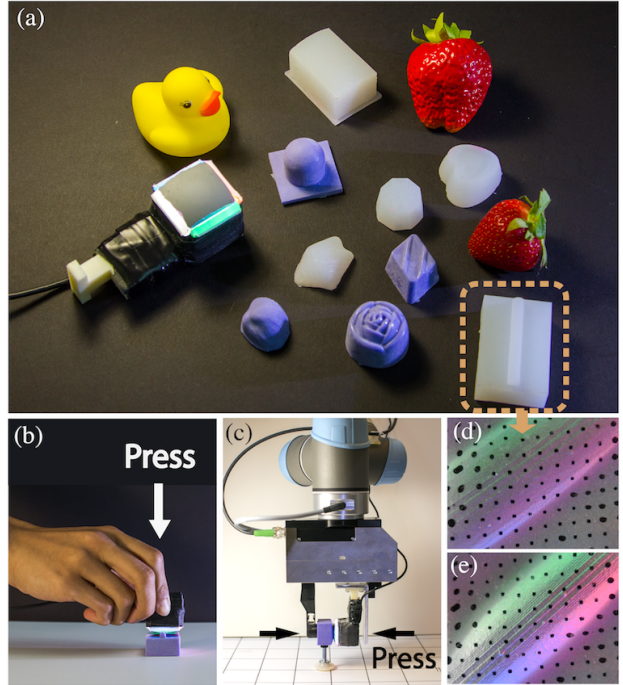
Fig. 1. GelSight sensor and test samples. GelSight is an optical tactile sensor that gets high resolution tactile images of contact surface topography and approximate contact force. (b) and (c) show GelSight contacting a deformable silicone sample: either a human tester presses the sensor on the object, or a robot gripper squeezes the object with the sensor. The right two pictures in the bottom shows the images captured by GelSight when contacting a soft cylinder and a hard cylinder: the color indicates the surface normal on the contact surface, the black dots are markers on the sensor surface, and their displacement is related to contact force.

measuring the force changes when contacting a sample with strictly controlled movement. The method is limited in that it requires strict control on both object geometry and robot manipulation.

Taking inspiration from human tactile sensing, we apply a soft tactile sensor called GelSight[3][4]. A GelSight sensor uses a thin piece of soft elastomer to contact external objects. An embedded camera takes pictures of the elastomer surface, which performs shading that varies with geometry under the specially designed optical system. The image shows the contact geometry with high-resolution, as well as general contact force. The sensor is similar to cutaneous sensing on human fingertips, in that it infers touch information from the deformation of the soft tissue.

When the sensor is pressed against an object, they both deform according to the hardness, which results in the difference in the surface curvature and contact force. We
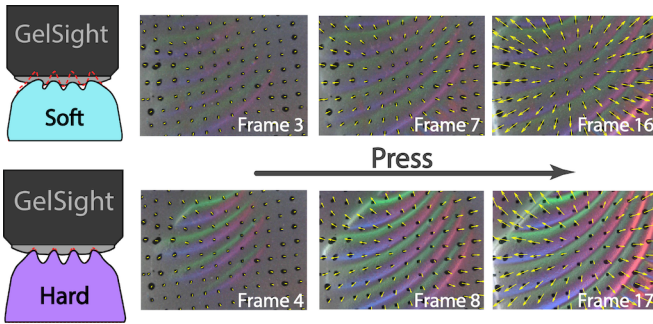
Fig. 2. The comparison of GelSight contacting a soft and a hard samples with the same shape. During the contact, soft object deforms more, with the ridges flattened, producing a smoother surface; hard sample deforms less, and ridges remain sharp. GelSight measures the 3D geometry of the surface, where larger height gradient causes bright color in the image, and the arrow field of black markers indicates contact force. We take a sequence of images during the contact procedure as our input signal.

showed in our previous work [5] that these physical changes could be measured using simple image cues: namely, changes in intensity and motion of the black markers embedded in the gel. Figure 2 shows an example.

While there is a strong physical motivation for these features, they make restrictive assumptions about the objects' geometry – namely, they assume that the objects are spherical. We'd like to use this sensor with a wider variety of object types. This, however, requires us to develop features that are applicable in more general settings. To address this problem, we take inspiration from recent work in computer vision that learns these image features in lieu of hand-crafted features. These methods, starting with the seminal work of Krizhevsky et al. [6], have recently achieved state-of-the-art performance on many recognition tasks. Here we apply these ideas to the domain of tactile perception by training a deep neural network to directly regress hardness directly from a raw GelSight video. To do this, we represent frames of the GelSight video using a convolutional neural network [7], and we use a recurrent neural network [8] to model changes in the gel deformation over time.

In this project, we use GelSight to contact objects in a loosely controlled condition, and measure the object hardness through the sequence of tactile images from GelSight. The sensor is either manually pressed on the samples, or mounted on a robot gripper which squeezes the samples. We train a neural network based on the sequences of tactile images during the presses on samples of known hardness, and predict the hardness of samples with unknown shapes. The test samples are made of silicone, with hardness that ranges from 8 to 87 in the Shore 00 scale, which is generally the hardness of gummy bears and pencil erasers. The samples are either of basic shapes including hemispheres and cylinders of 10 different radii, flat surface, edges and corners; or complicated shapes like everyday objects. We also test the model on a set of natural objects like tomatoes, candies, rubber tubes. Experimental results show that the model can well predict the silicone samples of basic shapes, regardless of the radii of

the hemispheres or cylinders. For natural objects, the model can estimate different hardness levels.

II. RELATED WORK

A. Optical Tactile Sensors and GelSight

Over the past decades, researchers have proposed many techniques for tactile sensing [9] in robotic applications. Among these, piezoresistive, peizoelectric, and capacitive sensors that measure the local force distribution have emerged as the dominant approaches. While tactile sensors based on these techniques have been used successfully in many applications, they often result in sensors that are difficult to fabricate, and which have limited spatial resolution or area. Optical-based tactile sensors [10], [11], [12], on the other hand, overcome many of these limitations. Most of those optical tactile sensors use a piece of deformable material as contact medium, and add some visually trackable patterns or markers to the medium. The sensors then infer the deformation of the contact medium by observing the deformation of the patterns, thereby measuring the contact force.

In this work, we consider the GelSight touch sensor – an optical tactile sensor that is designed to measure high-resolution shape of the contact interface. It consists of a clear elastomeric slab with a reflective membrane on the surface, with an embedded camera and illumination system [3], [4]. When the membrane is in contact with an object, one can obtain a highly accurate height map (accurate within microns in [4]) of its deformation via photometric stereo. This sensor has been used to measure physical properties in many applications. Jia, et al. [13], for instance, showed that it was more accurate than human subjects in detecting hard lumps in soft tissues, and Li [14] proposed a fingertip GelSight device, which is small enough to be mounted on a robot fingertip. Yuan et al. [15] added markers to the GelSight surface and provided a method for tracking their movement, thus enhancing its ability to infer the approximate contact force on the sensor.

B. Hardness Measurement

Robotics researchers have long been interested in estimating object properties through active touch. Drimus et al. [16] and Chu et al. [17], for example, recorded a sequence of contact force signals and then inferred object properties from change in force. These methods often implicitly include hardness in the signal. However, research on directly estimating hardness from a tactile sensor has been more limited.

For a robot, the most straightforward way to measure object hardness is by applying controlled force and measuring deformation, or vice versa [18]. Su et al. [19] measured the hardness of flat rubber samples using a BioTac touch sensor, which measures multiple tactile signals including the pressure on fixed points. In the work, the sensor was installed on a robot fingertip, and it is pressed onto flat silicone samples in a strictly controlled motion. Changes in the force re then used to discriminate between 6 samples with different hardness values ranging from 30 Shore00

rubber to rigid aluminum. For this method to be applicable, however, the sensor movement and object geometry must be strictly controlled. In contrast, our input is more general; we directly estimate hardness (rather than distinguishing between samples); and we use more diverse geometries and hardness.

Another specific way is to design a touch sensor that has special mechanisms for explicitly measuring both the local force and the depth of the press. For example, Shimizu et al. [20] designed a piezo-resistant cell with a gas-filled chamber, which was used to measure the indentation of the mesa on its top surface. The cell thus measures the material's hardness from the force measured by pressure change in the chamber and the indentation depth measured by piezo-resistance. The sensor makes measurement easier, under the condition that the surface geometry is certain. These limitations constrain the use of the sensor for more general touch tasks. There are also sensors designed specifically for measuring hardness for medical uses, such as [21][22], or with ultrasonic signals instead of force signals [23]. However, these sensors are not designed for measuring hardness of general objects, and are not well suited to other tactile tasks.

In our previous work [5], we showed that GelSight can accurately estimate the hardness of a set of hemispherical silicone samples under loosely controlled contact conditions. As in this work, we asked a human tester and an open-loop robot to press on samples, and developed a model to predict the hardness according to both the changing brightness in the GelSight image and the displacement field of the surface markers. While the model successfully predicted the hardness, it was limited to hemispherical samples, and it required that the radii be provided as input.

C. Learning for tactile sensing

Recently, researchers have proposed several learning-based methods that estimate tactile properties from touch sensors. Gao et al. [24], for example, used a neural network to infer haptic adjectives (i.e., qualitative properties such as "bumpy" or "squishy") from a Biotac sensor. While their technical approach also uses a recurrent neural network, specifically a 1D convolutional network and recurrent network, the output of the sensor– an assortment of 32 time-varying physical measurements, such as fluid pressure and temperature – is substantially more limited. In contrast, our method starts with an input that is significantly lower level – namely, images of a deforming gel. From this rich signal, we infer metric hardness measurements, rather than qualitative properties.

III. ESTIMATING HARDNESS FROM GELSIGHT VIDEOS

Figure 2 shows the examples of videos recorded by a GelSight sensor. Here, the differences between hard and soft objects are readily apparent: a hard object deforms little during the contact, while a softer object may deform largely and makes a flatter surface under relatively smaller force. These subtle deformations in shape and force, in turn, are visible in the videos recorded by GelSight. As in [5], the
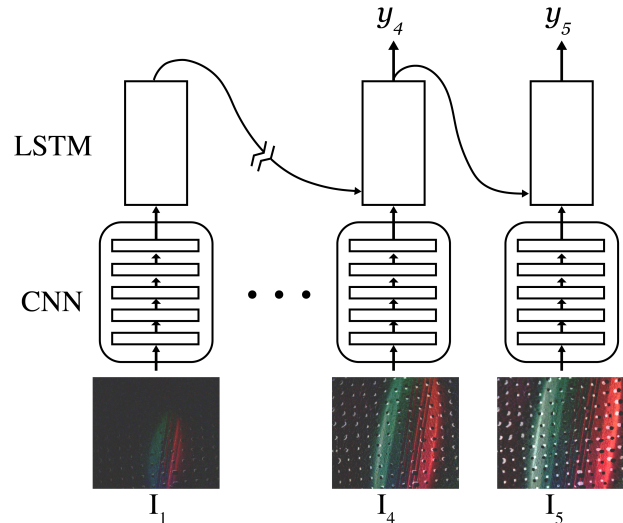

Fig. 3. We use a recurrent neural network to map a video recorded by the GelSight sensor to a hardness value. The network resembles [25]: images are represented using CNN features *fc7* from a VGG16 net, and feed into an LSTM net. The net input I_t is a sequence of GelSight image after subtracting the initial frame. In particular, we choose 5 frames evenly from a pressing sequence, and each frame is of one pressing stage. We only use the RNN output of the last 3 frames, i.e. y_3, y_4, y_5 to estimate the sample hardness.

GelSight video shows the surface normal of the contact surface through the inner surface's reflection change under the colored and directional light sources, which results in different color intensity. Furthermore, the black markers embedded on the surface help convey the contact force through their motion.

A. Neural network design

We ask in this work whether we can use these cues to predict an arbitrary object's hardness. To do this, we use a neural network that maps an image sequence to a scalar hardness value (measured on the Shore 00 scale). The design of the network is similar to recurrent networks considered in action recognition [25], and is shown in Figure 3. We represent each GelSight image I_i with convolutional network features $\phi(I_i)$. For these, we use the penultimate layer (*fc7*) of the VGG architecture[7]. We then use a recurrent neural network with long short-term memory units (LSTM) [8] to model temporal information.

At each timestep, we regress its output hardness value via an affine transformation of the current LSTM's hidden state h_t:

$$\begin{aligned} y_t &= Wh_t + b \\ h_t &= L(h_{t-1}, \phi(I_t)), \end{aligned} \quad (1)$$

where W and b define an affine transformation of the hidden state h_t, and L updates h_t based on the previous state h_{t-1} using the current image I_t (here we omit the LSTM's hidden cell state for simplicity).

The prediction y_t is the hardness estimate for the current timestep. We estimate a hardness value for the object as a whole by averaging the predictions from the final 3 frames.

We perform the regression on a per-frame basis to add robustness to videos in which the pressing motion differs significantly from those of the training set. During training, we minimize a loss that penalizes the difference between the predicted and ground-truth hardness values, using a Huber loss.

B. Choosing input sequences

We'd like to make our method invariant to the speed of the pressing motion and to the maximum contact force. For example, different human testers or robots may manipulate objects with different loading speed or maximum force. Therefore, we constrain the video sequence so that it begins and ends at times that are consistent across manipulation conditions. Specifically, we choose a 5-frame sequence of images in the loading period of the press, such that the sequence starts after the object being pressed. We determine this starting point by finding the frame in which the mean intensity of the GelSight image (a proxy for the force of the press) exceeds a threshold. For the end point of the sequence, we choose the last frame whose intensity change is the peak of the sequence. The other 3 frames are evenly chosen in the middle according to the intensity change.

We subtract the first frame in the sequence from the chosen frames to account for preexisting deformations in the elastomer. Unlike in [5], we do not explicitly model the motion of the markers that are embedded in the gel – relying instead on the network to learn about their motion via the raw images (in which the motion of the marker is implicitly visible due to subtracting the initial frame).

C. Training

The training dataset contains about 7000 videos obtained by having human testers press GelSight on different silicone samples, each video is an independent pressing sequence. The training dataset contains mainly the basic object shapes(Group 1 in Figure 5), but also with a large portion of complicated shapes or bad contact conditions(Group 2 and 4 in Figure 5). Those irregular data greatly help to prevent overfit of the model. A single video is used for multiple times during the training, with different end point for sequence extract, so that the contact situations with different maximum forces are included. In other words, when choosing a sequence ends at the middle of the loading process, the sequence equals to the case of pressing on the same object in the same way, but with smaller maximum force. We train the model using stochastic gradient descent, initializing the CNN weights with ImageNet [6] pretraining, jointly training the CNN and LSTM. We train the algorithm for 10,000 iterations, a learning rate of 0.001, and a step size of 1000.

IV. EXPERIMENTAL SETUP

The touch sensor we used is a Fingertip GelSight device [14], with black markers on the surface to track the displacement field. The elastomer on GelSight is of the dome shape, with the area of $25\text{mm} \times 25\text{mm}$, and the maximum
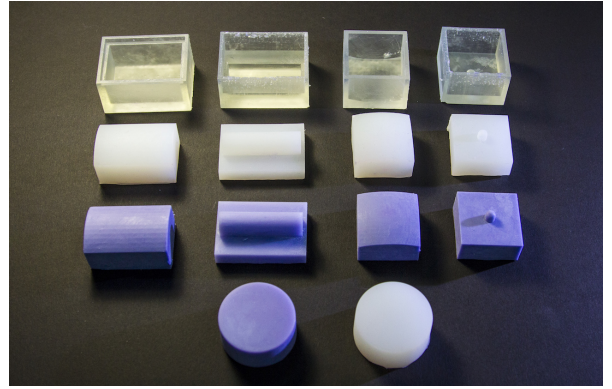

Fig. 4. Some 3D printed molds and silicone samples in hemispherical and cylindrical shapes. Samples are of different radii and hardness, and hardness is controlled by changing silicone ingredient mixing ratio. The material hardness is tested by a durometer on standardized silicone samples, as the bottom two samples.

thickness in the center is around 2.4mm. The average interval between the black markers is about 1.3 mm. The hardness of the elastomer is 17 in Shore 00 scale. The camera in the sensor takes original images of size 960×720 pixels over an area of $18.4\text{mm} \times 13.8\text{mm}$ at a speed of 30Hz.

The silicone samples that we used in our experiments have the varied shapes of 1) hemispheres of different radii, 2) cylinders of different radii, and 3) some arbitrary shapes. The silicone materials were Ecoflex® 00-10, Ecoflex® 00-50 and Smooth-Sil® 945 from Smooth-On Inc. The three raw materials are mixed by different ratio to make the samples into 16 different hardness levels between 8 in Shore 00 scale to 45 in Shore A scale (equals to 87 in the Shore 00 scale). The silicone materials are in liquid phase, and they will solidify in several hours after Part A and Part B are mixed and poured into the mold. The hemispherical and cylindrical samples have 9 different radii ranging from 2.5mm to 50mm, but they are of similar height (around 25mm). This design is to ensure the sample thickness would have limited influence of on the shape change during the press. We fabricate the mold with a Form 2 3D printer from Formlabs. For the arbitrary shapes, one group is of simple and common shapes, that is casted from some daily vessels, like square shaped ice box, truncated cone shaped measuring cups, small beakers; another group are of complicated and special shaped, casted from assorted chocolate molds. They include the shapes of different emboss textures or complicated curvatures, like the shape of shells. In total, we made 95 hemispherical samples, 81 cylindrical samples, 15 flat samples and 160 samples of arbitrary shapes for experiments. We also collect a set of 20 rigid objects in the database.

The sample hardness is measured by a PTC® 203 Type OO durometer. We made a set of flat thick silicone samples(the bottom two samples in Figure 4) of each mixing portion as the test sample from durometer test. To reduce the measurement error, we took 5 tests and use the mean value for each test sample . The real samples are considered as of the same hardness of the samples being tested.
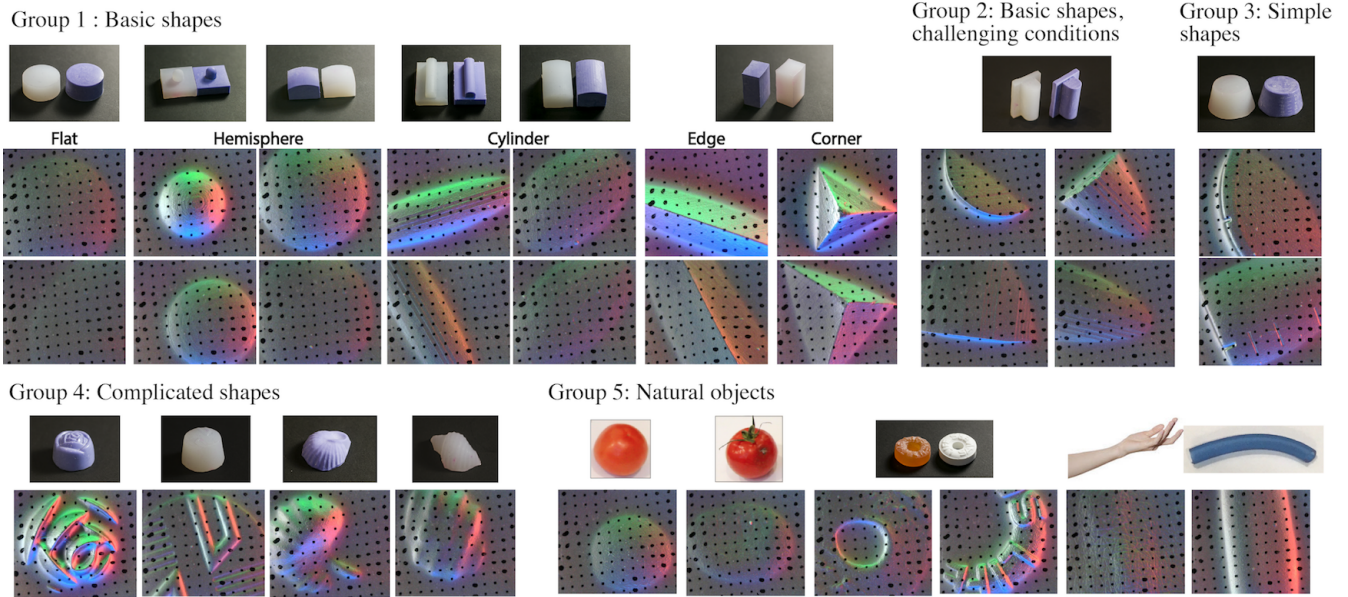
Fig. 5. Examples of GelSight dataset. Group 1 are of the basic shapes, where the second row is of the tactile images of softer samples with same shape; Group 2 are of contacting the same set of samples, but in a undesired contact condition; Group 3 are the same of some natural object shapes but with simple geometry; Group 4 are complicated shapes made by chocolate molds; Group 5 are natural objects, and in this figure the left two are from tomatoes, right two are from candies.

The hardness of the elastomer on the GelSight sensor may influence the sensitive range of the sensor. The sensor can better estimate the objects' hardness when it is closer to the sensor's hardness. In this project, we choose a soft elastomer of 17 in Shore 00 scale, in order to get a better discrimination of the very soft objects, but the sensor cannot well differentiate the hardness when it is harder than 70 in Shore 00 scale.

V. EXPERIMENTAL PROCEDURE

When we press the GelSight sensor into a sample object, we push it in the normal direction. We conducted two sets of experiments: human testers pressing the sensor, and a robot gripper squeezing on the samples. In the human testing scenario, the test object is placed on a flat hard surface, and a tester holds the GelSight sensor and presses on the object vertically, as shown in bottom left figure in Figure 1. For the robot test, we used a Weiss WSG 50 gripper which has GelSight as one finger. The robot gripper closes in a slow and constant speed until the gripping force reaches the threshold, making the GelSight sensor squeezing on the object. The speed was randomly chosen between 5 to 7 mm/s, and the gripping force threshold is a random between 5 to 9N. The procedure is shown in the bottom middle figure in Figure 1. In both cases, the sensor is pressed into the object, the contact force grows, and the deformation of both the GelSight elastomer and the object increases. We record the GelSight video during the press. In average, there are 20 to 30 frames in the pressing period.

Note that in both experimental sets, the contact condition is highly varied. In the human tester experiment, the trajectory and pressing velocity are both unknown and uncontrolled, and small amount of shear force and torque exists in the contact. Our goal was to model a "natural" tactile interaction – similar to what one would do in daily environments, or when a robot with a touch sensor contacting an arbitrary object in a complicated environment.

The collected GelSight data are of the following geometry types, as shown in Figure 5:

1) *Basic shapes*. Samples of the simplest geometries, including flat surface, spherical surface, cylindrical surface, straight sharp edges, sharp corners. Spherical and cylindrical shapes are of 10 levels of radii from 2.5mm to 50mm.
2) *Bad basic shapes, challenging contact conditions*. The contact objects are the same as in the previous group, but the contact condition is undesired. For example, the sensor contacts the silicone samples in tilted angles, or the sample is included in the contact area.
3) *Simple shapes*. The samples are made of silicone with known hardness and basic shape, such as shapes of frustum measuring cup.
4) *Complicated shapes*. The samples are made of silicone but with complicated textures or shapes. They are made from the chocolate molds.
5) *Soft objects in everyday life*. These are the soft objects in the everyday life, mostly with relatively simple shapes. Human can roughly feel whether they are 'soft', or 'very soft', or 'hard'.

We selected some samples in Group 1, Group 2, and Group 4 as the training set, and tested the model's prediction on Group 1, 3, 4 and 5. The data in the training set is the data collected by human testers; in the test set, some data is collected by human testers, while data is collected

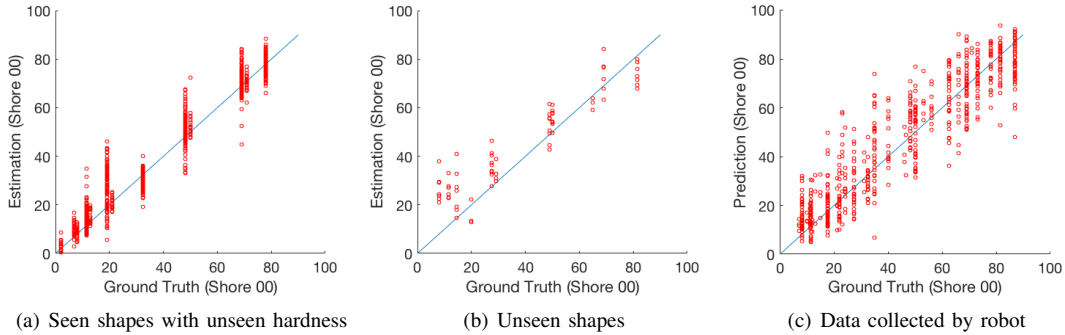| (a) Seen shapes with unseen hardness | (b) Unseen shapes | (c) Data collected by robot |

Fig. 6. Prediction results on samples of basic shapes, including hemispheres, cylinders, flat surface, edges, corners.

TABLE I

NETWORK PREDICTION ON BASIC SHAPES

	number of videos	R^2	RMSE
Trained shape, novel hardness	1398	0.9564	5.18
Novel shapes	73	0.7868	11.05
Trained samples, robot gripper	683	0.8524	10.28

by a robot. The dataset is published at `http://people.csail.mit.edu/yuan_wz/hardnessdataset/`.

VI. EXPERIMENTAL RESULTS

A. Basic Shapes

In the first experiment, we wanted to test whether a model could generalize to new hardness values. Therefore, the objects in the test set had the same shape as those in the training set (Group 1 in Figure 5 and Section V]), but with different hardness ratings. In the second experiment, the samples are of cylindrical or spherical shapes, but are not included in the training set; in the third group, the samples are of the Group 1 basic shapes that have been seen, but the experiment is conducted by the robot, which makes different contact motion than human testers. The estimation result of the three groups is shown in Figure 6 and Table I.

B. Arbitrary Shapes

In this challenging experiment, we test our model on the arbitrary silicone shapes that the algorithm has not seen during training, as mentioned as Group 3 and Group 4 in Section V and Figure 5. For Group 3 samples, the R^2 of measurement is 0.57, and RMSE is 19.3; for Group 4 samples, the R^2 decreases to 0.39, and the RMSE is 18.2. In most large-error measurement cases, the model tends to estimate the hardness much higher than the ground truth. In these cases, the objects mostly have some sharp surface curvatures but not the same as as those in the training set. Moreover, there may be multiple ridges present in a sample. The network tends to consider the object of higher hardness when there are sharp curvatures on the contact surface. On the other hand, for the object shape that is included in the training set, the neural net can well estimate the target hardness.

when the training set contains basic shapes, the model can generalize to more complex shapes – albeit with decreased performance. For the method of deep neural network, a tough learning through all the possible data is required. To make the network more generalized in estimating the hardness for objects with more arbitrary shapes, a much larger training set that contains more changes in the sample shapes is required.

C. Estimation of Natural Objects

We use GelSight to contact natural objects and use the model to estimate the hardness of them. As it is difficult to measure their ground-truth hardness with standard method, we ask human subjects to rank the hardness of similar objects, and compare the readings from GelSight measurement. We compared the estimated hardness of several plum tomatoes and round tomatoes with different ripeness, as shown in Figure 7. Although the tomatoes are of close hardness, the GelSight prediction is similar to human estimation. The GelSight measurement of different candies, elastomer tubes, and some random daily objects are shown in Figure 8.

From the figures we can see that, for natural objects of simple geometry and smooth surfaces, our network can well estimate their hardness level. The estimation can be used to differentiate ripeness levels of some fruits, like tomatoes. Rigid objects will be estimated with a number larger than 80. However, similar to the experiments in Section VI-B, the geometry of the natural objects may also have an influence on the measurement, in that the objects with textures, like a ridges surface, tend to be estimated harder. The reason is mainly due to the incompleteness of the training set for the network.

VII. CONCLUSION

In this paper, we proposed a deep neural network model to estimation hardness of objects with multiple shapes using a GelSight tactile sensor. We mainly considered the commonly seen shapes of natural objects, including spheres, cylinders, flat surface, edges and corners. The GelSight sensor records the objects' deformation during the contact process in high-resolution image sequences, which contain information about both the shape change and contact force. The biggest challenge lies in that the different geometry of the object may have very complicated influence on the sensor output, and is
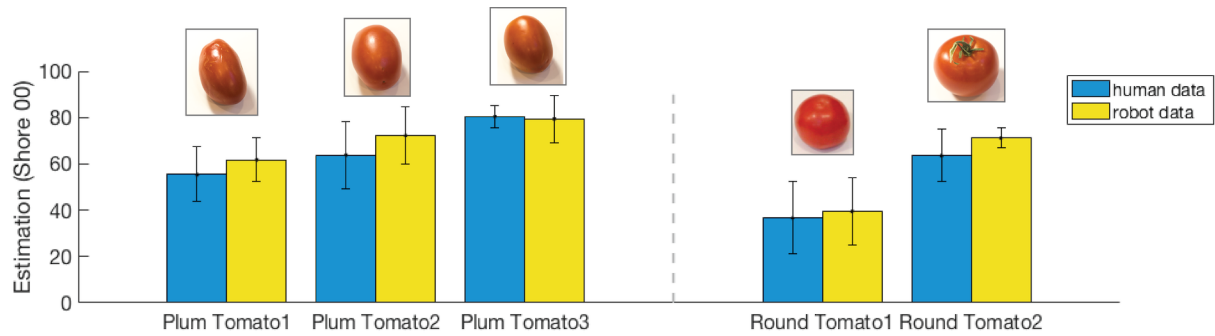
Fig. 7. GelSight estimation of tomato hardness, experiments executed by either human or robot. For each tomato shape, the display ranking is according to human estimation: larger index indicates harder and rawer tomatoes. In average, each tomato is pressed for 18 times on different parts, and the hardness on different parts is slightly different.
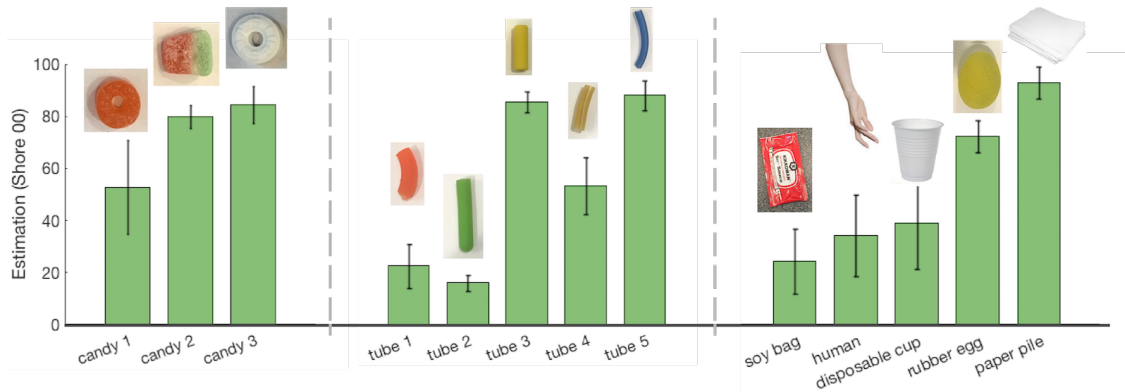

Fig. 8. Measurement of some natural objects. In average, each object is pressed on for 5 times by a human tester. In each group, the object order on x axis corresponds to human ranking of hardness. In the first group, candy 1 is a soft gummy candy, candy 2 is a hard gummy candy with rigid sugar particles on the surface, candy 3 is a rigid candy. Hardness of candy 2 is over estimated because it has dense particles on the surface. Similarly, in the 2nd group, hardness of tube 1 and tube 3 is overestimated because they have textures, like a ridged surface.

nearly impossible to model with linear relationships. To solve this problem, we apply a convolutional neural network and recurrent neural network to extract the hardness information from the GelSight video sequence. Experimental results show that the network can well predict hardness of silicone samples with similar shape in the dataset, regardless of the loading conditions; for the objects with complicated shapes or ridged surface, the model does not estimate the hardness well, which caused by the incompleteness of the training data set. For many natural objects with smooth surface and simple geometries, the model can roughly measure their hardness level, thus can be used for recognize objects with special hardness, or choose fruits with preferable ripeness level.

ACKNOWLEDGMENT

REFERENCES

[1] S. J. Lederman and R. L. Klatzky, "Extracting object properties through haptic exploration," *Acta psychologica*, vol. 84, no. 1, pp. 29–40, 1993.

[2] M. A. Srinivasan and R. H. LaMotte, "Tactual discrimination of softness," *Journal of Neurophysiology*, vol. 73, no. 1, pp. 88–101, 1995.

[3] M. K. Johnson and E. Adelson, "Retrographic sensing for the measurement of surface texture and shape," in *Computer Vision and Pattern Recognition (CVPR), 2009 IEEE Conference on*. IEEE, 2009, pp. 1070–1077.

[4] M. K. Johnson, F. Cole, A. Raj, and E. H. Adelson, "Microgeometry capture using an elastomeric sensor," in *ACM Transactions on Graphics (TOG)*, vol. 30, no. 4. ACM, 2011, p. 46.

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

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[7] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[8] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[9] R. S. Dahiya, G. Metta, M. Valle, and G. Sandini, "Tactile sensingfrom humans to humanoids," *IEEE Transactions on Robotics*, vol. 26, no. 1, pp. 1–20, 2010.

[10] N. J. Ferrier and R. W. Brockett, "Reconstructing the shape of a deformable membrane from image data," *The International Journal of Robotics Research*, vol. 19, no. 9, pp. 795–816, 2000.

[11] K. Kamiyama, K. Vlack, T. Mizota, H. Kajimoto, K. Kawakami, and S. Tachi, "Vision-based sensor for real-time measuring of surface

traction fields," *IEEE Computer Graphics and Applications*, vol. 25, no. 1, pp. 68–75, 2005.

[12] C. Chorley, C. Melhuish, T. Pipe, and J. Rossiter, "Development of a tactile sensor based on biologically inspired edge encoding," in *Advanced Robotics, 2009. ICAR 2009. International Conference on*. IEEE, 2009, pp. 1–6.

[13] X. Jia, R. Li, M. A. Srinivasan, and E. H. Adelson, "Lump detection with a gelsight sensor," in *World Haptics Conference (WHC), 2013*. IEEE, 2013, pp. 175–179.

[14] 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 *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*. IEEE, 2014, pp. 3988–3993.

[15] W. Yuan, R. Li, M. A. Srinivasan, and E. H. Adelson, "Measurement of shear and slip with a gelsight tactile sensor," in *Robotics and Automation (ICRA), 2015 IEEE International Conference on*. IEEE, 2015, pp. 304–311.

[16] A. Drimus, G. Kootstra, A. Bilberg, and D. Kragic, "Design of a flexible tactile sensor for classification of rigid and deformable objects," *Robotics and Autonomous Systems*, vol. 62, no. 1, pp. 3–15, 2014.

[17] V. Chu, I. McMahon, L. Riano, C. G. McDonald, Q. He, J. Martinez Perez-Tejada, M. Arrigo, N. Fitter, J. C. Nappo, T. Darrell, *et al.*, "Using robotic exploratory procedures to learn the meaning of haptic adjectives," in *Robotics and Automation (ICRA), 2013 IEEE International Conference on*. IEEE, 2013, pp. 3048–3055.

[18] W. M. B. Tiest and A. M. Kappers, "Cues for haptic perception of compliance," *IEEE Transactions on Haptics*, vol. 2, no. 4, pp. 189–199, 2009.

[19] Z. Su, J. A. Fishel, T. Yamamoto, and G. E. Loeb, "Use of tactile feedback to control exploratory movements to characterize object compliance," *Active Touch Sensing*, p. 51, 2014.

[20] T. Shimizu, M. Shikida, K. Sato, and K. Itoigawa, "A new type of tactile sensor detecting contact force and hardness of an object," in *Micro Electro Mechanical Systems, 2002. The Fifteenth IEEE International Conference on*. IEEE, 2002, pp. 344–347.

[21] S. Omata and Y. Terunuma, "New tactile sensor like the human hand and its applications," *Sensors and Actuators A: Physical*, vol. 35, no. 1, pp. 9–15, 1992.

[22] O. Lindahl, S. Omata, and K.-A. Ängquist, "A tactile sensor for detection of physical properties of human skin in vivo," *Journal of medical engineering & technology*, vol. 22, no. 4, pp. 147–153, 1998.

[23] M. Takei, H. Shiraiwa, S. Omata, N. Motooka, K. Mitamura, T. Horie, T. Ookubo, and S. Sawada, "A new tactile skin sensor for measuring skin hardness in patients with systemic sclerosis and autoimmune raynaud's phenomenon," *Journal of international medical research*, vol. 32, no. 2, pp. 222–231, 2004.

[24] Y. Gao, L. A. Hendricks, K. J. Kuchenbecker, and T. Darrell, "Deep learning for tactile understanding from visual and haptic data," *arXiv preprint arXiv:1511.06065*, 2015.

[25] J. Donahue, L. A. Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell, "Long-term recurrent convolutional networks for visual recognition and description," *CVPR*, 2015.

[26] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural computation*, vol. 1, no. 4, pp. 541–551, 1989.