Active Sampling for Hardness Classification with Vision-Based Tactile Sensors

Junyi Chen[1], 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— **Hardness is a key tactile property perceived by humans and robots. In this work, we investigate information-theoretic active sampling for efficient hardness classification using vision-based tactile sensors. We assess three probabilistic classifiers and two uncertainty-based sampling strategies on a robotic setup and a human-collected dataset. Results show that uncertainty-driven sampling outperforms random sampling in accuracy and stability. While human participants achieve** 48.00% **accuracy, our best method reaches** 88.78% **on the same objects, highlighting the effectiveness of vision-based tactile sensors for hardness classification.**

I. INTRODUCTION

Robots are increasingly deployed across various domains, from manufacturing to healthcare, where they interact with objects and rely on sensory feedback to guide their actions. A key challenge in robotics is accurately perceiving object properties. This work focuses on one fundamental property sensed through touch: hardness. Specifically, we explore active sampling strategies for efficient hardness classification using a Vision-Based Tactile Sensor (VBTS). VBTSs, such as GelSight Mini [1] and FingerVision [2], offer a high-resolution, cost-effective alternative to traditional tactile sensors while benefiting from advancements in camera technology and computer vision.

In this work, we investigate information-theoretic active sampling strategies for efficient hardness classification. The robot is tasked with classifying a "test object" into one of the reference classes based on its hardness, using as few touches of a VBTS as possible (see Fig.1). The robot has no prior knowledge of the objects and can collect multiple samples by touching both the test and reference objects. In prior work, Boehm et al. [3] explored active sampling strategies for texture recognition using a VBTS. While Boehm et al. framed texture recognition as an image classification problem, hardness classification with a VBTS involves processing a sequence of frames [4], [5], making it a more complex task.

We provide new empirical evidence on the effectiveness of model-uncertainty-based active sampling strategies for hardness classification with a VBTS. We first compare three probabilistic classifier architectures on two datasets: our dataset containing videos collected by a robot pressing a GelSight Mini sensor on five silicone objects of different hardness levels (see Fig. 1) and the dataset created by Yuan et al. [6] containing videos obtained by human testers pressing an older version of the GelSight sensor on different silicone objects. Then, we implement and evaluate the active sampling strategies with the most promising model for each dataset. We also conduct a human-participant study to assess the performance of our method vis-à-vis human performance on the hardness classification task.

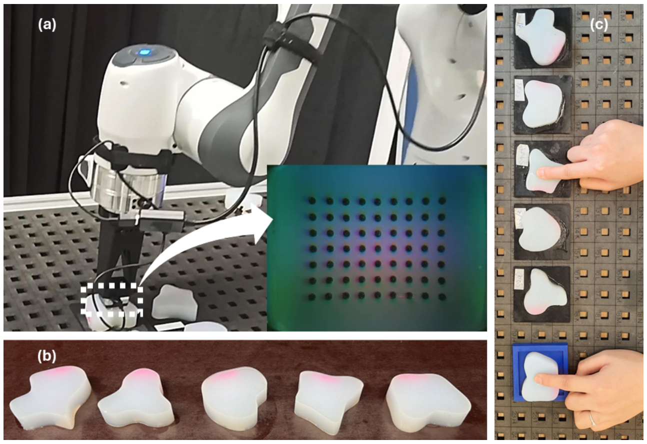

Fig. 1: The hardness classification task requires the agent to classify the test object into one of the reference classes based on hardness level. (a) The robot uses a GelSight Mini sensor mounted on its end-effector to explore the hardness of objects. The image captured by the sensor is shown in the inset. (b) Our dataset consists of GelSight Mini videos collected from five silicone objects of different shapes in increasing hardness from left to right. (c) In our human participant study, participants explored the test object (blue plate) and the reference objects (black plates) with their index fingers to compare hardness.

II. METHOD

We seek to investigate sample-efficient hardness classification using vision-based tactile sensors. The agent first trains the classifier model on a small number of initial samples collected from the reference objects (5 per object in our experiments). Following this initial training, the agent decides on the next reference object to collect more samples based on the model's uncertainty. The agent queries the model

[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, TU 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.

repeatedly with the same test samples and estimates the model's uncertainty from the distribution of the predictions generated due to the dropout layers. Similar to Boehm et al. [3], we compare two uncertainty metrics—*entropy* and *variance*—obtained from the classifier's predictions on the samples collected from the test object. To manage the size of training data, we introduce reservoir sampling to the active sampling strategies considered in [3]. Initially, the model is trained on a training reservoir and subsequently retrained in execution time with selected reference samples added to the reservoir to adapt to new data.

III. EVALUATION

We seek to evaluate the effectiveness of active sampling strategies for the hardness classification task and also compare their performance with human performance. To do so, we first evaluate the three classifier models on our dataset and a previously published dataset. Then, we test the active sampling strategies with the best model for each dataset. Finally, we conduct a human-participant study to estimate the hardness classification accuracy of humans on the objects in our dataset.

A. Datasets

We tested our hardness classification method on two datasets: a dataset that we created containing five hardness classes, where we sample 5 frames for each video, and a dataset created by Yuan et al. [6] containing nine hardness classes, where we sample 3 frames per video.

B. Classifier Architectures

We implement three probabilistic classifiers for the task of hardness classification from VBTS videos. The first model, Optical Flow Features and Neural Network Classifier (OFNN), classifies video data using Lukas-Kanade motion features [7] and a 2-layer neural network with dropout. The second model, DINOv2 Features and Neural Network Classifier (DINOv2NN), uses pre-trained DINOv2 [8] to extract visual features from video frames, which are classified by a 2-layer neural network with dropout. The third model, Convolutional Long-Short Term Memory Network (ConvLSTM) [9], combines CNN for spatial feature extraction and LSTM for temporal learning.

C. Hyperparameters

We set the initial and per-iteration sample count to five. Our dataset was harder to classify than the Yuan dataset, so we used 100 initial training epochs and 50 epochs per iteration for our dataset and 50 and 25 epochs, respectively, for the Yuan dataset. Due to GPU constraints, we limited training samples per iteration to 80 for OFNN and DINOv2NN, and 40 for ConvLSTM, which required more memory.

D. Model Baselines

We first compare the performance of the three classifier models—*OFNN*, *DINOv2NN*, and *ConvLSTM*—trained with the initial samples without any active sampling on both datasets, to decide which model is used on the respective dataset. Each model is trained on five samples per class and evaluated on five validation samples per class. We find that *ConvLSTM* outperforms the other two models on our dataset, achieving a validation accuracy of 42%, *DINOv2NN* follows close behind with an accuracy of 38%, and *OFNN* performs the worst with a validation accuracy of only 27%. On the Yuan dataset, *DINOv2NN* outperforms the other two models, achieving a validation accuracy of 59%, *ConvLSTM* is next with an accuracy of 46%, and *OFNN* performs the worst with a validation accuracy of only 24%.

E. Active Sampling Performance

We test both information-theoretic active sampling strategies (*entropy* and *variance*) and compare them to the *random* sampling baseline on our dataset and the Yuan dataset, with the best-performing classifier models, i.e., *ConvLSTM* for our dataset and *DINOv2NN* for the Yuan dataset. On both datasets *variance* strategy achieved the highest accuracy and the lowest MAE.

F. Human Study

We conducted a human study with ten participants for hardness classification with the objects in our dataset. Each trial included a practice round and five test rounds, where blindfolded participants used only their index fingers to identify one of five reference objects that matched the hardness of a test object. Each participant classified five different test objects with shuffled reference sequences. As shown in Table I, the Convolutional Long Short Term Memory (ConvLSTM) classifier outperforms the human participants, both with and without the active sampling strategies, on the same set of objects.

Humans	*No Resampling*	*Variance*	*Entropy*	*Random*
48.00%	57.20%	**88.78**%	85.79%	83.26%
±22.27%	±37.25%	±26.85%	±25.85%	±30.84%

TABLE I: Comparison of the hardness classification accuracies on our dataset. *Humans* denotes the accuracy of the human participants in our study (see Sec. III-F). *No Resampling* denotes the accuracy of the ConvLSTM classifier with five initial samples per class (see Sec. III-D). For the three active sampling strategies, the accuracies shown are after 5 sampling iterations.

IV. CONCLUSION

We investigated the performance of a model-uncertainty-based approach to active sampling for hardness classification with a VBTS. We evaluated different classifier architectures and active sampling strategies to find out the choices that maximize classification accuracy. Future works could focus on sample-efficient hardness classification of complex objects with multi-modal data such as proprioceptive, visual, and tactile data.

REFERENCES

[1] "GelSight Mini System - GelSight," April 2023, [Online; accessed 15. Sept. 2024]. [Online]. Available: https://www.gelsight.com/product/gelsight-mini-system

[2] A. Yamaguchi and C. G. Atkeson, "Recent progress in tactile sensing and sensors for robotic manipulation: can we turn tactile sensing into vision?" *Advanced Robotics*, vol. 33, no. 14, pp. 661–673, 2019.

[3] 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, pp. 15 099–15 105.

[4] S. Fang, T. Mi, Z. Zhou, C. Ye, C. Liu, H. Wu, Z. Yi, and X. Wu, "Tactcapsnet: Tactile capsule network for object hardness recognition," *IEEE International Conference on Real-Time Computing and Robotics*, pp. 1128–1133, July 2021.

[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*, vol. 20, pp. 2784–2794, 2023.

[6] 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*, July 2017, pp. 951–958.

[7] 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, 1981, pp. 674–679.

[8] M. Oquab, T. Darcet, T. Moutakanni, H. Vo, M. Szafraniec, V. Khalidov, P. Fernandez, D. Haziza, F. Massa, A. El-Nouby *et al.*, "DINOv2: Learning robust visual features without supervision," *arXiv preprint arXiv:2304.07193*, 2023.

[9] 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.