2024 IEEE International Conference on Robotics and Automation (ICRA 2024)
May 13-17, 2024. Yokohama, Japan

What Matters for Active Texture Recognition With Vision-Based Tactile Sensors

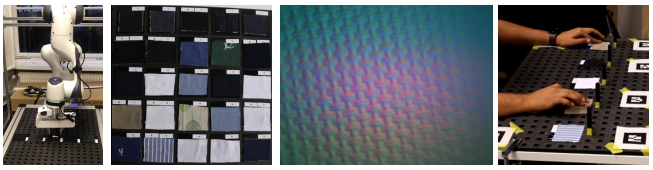
Alina Böhm[1], Tim Schneider[1], Boris Belousov[2], Alap Kshirsagar[1],
Lisa Lin[3], Katja Doerschner[3], Knut Drewing[3], Constantin A. Rothkopf[4,5], Jan Peters[1,2,4,5]

***Abstract*— This paper explores active sensing strategies that employ vision-based tactile sensors for robotic perception and classification of fabric textures. We formalize the active sampling problem in the context of tactile fabric recognition and provide an implementation of information-theoretic exploration strategies based on minimizing predictive entropy and variance of probabilistic models. Through ablation studies and human experiments, we investigate which components are crucial for quick and reliable texture recognition. Along with the active sampling strategies, we evaluate neural network architectures, representations of uncertainty, influence of data augmentation, and dataset variability. By evaluating our method on a previously published Active Clothing Perception Dataset and on a real robotic system, we establish that the choice of the active exploration strategy has only a minor influence on the recognition accuracy, whereas data augmentation and dropout rate play a significantly larger role. In a comparison study, while humans achieve 66.9% recognition accuracy, our best approach reaches 90.0% in under 5 touches, highlighting that vision-based tactile sensors are highly effective for fabric texture recognition.**

I. Introduction

Touch is a crucial sensing modality that helps humans perceive object properties and perform dexterous manipulation tasks. Without tactile feedback, even simple tasks such as lighting a match become harder to perform [1], [2]. Therefore, incorporating tactile sensing into robotics is an important step towards making robots more versatile and dexterous [3]. In this paper, we focus on the problem of tactile perception of object properties, and in particular, on the recognition of fabric textures. Various applications, such as laundry separation and fabric recycling, waste sorting, and material handling, can benefit from rapid texture classification, as discussed in [4].

Classification of fabrics has been tackled with different types of sensors using both supervised and active methods. *Vibration/force-based tactile sensors* of different types—such as the iCub sensors [5]–[7], BioTac sensors [6], [7], and custom-designed ones [8]–[10]—have been used for supervised texture classification, using spiking neural networks [6], modified RNNs [7], and k-NN classifiers [10]. All these methods rely on high-frequency temporal data, requiring RNNs or spatio-temporal subsampling to keep the input dimensionality low. In contrast, *vision-based tactile sensors* provide high-resolution data but at a lower rate, thereby requiring less history as input. Supervised classification of fabrics was successfully showcased using GelSight heightmap patterns [11] and more advanced spatio-temporal attention features [12]. Furthermore, *active sampling methods* have been developed for GelSight to 'actively' collect the data [4], in the sense of repeating touches until a 'good' tactile image is obtained, or for material roughness classification [13], where predictions on image patches were weighted by the output variance of a Bayesian CNN to improve the overall label prediction accuracy.



(a) (b) (c) (d)

Fig. 1: The texture recognition task requires identifying a given fabric among four comparison samples: (a) robot arm exploring sample fabrics; (b) dataset of 25 fabrics; (c) example tactile image; (d) human participant using index fingers to compare fabric samples.

In this paper, we are tackling the problem of *tactile active texture recognition* (see Fig. 1). With no pre-training, a robot is given a 'reference' texture and asked to identify it among four comparison textures using as few touches as possible. This problem setup models applications where the robot needs to quickly identify an object provided only a few touches. Unlike [4], we do not want to pre-train on a large dataset but rather quickly adapt on-the-fly, and we do not aim to 'classify' but only 'recognize' fabrics. In contrast to [13], we do not use uncertainty for label prediction but rather for action selection, choosing which fabric to touch next. This setup also allows us to compare robot vs. human tactile exploration strategies.

In the next sections, we formalize the tactile active texture recognition problem, present a general Bayesian decision-theoretic framework for action selection, describe our implementation which leverages probabilistic NNs for uncertainty quantification, and provide extensive empirical studies and analysis of different components of the algorithm, including the comparison to human exploration strategies and ablations on two datasets and experiments on a real robot.

II. Problem Setup and Task Formalization

We investigate sample-efficient texture recognition using vision-based tactile sensors such as GelSight Mini [14],

[1]Intelligent Autonomous Systems Lab, Department of Computer Science, TU Darmstadt, Germany, tim.schneider1@tu-darmstadt.de
[2]German Research Center for AI (DFKI) http://dfki.de/sairol
[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.

Digit [15], or FingerVision [16]. In this paper, we focus on the GelSight Mini sensor, as it provides high-resolution, high-quality images, independent of the external lighting. The sensor is held by a Franka Panda [17] robotic arm (see Fig. 1) and pressed against pieces of fabrics on plastic platforms at predefined locations with randomized amounts of pressure and rotation around the vertical axis to provide more variability in the data. The leftmost platform holds the *reference texture*, while the remaining four platforms hold randomly chosen *comparison textures*, one of which is equal to the reference.

The agent's goal is to identify the reference among the comparison textures using as few touches as possible. Crucially, the agent has no prior knowledge of the textures, and therefore has to learn about them within one *trial*, i.e., one fixed selection of five fabrics in a particular order. One trial consists of multiple touches and ends after a predefined number of touches in our robot experiments or once the participant has made a decision in our human study. The *action* of the agent is the high-level choice which platform to approach next (the low-level robot control is handled by a Cartesian position controller). We call each step of this action-observation loop a *round*, and we start counting rounds after each object has been touched once, i.e., if the process has terminated after one round, it means the agent has touched all four comparison fabrics and the reference fabric once and then did just one additional touch. Thus, one trial consist of several rounds (up to 20). We perform multiple trials with different textures, and multiple *runs* for each trial to reduce the statistical error.

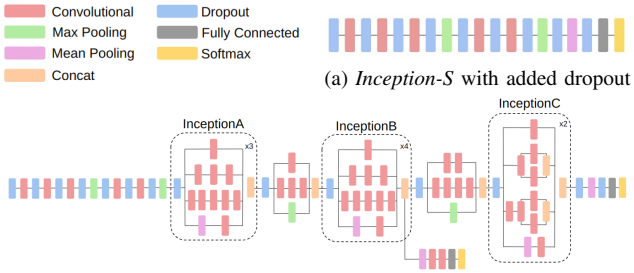
To provide textures for our experiments, we created a dataset of 25 denim and cotton fabrics, chosen to be particularly hard to distinguish by touch, as confirmed by our human study in Sec. IV-C. For each fabric, we collected 200 *samples* with randomly perturbed positions and rotations around the vertical axis. A sample of this dataset can be seen in Fig. 1c. Our complete dataset is available online.[1]

III. Tactile Active Texture Recognition Method

Consider one round of the agent's decision making. Having touched each of the five platforms one or more times, the agent needs to make a decision which platform to touch next. The Bayesian approach to this problem is to build a probabilistic model and to choose the action that provides the most information to support the final decision (i.e., the decision which fabric is identical to the reference) [18]. To implement this approach, we specify the model, describe how it is updated using the new data, and define the *acquisition function*, i.e., the action selection strategy.

A. Probabilistic Model Specification

As the output of the GelSight Mini sensor is a 320×240 RGB image, one either needs to manually extract features or employ a CNN. In [11], heightmap patterns were used, but with the advent of deep learning, automated feature extraction is prevalent. Therefore, in this paper, we employ a CNN with dropout layers to implement a probabilistic classifier [19]. Dropout has been shown to provide a viable approach for uncertainty quantification with neural networks [20]. Our experiments with an ensemble of CNNs have shown similar performance to dropout, albeit at a higher computational cost.



(a) *Inception-S* with added dropout

(b) Inception-v3 with added dropout

Fig. 2: The considered architectures of the probabilistic classifier: Inception-v3 and small Inception-v3 (*Inception-S*) with dropout.

We consider three CNN variants: i) Inception-v3 [21] pre-trained on the ImageNet [22] (*Inception-PT*); ii) randomly initialized Inception-v3 (*Inception-RI*); iii) small unpretrained version of Inception-v3 (*Inception-S*), which drops all the layers after the first *InceptionA* block and before the last *InceptionC* block (see Fig. 2). Considering these network variants allows us to evaluate the effects of pretraining and the network depth. We furthermore add dropout layers and evaluate different dropout rates in our ablation studies.

B. Model Update

Once a new tactile image is obtained, the model needs to be updated to incorporate the new evidence. As is common in deep learning, we employ *data augmentation* [23], by generating 10 randomly rotated versions of the same tactile image. Using all the samples collected during the current trial, we retrain the probabilistic NN classifier: the samples of the comparison textures serve as inputs and the respective platform positions serve as labels. The output of the classifier is a probability distribution $p_\theta(i|o)$ over the platform labels $i \in \{1, 2, 3, 4\}$, given an image o and the model parameters θ.

Hence, the model learns to map texture samples to platform labels. When queried with the reference texture (unseen during training), the model outputs a 'probability distribution' over the labels. To obtain a more robust estimate, we apply the model to 10 randomly rotated copies of the reference image and average the probabilities

$$i^* = \arg\max_i \frac{1}{n_{\text{ref}}} \sum_{k=1}^{n_{\text{ref}}} p_\theta(i|o_k^{\text{ref}})$$

where n_{ref} is the number of samples of the reference texture and o_k^{ref} is the k-th reference sample. This response is correct if the platform with the same fabric as the reference is predicted.

C. Active Sample Selection Strategy

The decision which platform to explore next is made based on the model uncertainty. As described in Sec. III-A, we add dropout layers to Inception-v3 (see Fig. 2) to model the

[1]Our Tactile Active Recognition of Textures (TART) Dataset can be downloaded at `https://drive.google.com/drive/folders/1S_2PLKV-Ap2tifV1gvMfCYjoaNyQaF9z?usp=sharing`

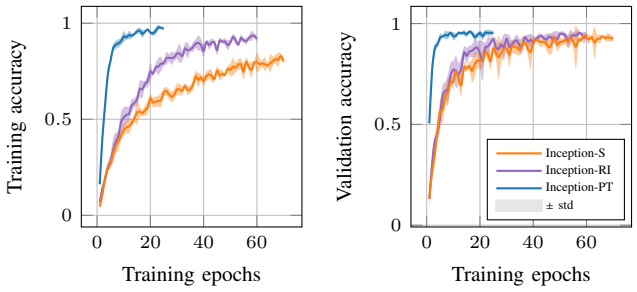


Fig. 3: Training and validation accuracy of the Inception-v3 models (see Sec. III-A) on the non-interactive 25-fabric classification task, averaged over five runs. Each model is trained until the validation accuracy converges. The final validation accuracies are 95.2% for *Inception-PT*, 94.2% for *Inception-RI*, and 92.6% for *Inception-S*.

epistemic uncertainty [20]. By querying the dropout network with the same input multiple times, we obtain different output samples and can gauge the uncertainty by their distribution.

We compare four sample selection strategies: *Random*, *Variance*, *Entropy*, and *You Only Touch Once (YOTO)*.
i) *Random* strategy is a naive non-active baseline that selects the next texture to touch according to a uniform distribution.
ii) *Variance* strategy selects the platform for which the variance of the class probability predictions is the highest

$$i_{\text{next}} = \arg\max_i \frac{1}{n_{\text{ref}}} \sum_{k=1}^{n_{\text{ref}}} \underset{m\sim p(m)}{\text{Var}} [p_\theta(i|o_k^{\text{ref}}, m)]$$

where n_{ref} is the number of collected samples of the reference object and $p(m)$ is the distribution of the dropout masks.
iii) *Entropy* strategy selects the platform that contributes the most to the class distribution entropy for the reference object

$$i_{\text{next}} = \arg\max_i \frac{1}{n_{\text{ref}}} \sum_{k=1}^{n_{\text{ref}}} \underset{m\sim p(m)}{\mathbb{E}} [-p_i^k \ln p_i^k]$$

where n_{ref} and $p(m)$ defined as before and $p_i^k \coloneqq p_\theta(i|o_k^{\text{ref}}, m)$.
iv) *You Only Touch Once (YOTO)* is a trivial baseline that makes a decision immediately after the initial five touches, i.e., each object touched once. This baseline provides a reference to quantify the 'value' of the actively gathered data.

IV. Experimental Results

Our experiments aim at identifying what components of the algorithmic architecture matter for active texture recognition with vision-based tactile sensors. For that, we first compare the three probabilistic classifier architectures introduced in Sec. III-A in a classical, non-interactive supervised learning setting on all 25 fabrics in our dataset. Second, we evaluate the active sample selection strategies from Sec. III-C. Third, we present a human study in which we investigate human exploration strategies on the same task in order to identify potential improvements to the robotic policies. Fourth, we provide ablation studies of the effects of hyperparameters on a bigger *Active Clothing Perception Dataset* [4].

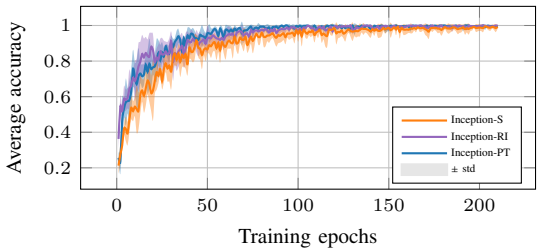


Fig. 4: Comparing the performance of the Inception-v3 models on the active texture recognition task. Notably, the small Inception network *Inception-S* performs as well as the larger *Inception-PT*.

A. Supervised Texture Classification

In this experiment, the three models introduced in Sec. III-A are trained to distinguish all 25 fabrics of our dataset. Since we are not interested in uncertainty quantification in this experiment, we set the dropout probability to 0%. For this experiment, we use 10 images per class for training and 20 images for validation. We see considerable differences within the performance of the three models. Fig. 3 shows that pre-training helps to learn to recognize the textures quickly and that *Inception-S* needs more training time to reach the same performance as the other two models. At the same time, the performance on the validation data is close to *Inception-RI*, indicating that *Inception-S* still generalizes well.

B. Active Texture Recognition

The objective of this experiment is to compare the three network architectures (Sec. III-A) and the four active sampling strategies (Sec. III-C). Each model is trained for 210 epochs, 10 epochs after creating a baseline and then 10 more epochs after resampling in each of the 20 rounds. For all three models, we collect the results of five subsets of fabrics using the four different strategies and average the performance of each model. In Fig. 4, it can be seen that the models perform similarly, unlike in the non-active experiment in Sec. IV-A. Especially the large advantage of *Inception-PT* seen in Sec. IV-A cannot be observed in this experiment. We believe the reason why pre-training is not advantageous in this case is the retrospective addition of dropout layers, which the model was not trained for. Considering that *Inception-S* can solve the task on par with the other models while being substantially less computationally expensive, we use this model for our further comparisons.

In Fig. 5, the influence of the different strategies on the performance of *Inception-S* is shown. When we run the experiment for 20 rounds, sampling offers an advantage, as *YOTO* has the lowest performance on the training data. On average, the model performs best using *Variance*, closely followed by *Entropy* and *Random*. However, the active sampling strategies generally have a lower predictive uncertainty than *YOTO*. While *YOTO* performs quite well on the training data after 20 rounds, its accuracy in predicting the label of the reference object of each trial is only 80% on average.

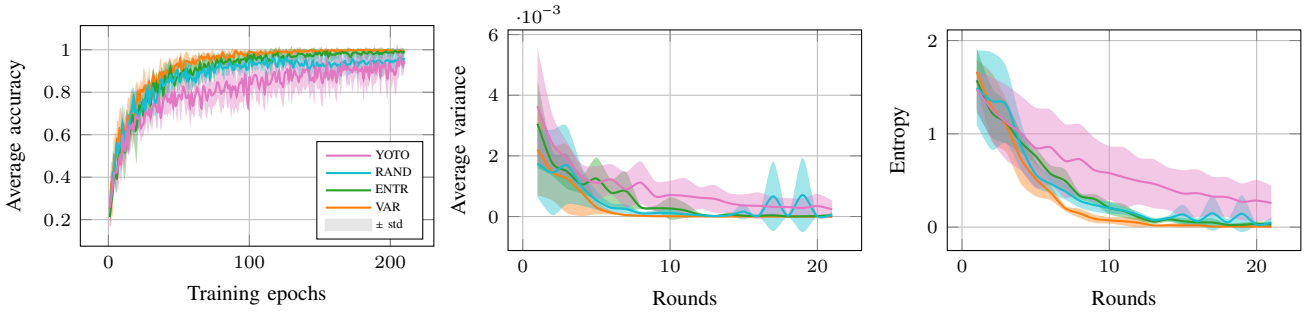


Fig. 5: Comparison of the exploration strategies on the *tactile active texture recognition* task. Average prediction accuracy, average variance, and entropy of the predictions are shown. *Inception-S* is used in all experiments. The *Variance* strategy achieves the highest accuracy, closely followed by *Entropy* and *Random*. Interestingly, the *Variance* strategy leads to a faster entropy decrease even than *Entropy* (rightmost plot).

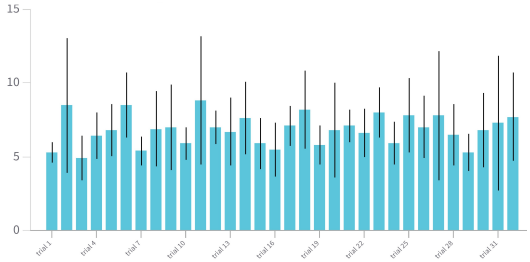


Fig. 6: The average number of touches made by the human participants in each trial. Error bars denote the standard error. Humans do 5–9 touches (i.e., 1–5 rounds) in each trial before giving the final response. Some harder trials with similar objects show high variance.

Humans	*Variance*	*Entropy*	*Random*	*YOTO*
66.88% ±16.93%	**90.00%** ±15.24%	88.13% ±14.24%	89.38% ±14.35%	80.63% ±22.42%

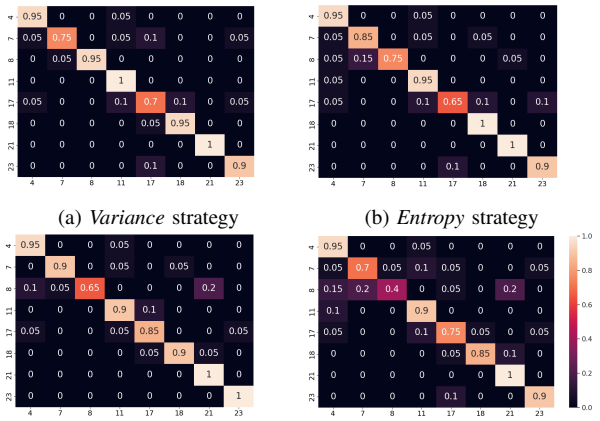
TABLE I: Comparison of the final accuracies achieved by the different exploration strategies. *Humans* denotes the average human performance. Notably, all robotic strategies are superior to humans, showing that the vision-based tactile sensor alone provides an advantage over the human touch in this task. The type of the exploration strategy, however, seems to play a minor role, since *Variance*, *Entropy*, and *Random* all achieve about 90% accuracy.

C. Human Study

In order to find out how well the proposed tactile active recognition method performs compared to humans, we carry out an experiment with ten human participants. Ideally, we would like to see whether human exploration strategies can be characterized using information-theoretic metrics, and whether insights from humans can be transferred to the robots.

During a trial, the participants are blindfolded and can only use the tips of either their index or middle fingers to explore the fabrics (see Fig. 1d). All participants have passed the two-point discrimination test with a point distance of 2mm, confirming that their tactile perception capabilities are not impaired [24]. They indicate their response to the experimenter by resting their finger on the chosen fabric and by verbally confirming their selection. No feedback on the participant's performance is provided during the experiment.

We select 8 out of 25 fabrics which were mostly confused by the neural network and we use each fabric as the reference object four times, with the corresponding comparison object



(a) *Variance* strategy (b) *Entropy* strategy

(c) *Random* strategy (d) *YOTO* strategy

Fig. 7: Confusion matrices for the 8 fabrics included in the human study. Some fabrics are consistently misclassified by all strategies, e.g., 17 and 7, while others are always classified correctly, e.g., 21.

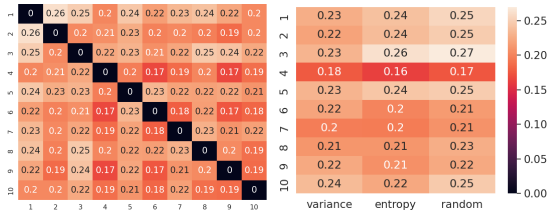
placed in each of the four possible locations once, resulting in 32 trials in total. To analyze the number and time of revisits per object, we record videos of each participant's hand movements. Each time a participant switches between two objects is counted as a new revisit. The data of this experiment and the code for analysis are available online.[2]

Figure 6 shows the average number of revisits before giving a response in each trial, ranging from five to nine revisits needed per trial. In Table I, we compare the prediction accuracy of the human participants and the robot. To ensure fairness, in each trial, we only allow the robot to use the same number of touches that humans used (Fig. 6). The low prediction accuracy of 66.88% achieved by the humans indicates that the task is quite non-trivial, and it provides a reference point for the 90% accuracy achieved using the vision-based tactile sensor. In the next section, we take a closer look at the exploration strategies employed by the humans in comparison to the information-theoretic strategies.

D. Behavior Comparison Between Participants and the Robot

To gain insight into the difficulty of differentiating the fabrics, we plot the confusion matrices in Fig. 7. Some fabrics

[2]Our study data `https://drive.google.com/drive/folders/1S_2PLKV-Ap2tifV1gvMfCYjoaNyQaF9z?usp=sharing`



(a) inter-participant (b) participant-robot

Fig. 8: Comparing the exploration strategies among participants and against the information-theoretic strategies. The numbers indicate the Jensen-Shannon divergence between the distributions of time spent over objects, averaged over trials. The inter-subject variability is comparable to the subject-robot variability, therefore no uniform judgement about what strategy all humans use can be made. Instead, each participant seems to follow a personal exploration strategy.

have close to 1 recognition accuracy, whereas others are misclassified more often. Notably, the confusion matrix of *YOTO* has the lowest values on the diagonal, in accordance with the results in Table I. For humans, an average confusion matrix is not very informative since no single fabric was inherently harder to recognize for all participants, i.e., the inter-participant variance was relatively high.

To compare the exploration strategies, we formalize the problem by normalizing the time spent by the human participants on each object per trial, to get a distribution of relative times per fabric. This gives us a distribution of time spent over objects, and subsequently we can compute a distance between these distributions to judge how close they are. We employ the symmetric Jensen-Shannon divergence. Thus, we can compare both human and robotic strategies to each other at least in this restricted sense. *YOTO* is excluded from this comparison as it performs no exploration.

The Jensen-Shannon distance takes values in the range $[0, 1]$, with lower values indicating greater similarity between strategies. The resulting distances of comparing the robotic strategies to each other are 0.14 between *Variance* and *Entropy*, 0.12 between *Entropy* and *Random*, and 0.16 between *Variance* and *Random*. Thus, according to the Jensen-Shannon divergence, the two uncertainty-based strategies *Variance* and *Entropy* are not the most similar, and *Entropy* produces exploratory behavior that is more similar to *Random* than to *Variance* on average. While the distances between the robot strategies are in the range 0.12–0.16, the inter-participant (Fig. 8a) and the participant-robot (Fig. 8b) distances are around 0.2 and higher.

If we average the distances of each participant, we get a mean Jensen-Shannon distance of 0.219 for the variance strategy, 0.218 for the entropy strategy, and 0.231 for the random sampling strategy, meaning that the uncertainty-based strategies are on average slightly more similar to human exploration under the Jensen-Shannon distance. On the other hand, we again observe a high variance between trials. There are some trials where the same two participants follow a very similar exploration strategy and others where they choose different approaches with a higher Jensen-Shannon distance.

Finally, we observe a similar pattern when it comes to the *Variance* strategy of the robot and the human participants. In 63.75% of the trials, the participants touched that object most often which they predicted to be the reference object. With *Variance*, we get 56.25%, and with *Entropy*, only 32.5% of the trials where the predicted object is touched most often.

	Variance	*Entropy*	*Random*	*YOTO*
DA, DR=0.25	68.13% ±46.8%	78.44% ±40.81%	81.87% ±37.02%	46.88% ±50.7%
No DA	50% ±50.8%	59.22% ±48.33%	62.66% ±47.38%	49.22% ±50.19%
DR=0.5	55.94% ±50.15%	80.94% ±37.01%	79.06% ±39.54%	49.06% ±50.12%
DR=0.15	70.94% ±45.25%	80.31% ±39.55%	79.69% ±36.23%	47.19% ±50.43%
DR=0.05	68.75% ±47.09%	84.38% ±36.89%	86.56% ±33.47%	50% ±49.25%

TABLE II: Ablations on the influence of Data Augmentation (DA) and Dropout Rate (DR) on the final prediction accuracy of different exploration strategies, evaluated on the *Active Clothing Perception Dataset* [4]. Adding data augmentation boosts the performance by 20% for all strategies except *YOTO*. Decreasing the dropout rate generally improves the results, but to a lesser degree.

E. Ablation Study on the Active Clothing Perception Dataset

In the ablation study, we investigate the role of other hyperparameters and design choices on the performance of the considered tactile active texture recognition algorithm. To make sure that the results are not specific to our dataset, we perform the ablations on the images from the *Active Clothing Perception Dataset* [4]. To make the setup comparable to our experiments, we randomly select 8 of the fabrics and create 32 trials, with 4 fabrics each.

Table II and Fig. 9 show the results of the ablation studies on the Dropout Rate (DR) and Data Augmentation (DA). In general, the dataset [4] contains more variable data compared to ours, because the data was collected by autonomously grasping real clothes at wrinkle locations. Furthermore, a prior version of the GelSight sensor was used, that exhibits a higher variance in the light distribution upon contact. For these reasons, the variance in the performance of our method is higher on this dataset (Table II) compared to ours (Table I).

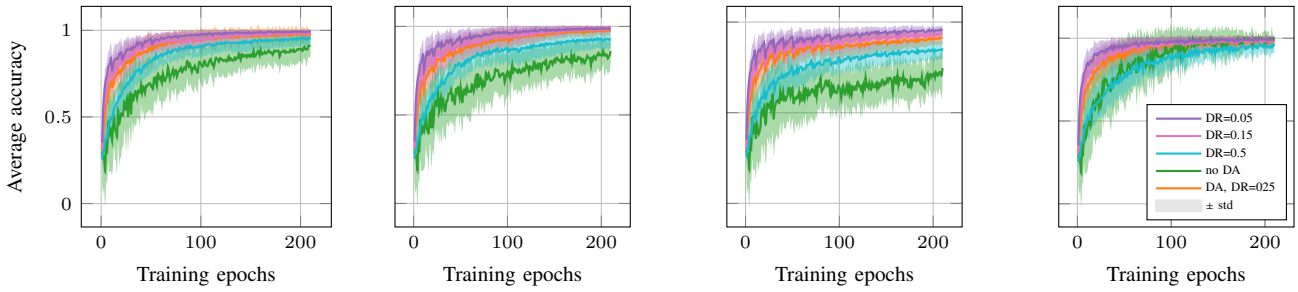
In the first two rows in Table II, we compare the final average test accuracies with and without DA, at a fixed DR = 0.25. For all methods except *YOTO*, data augmentation adds about 20% to the accuracy. This confirms our observation that data augmentation is essential for the good performance of the CNN-based model.

In the bottom three rows in Table II, we compare different dropout rates. The general trend is that the smaller DR results in a higher average accuracy, though the improvement is minor in the range DR ≤ 0.25. Therefore, in our main experiments, we used 0.25 as it performed sufficiently well in all tests.

In Fig. 9 we observe the same trends during training as in Table II. Namely, removing data augmentation leads to a significant drop in performance, and decreasing the dropout rate provides an improvement to the accuracy of all strategies.

V. Discussion & Conclusion

We have investigated the performance of a Bayesian approach to active sampling for fabric texture recognition



(a) Variance (b) Entropy (c) Random (d) YOTO

Fig. 9: Ablations on the influence of Data Augmentation (DA) and Dropout Rate (DR) on the prediction accuracy during training for different exploration strategies, evaluated on the *Active Clothing Perception Dataset* [4]. Adding data augmentation significantly improves the performance for all strategies except *YOTO*. Decreasing the dropout rate generally improves the results, but to a lesser degree.

with vision-based tactile sensors using variance and entropy criteria. We performed ablation studies on different model architectures and hyperparameters, and we identified which choices have the largest impact on the recognition accuracy.

First, we found that an ImageNet-pretrained Inception-v3 network allows for a significantly faster training of a 25-class classifier on our dataset of denim and cotton fabrics (see Fig. 3), achieving 95.2% accuracy after 20 epochs of training. However, on our main 4-class recognition task, where the network needs to adapt very quickly with only a handful of training samples, we found that even a much smaller model *Inception-S* performs similarly, while being more computationally efficient (see Fig. 4). Therefore, we conclude that big pretrained networks are not necessary for few-shot recognition tasks with vision-based tactile sensors.

Second, we investigated the importance of the exploration strategy on the recognition accuracy, and we did not find a significant difference between the strategies that sample the objects with highest predictive variance or entropy contribution. Furthermore, even a random sampling strategy has shown similar performance, which suggests that the texture recognition task is relatively straightforward for the vision-based tactile sensors such as GelSight Mini, despite the fact that humans only achieved 66.88% average accuracy on this task. These results are in agreement with the work on active classification of material roughness [13], where the algorithm using vision-based tactile sensors was shown to achieve significantly higher accuracy than human participants.

Third, we performed a human study with the goal of quantifying the human performance on the texture recognition tasks, and we performed analysis of exploratory behaviors that humans employ in order to compare them to the behavior of our exploration strategies. Apart from confirming that the task is quite hard for humans, we found out that there is a significant variability among the participants with regards to the exploration strategy, as evidenced by our analysis in Fig. 8. Moreover, the inter-participant variability was found to be similar to the participant-robot variability, meaning that there is no universal exploration strategy that all participants have followed. Nevertheless, on average, human exploration behavior was closer to the information-theoretic strategies, *Variance* and *Entropy*, than to random exploration.

Fourth, we reported the results of ablation studies on the effect of data augmentation and dropout rate on the model performance. Most importantly, we found data augmentation to significantly improve the performance of all exploration strategies (see Table II) by almost 20% on average. The dropout rate, on the other hand, had a relatively smaller influence, well within the standard error range. Combined with our observations about the influence of exploration strategies, this result allows us to conclude that the quality of the data and data augmentation, together with the network architecture, play a more significant role in improving the performance compared to the choice of the exploration strategy.

Limitations: Comparing human and robotic tactile perception in our study is limited due to the different nature of the sensors. The vision-based tactile sensor achieving a higher performance on the texture recognition task could in principle be attributed to using a different sensing modality rather than to a better representation or sample selection strategy. However, this concern was partially addressed in [13], where human performance using touch was compared to using GelSight images for material roughness classification. Interestingly, they found that humans are much better at classification using their sense of touch rather than vision.

Our results furthre suggest that the choice of which object to touch next in the fabric recognition task may depend on the nature of the internal representation, and not on the sampling strategy. It is, however, not straightforward to compare internal representations of humans and robots, especially given that there seems to be no 'average' human representation, i.e., humans differ substantially in terms of which fabrics they confuse and which exploration strategy they follow.

Outlook: While we evaluated ImageNet pre-trained networks on texture recognition, it would be of interest to develop a "tactile ImageNet" dataset and a network pre-trained only on textures. Such a network would potentially further improve the active texture recognition performance. Tackling more challenging tasks, such as contour/shape exploration and object pose estimation would provide further insights into active tactile sensing, as well as integrating multiple sensing modalities, such as touch, vision, and proprioception. On the human side, a study on the representations of object properties that humans utilize would be especially relevant.

REFERENCES

[1] R. S. Johansson and J. R. Flanagan, "Coding and use of tactile signals from the fingertips in object manipulation tasks," *Nature Reviews Neuroscience*, vol. 10, no. 5, pp. 345–359, 2009.

[2] H. Dang and P. K. Allen, "Stable grasping under pose uncertainty using tactile feedback," *Autonomous Robots*, vol. 36, pp. 309–330, 2014.

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

[4] W. Yuan, Y. Mo, S. Wang, and E. H. Adelson, "Active clothing material perception using tactile sensing and deep learning," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 4842–4849.

[5] T. Taunyazov, H. F. Koh, Y. Wu, C. Cai, and H. Soh, "Towards effective tactile identification of textures using a hybrid touch approach," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 4269–4275.

[6] T. Taunyazov, Y. Chua, R. Gao, H. Soh, and Y. Wu, "Fast texture classification using tactile neural coding and spiking neural network," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 9890–9895.

[7] R. Gao, T. Tian, Z. Lin, and Y. Wu, "On explainability and sensor-adaptability of a robot tactile texture representation using a two-stage recurrent networks," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 1296–1303.

[8] B. M. R. Lima, V. P. da Fonseca, T. E. A. de Oliveira, Q. Zhu, and E. M. Petriu, "Dynamic tactile exploration for texture classification using a miniaturized multi-modal tactile sensor and machine learning," in *2020 IEEE International Systems Conference (SysCon)*. IEEE, 2020, pp. 1–7.

[9] S. Huang and H. Wu, "Texture recognition based on perception data from a bionic tactile sensor," *Sensors*, vol. 21, no. 15, p. 5224, 2021.

[10] S.-a. Wang, A. Albini, P. Maiolino, F. Mastrogiovanni, and G. Cannata, "Fabric classification using a finger-shaped tactile sensor via robotic sliding," *Frontiers in Neurorobotics*, vol. 16, p. 10, 2022.

[11] R. Li and E. H. Adelson, "Sensing and recognizing surface textures using a gelsight sensor," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 1241–1247.

[12] G. Cao, Y. Zhou, D. Bollegala, and S. Luo, "Spatio-temporal attention model for tactile texture recognition," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 9896–9902.

[13] A. Amini, J. I. Lipton, and D. Rus, "Uncertainty aware texture classification and mapping using soft tactile sensors," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 4249–4256.

[14] "GelSight Mini System - GelSight," Apr. 2023, [Online; accessed 1. Jun. 2023]. [Online]. Available: https://www.gelsight.com/product/gelsight-mini-system

[15] M. Lambeta, P.-W. Chou, S. Tian, B. Yang, B. Maloon, V. R. Most, D. Stroud, R. Santos, A. Byagowi, G. Kammerer *et al.*, "Digit: A novel design for a low-cost compact high-resolution tactile sensor with application to in-hand manipulation," *IEEE Robotics and Automation Letters*, vol. 5, no. 3, pp. 3838–3845, 2020.

[16] 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.

[17] "Franka Emika - Next Generation Robotics." May 2023, [Online; accessed 1. Jun. 2023]. [Online]. Available: https://www.franka.de

[18] B. Settles, "Active learning literature survey," *Machine Learning*, vol. 15, no. 2, pp. 201–221, 1994.

[19] M. Abdar, F. Pourpanah, S. Hussain, D. Rezazadegan, L. Liu, M. Ghavamzadeh, P. Fieguth, X. Cao, A. Khosravi *et al.*, "A review of uncertainty quantification in deep learning: Techniques, applications and challenges," *Information Fusion*, vol. 76, pp. 243–297, 2021.

[20] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," in *International Conference on Machine Learning (ICML)*. PMLR, 2016, pp. 1050–1059.

[21] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2818–2826.

[22] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2009, pp. 248–255.

[23] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, pp. 1–48, 2019.

[24] D. Shooter, "Use of two-point discrimination as a nerve repair assessment tool: preliminary report," *ANZ Journal of Surgery*, vol. 75, no. 10, pp. 866–868, 2005.