# Hardness Similarity Detection Using Vision-Based Tactile Sensors

Alap Kshirsagar[1], Frederik Heller[1], Mario Gómez Andreu[1], Boris Belousov[2], Tim Schneider[1], Lisa P. Y. Lin[3], Katja Doerschner[3], Knut Drewing[3] and Jan Peters[1,2,4,5]

***Abstract*— Humans can classify deformable materials according to their hardness similarity, but existing robotic approaches focus on hardness recognition or absolute hardness prediction. In this work, we investigate hardness similarity detection using a vision-based tactile sensor (VBTS) and evaluate three methods: optical flow features and support vector machine (SVM) classifier, DINOv2 features and SVM classifier, and convolutional long short term memory (ConvLSTM) network trained with categorical cross-entropy loss. To evaluate these methods, we created a dataset of over 200 videos by pressing a GelSight Mini sensor, attached to a Franka-Panda robot, on five silicone objects of varying hardness, and also conducted a human-participant study showing humans achieved 80.25% average accuracy in hardness similarity detection. The three methods achieved average accuracies of 77.66%, 67.00%, and 70.00% with 15 samples per object, demonstrating that a VBTS can effectively classify objects based on hardness similarity.**

## I. INTRODUCTION

Touch sensing is essential for humans and robots, enabling the perception of object properties and precise manipulations. One of the most important properties perceived through touch is *hardness*. The ability to detect an object's hardness can help robots to effectively perform several tasks such as sorting waste, harvesting fruits, and handling delicate objects. Prior works have addressed tactile hardness detection through two problems: Hardness prediction, which estimates the absolute hardness of a test object (e.g., on the Shore hardness scale), and hardness recognition, which identifies a reference hardness class with the same hardness as the test object. We consider the problem of hardness similarity detection, where an agent selects the reference object closest in hardness to a given test object from a set of reference objects with varying hardness levels.

Prior works on the hardness prediction task [1]–[3] used labelled datasets of thousands of tactile samples with corresponding ground-truth absolute hardness values and trained a deep neural network to regress the absolute hardness of objects. However, given the variety of tactile sensors, creating such labelled datasets is difficult. Prior works on the hardness recognition task [4]–[7] have used piezoelectric sensing patches [4], pressure sensors [5], [6], or tactile array sensor [7] to collect 50–100 tactile samples from a set of reference objects and trained classifiers with these samples. In this work, we investingate the effectiveness of a Vision-Based Tactile Sensor (VBTS) for the task of sample-efficient hardness similarity detection. VBTSs provide a cost-effective and high-resolution alternative to traditional tactile sensors. They also allow leveraging the advancements in camera technology and computer vision.

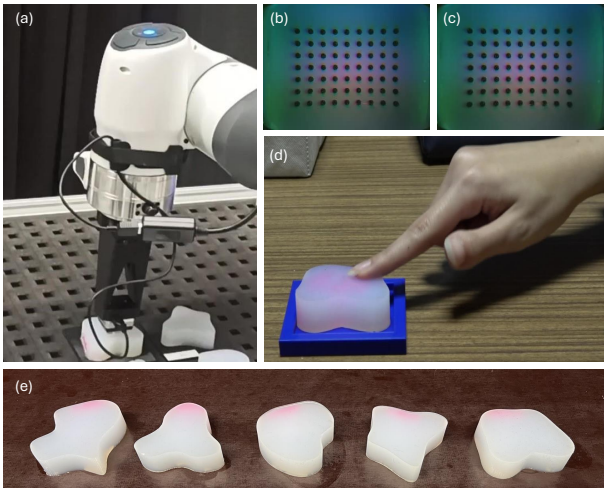



Fig. 1. (a) In our robot experiments, a Franka-Emika Panda robot uses a GelSight Mini sensor mounted on the robot's end-effector to explore hardness of objects. (b) The image captured by GelSight Mini sensor before making contact with the object. (c) The image captured by GelSight Mini sensor after contact with the object. (d) In our human participant experiments, the humans explored hardness of objects using their index finger. (e) The set of sample objects used in our experiments: five silicone objects of different shape in increasing hardness from left to right.

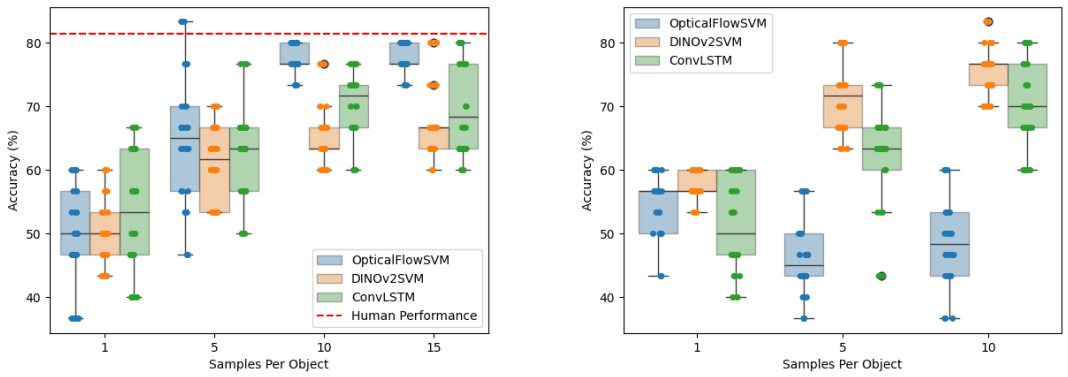
## II. HARDNESS SIMILARITY DETECTION WITH VISION-BASED TACTILE SENSORS

When a VBTS like the GelSight Mini is pressed against a deformable object, changes in color intensity and displacement of embedded markers on the gelpad (Fig. 1b-c) reflect the object's hardness, with the recorded videos also encoding temporal information. We investigate three methods for the task of hardness similarity detection from VBTS videos.

### *Method 1: Optical Flow and Support Vector Machine (SVM) Classifier*

Optical flow aims to capture the apparent motion of each patch of pixels between subsequent frames. We utilize the Shi-Tomasi corner detector [8] to select points of interests, i.e. dots on the gelpad, and apply the optical flow estimation method proposed by Lucas and Kanade [9]. We concatenate

[1]Intelligent Autonomous Systems Lab, Department of Computer Science, TU Darmstadt, Germany, alap@robot-learning.de
[2]German Research Center for AI (DFKI)
[3]Department of Psychology, University of Giessen, Germany
[4]Centre for Cognitive Science, Technical University of Darmstadt
[5]Hessian Center for Artificial Intelligence (Hessian.AI), Darmstadt
We thank Hessisches Ministerium für Wissenschaft & Kunst for the DFKI grant and "The Adaptive Mind" grant.




(a) Evaluation on our dataset

(b) Evaluation on Yuan et al. [1]'s dataset

Fig. 2. Boxplots displaying the hardness similarity detection accuracies from 10 evaluation rounds (30 trials in each round). Each trial consists of two reference objects and one test object. The colored dots indicate the individual accuracies for each of the 10 rounds per method. The red dotted line shows the average human performance on the objects in our dataset.

the optical flow features for each frame in a video and train a Support Vector Machine (SVM) classifier on the videos from the reference objects. We then use the trained SVM classifier to predict the hardness class of videos collected from the test object and apply majority voting to assign the test object to the closest reference object.

*Method 2: DINOv2 Features and SVM Classifier*

Instead of employing pre-defined feature representations, such as optical flow, we can utilize pretrained large-scale vision foundational models to extract learned features. In this method, we utilize the DINOv2 [10] model for feature extraction from VBTS videos. DINOv2 is a self-supervised model utilizing Vision Transformer architecture, trained to extract high-dimensional features from images through self-distillation and clustering techniques. Similar to method 1, we train an SVM classifier on the concatenated DINOv2 features extracted from the VBTS videos of reference objects.

### *A. Method 3: Convolutional Long Short Term Memory (ConvLSTM) Network*

In this method, we use a deep neural network based on the ConvLSTM architechure to extract features from VBTS videos. A ConvLSTM [11] network combines the strengths of Convolutional Neural Network (CNN)s and Long Short Term Memory (LSTM) networks, enabling the model to handle spatiotemporal data by capturing both spatial features and temporal dependencies. We design a neural network of four stages, each consisting of a ConvLSTM layer, a MaxPooling layer as well as a Dropout of 20% on the recurrent state during training. To generate class probabilities, we flatten the output of the last stage and feed it into a dense layer with softmax activation. We train this network using categorical cross entropy loss for 100 epochs on the videos collected from the reference objects. In contrast to Method 1 and Method 2, this method learns the feature extraction and classification steps simultaneously.

## III. EVALUATION

We evaluate the methods described in Section II for the task of hardness similarity detection. For our experiments, we consider the following scenario: the agent needs to select the reference object closest in hardness to a test object from a set of two reference objects. We created a set of five silicone objects, shown in Fig. 1e, of varying hardness levels. As shown in Fig. 1a, we mounted a GelSight Mini sensor on a robot arm to record videos of the sensor pressing on object surfaces with varying velocities and forces. We conduct three sets of experiments: first, evaluating the three methods on our dataset, second, a human-participant experiment to assess the human performance on the same task [12], and third, evaluating the three methods on a dataset collected by Yuan et al. [1], specifically on samples from the cuboid objects.

Fig. 2 shows the accuracies the of the methods for different sample sizes. We find that the optical flow features and SVM classifier outperform other methods on our dataset, but none surpass the human performance level. We also observed that all three methods and the human participants perform poorly when the hardness of the test object is between the hardness levels of the reference objects. Surprisingly, on Yuan et al. [1]'s dataset, DINOv2 features and SVM classifier performs the best. One possible explanation is that Yuan et al.'s dataset has more color intensity fluctuations and edges, leading to better DINOv2 features.

## IV. CONCLUSION AND FUTURE WORK

We evaluated three methods for the task of hardness similarity detection with a VBTS: a traditional video classification method, a deep-learning method, and a vision foundation model. Our evaluations on two datasets demonstrate the strengths and weaknesses of these methods. In the future, we will evaluate the methods in multi-class hardness similarity detection scenarios. Also, we will investigate active sampling strategies (e.g. entropy and variance strategies from [13]) for improving the sample efficiency of these methods.

## REFERENCES


[1] W. Yuan, C. Zhu, A. Owens, M. A. Srinivasan, and E. H. Adelson, "Shape-independent hardness estimation using deep learning and a gelsight tactile sensor," in *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 951–958, 2017.

[2] Y. Chen, J. Lin, X. Du, B. Fang, F. Sun, and S. Li, "Non-destructive fruit firmness evaluation using vision-based tactile information," in *International Conference on Robotics and Automation (ICRA)*, pp. 2303–2309, 2022.

[3] S. Nam, T. Jack, L. Y. Lee, and N. F. Lepora, "Softness prediction with a soft biomimetic optical tactile sensor," in *IEEE International Conference on Soft Robotics (RoboSoft)*, pp. 121–126, 2024.

[4] Y. Amin, C. Gianoglio, and M. Valle, "Embedded real-time objects' hardness classification for robotic grippers," *Future Generation Computer Systems*, vol. 148, pp. 211–224, 2023.

[5] S. Fang, Z. Yi, T. Mi, Z. Zhou, C. Ye, W. Shang, T. Xu, and X. Wu, "Tactonet: Tactile ordinal network based on unimodal probability for object hardness classification," *IEEE Transactions on Automation Science and Engineering*, 2022.

[6] X. Qian, E. Li, J. Zhang, S.-N. Zhao, Q.-E. Wu, H. Zhang, W. Wang, and Y. Wu, "Hardness recognition of robotic forearm based on semi-supervised generative adversarial networks," *Frontiers in Neurorobotics*, vol. 13, p. 73, 2019.

[7] Z. Zhang, J. Zhou, Z. Yan, K. Wang, J. Mao, and Z. Jiang, "Hardness recognition of fruits and vegetables based on tactile array information of manipulator," *Computers and Electronics in Agriculture*, vol. 181, p. 105959, 2021.

[8] J. Shi *et al.*, "Good features to track," in *IEEE conference on Computer Vision and Pattern Recognition*, pp. 593–600, 1994.

[9] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *International Joint Conference on Artificial Intelligence*, vol. 2, pp. 674–679, 1981.

[10] M. Oquab, T. Darcet, T. Moutakanni, H. V. Vo, M. Szafraniec, V. Khalidov, P. Fernandez, D. HAZIZA, F. Massa, A. El-Nouby, *et al.*, "Dinov2: Learning robust visual features without supervision," *Transactions on Machine Learning Research*, 2023.

[11] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-c. Woo, "Convolutional lstm network: A machine learning approach for precipitation nowcasting," *Advances in Neural Information Processing Systems*, vol. 28, 2015.

[12] L. P. Y. Lin, A. Böhm, B. Belousov, A. Kshirsagar, T. Schneider, J. Peters, K. Doerschner, and K. Drewing, "Task-adapted single-finger explorations of complex objects," in *Eurohaptics Conference*, 2024.

[13] A. Böhm, T. Schneider, B. Belousov, A. Kshirsagar, L. Lin, K. Doerschner, K. Drewing, C. A. Rothkopf, and J. Peters, "What matters for active texture recognition with vision-based tactile sensors," in *IEEE International Conference on Robotics and Automation*, 2024.