Preprint. Cite as:
Kerzel M., Strahl E., Gaede C., Gasanov E. & Wermter S. (2019). Neuro-robotic haptic object classification by active exploration on a novel dataset, Proceedings of the International Joint Conference on Neural Networks (IJCNN), IN PRINT.

# Neuro-Robotic Haptic Object Classification by Active Exploration on a Novel Dataset

Matthias Kerzel, Erik Strahl, Connor Gaede, Emil Gasanov, Stefan Wermter
*Knowledge Technology, Department of Informatics,*
*University of Hamburg, Germany*
Hamburg, Germany
kerzel/strahl/4gaede/1gasanov/wermter@informatik.uni-hamburg.de

***Abstract*—We present an embodied neural model for haptic object classification by active haptic exploration with the humanoid robot NICO. When NICO's newly developed robotic hand closes around an object, multiple sensory readings from a tactile fingertip sensor, motor positions, and motor currents are recorded. We created a haptic dataset with 83200 haptic measurements, based on 100 samples of each of 16 different objects, every sample containing 52 measurements. First, we provide an analysis of neural classification models with regard to isolated haptic sensory channels for object classification. Based on this, we develop a series of neural models (MLP, CNN, LSTM) that integrate the haptic sensory channels to classify explored objects. As an initial baseline, our best model achieves a 66.6% classification accuracy over 16 objects. We show that this result is due to the ability of the network to integrate the haptic data both over time domain and over different haptic sensory channels. Furthermore, we make the dataset publically available to address the issue of sparse haptic datasets for machine learning research.**

***Index Terms*—haptic perception, embodied neuro-cognitive model, developmental robotics, haptic dataset**

## I. INTRODUCTION

Developmental robotics aims to develop intelligence in artificial agents through a rich interaction with complex, lifelike environments [6]. Often research focuses on the visual modality, but understanding non-visual concepts like *hard, soft, spongy or flexible*, is relevant for language acquisition [13] and development of motor abilities in the physical world. Recognition of such object properties is vital to avoid danger, e.g., a soft object needs to be grasped firmly to avoid slippage while fragile objects require a soft grasp. Haptic perception is also relevant when identifying objects in the absence of visual information, e.g., when a robot is picking up an occluded object. However, haptic perception is complex: in contrast to visual perception, haptic perception requires motor actions to actively explore an object. In humans, the resulting haptic perception is mediated by several different sensory subsystems ranging from different tactile sensing cells in the skin to proprioception of position and forces in joints, tendons, and muscles. Likewise, for robotic agents, signals from specialized types of tactile sensors, position and currents of motors need to be integrated to facilitate haptic perception. Artificial neural networks have shown great ability to integrate multimodal data [3], [10] and also process the sequential data from a haptic exploratory action. However, few established architectures exist which integrate multiple haptic sensory channels [19], which can in part be attributed to the lack of available haptic datasets.

Fig. 1. The humanoid robot NICO, Neuro Inspired COmpanion, in its experimental setup: NICO is squeezing 16 different objects to learn how to distinguish them by touch. Here, NICO is squeezing the object "red ball".

To address this issue, we developed a novel robotic setup for recording a large haptic dataset based on a single-point tactile sensor embedded in a robust three-fingered robotic hand. The hand is used to actively close around objects (squeezing) and record several data channels from the tactile sensors as well as motor positions and forces to perceive different haptic properties of the object like compliance and shape in one single active exploration. The dataset consists of 1600 samples of a set of 16 different objects that are explored in 100 repetitions with 52 measurements per sample, overall resulting in 83200 haptic measurements. It will be made publically available[1]. Based on this dataset, we developed and optimized a series of neural baseline approaches for haptic object classification

*The authors gratefully acknowledge partial support from the German Research Foundation DFG under project CML (TRR 169). The authors also thank Pedro Ramilo and the SeedRobotics team for the joint development of the tactile sensing robotic hand presented in this paper.

[1]The haptic dataset will be made available at https://www.inf.uni-hamburg.de/en/inst/ab/wtm/research/corpora.html.

that ranges from simple MLP architectures to convolutional and LSTM recurrent neural networks. Extending our previous work [15], we show that artificial neural networks can process and integrate raw sensory haptic data and do not require hand-crafted features. We reach a classification accuracy of over 66.6% on 16 different objects with the best architecture. In contrast to existing robotic haptic sensor setups, our design is robust, requires low-maintenance and relies on the neural integration of existing motor sensors with tactile information. Furthermore, we provide an analysis of different neural architectures with regard to involved sensory subsystems of the neuro-robotic platform. Our research contributes to the development of active robotic haptic perception with neural approaches.

## II. RELATED WORK

To develop embodied neuro-cognitive models based on active haptic exploration, understanding of both biological as well as technical sensory processes is necessary [13]. We will first report relevant findings of active human haptic sensing and robotic haptic perception before we discuss suitable neurocognitive models for processing this sensory information.

### A. Haptic Perception in Humans

Haptic perception is to a significant degree caused by active movements. Lederman and Klatzky [18] observed a set of haptic exploratory procedures that humans use to gain information about an object by touching or manipulating it in various ways: Humans press down an object to learn about its compliance; they slide a fingertip over the object to learn about its surface; objects are enclosed in a hand, or their contours are traced to learn about the object's shape. Objects are lifted to weigh them or objects are held in static grasp to gain information about their temperature. Research by Klatzky et al. [17] showed that sometimes these haptic exploratory procedures are used to compensate for the lack of visual information, e.g., when handling an object in a dark environment. At other times haptic exploratory procedures are employed to gather information about an object that is not visually available, e.g., finding out how soft or smooth an object is if this cannot be approximated from experience or context.

Haptic perception is mediated by a complex set of independent sensory channels [11]. Specialized nerve cells that are distributed in different densities in the epidermis are excited by pressure, skin deformation, different frequencies of vibration and temperature. This sensory subsystem is referred to as tactile sensing. It is complemented by proprioception: inside of the body, nerve cells register strain on and the state of muscles and tendons. Additionally, efferent signals from the motor cortex are used in processing haptic information. All of these signals converge in the somatosensory system [14], where they are integrated and processed. Though often thought of a singular sense, haptic perception can be seen as a complex multimodal perception process.

### B. Robotic Haptic Perception and Datasets

Different robotic sensor systems can be categorized according to the area and spatial resolution of the haptic sensor [19]. Furthermore, the sensors differ in their pressure sensitivity and sampling frequency. In the high-frequency range, these sensors pick up vibrations, which can, for instance, be caused by sliding the sensor over an object to differentiate materials.

One-Point tactile sensors measure the interaction forces between a single point of contact and an object. Jikvo Sinapov et al. [23] classify materials by scratching with an artificial fingernail. They use a k-nearest neighbors algorithm (k-NN) and a support vector machine (SVM) to classify the material based on magnitude and frequency components of the caused vibrations. In a similar approach, Romano and Kuchenbecker [21] move a pen-like recording tool over material samples using a PR2 robot. Materials are classified by Support Vector Machines.Kerzel et al. [15] use a one-point sensor for both sliding over a surface and for measuring compliance by pushing the sensor into the surface. They apply a convolutional neural architecture to classify materials based on the raw sensor data.

Several one-point sensors can be arranged into a dense array to form a fingertip-like sensor or a tactile sensitive surface.The SynTouch biomimetic tactile sensor (BioTac) [26] resembles an artificial fingertip with an array of 19 tactile sensors; additionally, it also registers vibration and temperature. Fishel and Loeb [9] use the BioTac sensor to slide it over different material samples based on Bayesian exploration. Gao et al. [10] use a robotic gripper with BioTac sensors to hold, squeeze and slide across different objects. Together with visual data, the resulting sensory information is used to train a deep neural network for haptic adjective classification. Tada et al. [24] developed a tactile-skin based on polyvinylidene fluoride (PVDF) films and strain gauges that covers the palm and the fingertips of a robotic hand. Takamuku et al. [25] use these sensors to squeeze and tap objects placed into the robotic hand. They employ a self-organized map (SOM) to cluster materials according to haptic properties like softness.

Additionally there are non-tactile sensors: The BioTac contains, for instance, a temperature sensor [26]. Finally, motor-system related sensors act like human proprioception by measuring position and forces in motors [12]. This functionality is regularly included in robotic motors for motion planning and prevention of motor overload. Similar to the presented dataset, the Open Access Haptic Database [5] focuses on multi-modal haptic datasets. However, non-hand-like end effectors are used for haptic exploration. In contrast, the dataset from Murali et al. [20] is more focussed on optimizing a grasping process. In summary, different approaches and datasets haptic sensing exist. However, the presented dataset explores everyday items that are characterized and distinguished by a set of haptic properties like softness and shape with a human haptic exploratory procedure. A combination of different sensors that cover the distinctive properties of objects will have a high utility for object classification.

### C. Neural models of haptic perception

In related work on robotic haptic perception neural (convolution, SOM) and non-neural approaches (SVM, k-NNm Bayesian classifier) are used. In contrast to non-neural approaches, neural networks can learn to extract meaningful features from raw sensor data [15] and can also be used to integrate different sensory channels [10], [15]. For some time recurrent models have been established as best practice when processing sequential data. Thus, such models should be applied to haptic sensory information collected during a temporally extended haptic exploration. However, recent advances question this status: Bai et al. [2] argue that convolutional architectures can outperform recurrent architectures on tasks like audio synthesis and machine translation. Likewise, Dauphin et al. [8] show for the domain of language modeling, that stacked convolutional architectures outperform recurrent architectures. In summary, there is not yet an established best practice for developing neural architectures for classification of haptic data. We will evaluate both convolutional and recurrent architectures and also include a simple MLP model that is based on a single measurement instead of a sequence of measurements.

## III. METHODOLOGY

We collect a dataset of 83200 haptic measurements, based on 100 samples of each of 16 different objects, every sample containing 52 measurements by applying a haptic exploratory procedure of squeezing an object in a robotic hand. We use the humanoid robot NICO and its three-fingered robotic hand with a tactile sensor for squeezing. During the closure of the hand, 52 measurements are taken, recording the following: the position of the fingers, current in the motors associated with the fingers, and tactile data from a sensor in the thumb. With the resulting data, a series of neural network classifiers are trained to evaluate the recorded dataset and to provide a baseline for further research.

### A. NICO Humanoid Robot Platform

We recorded the haptic dataset with NICO, the Neuro Inspired COmpanion, developed by the Knowledge Technology group in Hamburg[2] [16] for embodied multimodal neurocognitive models and social interaction. NICO is a child-sized humanoid with a height of one meter. Its size and design aim to create an approachable and non-threatening appearance while still being able to interact in a domestic environment. Distributed over its body, NICO features 30 degrees of freedom that enable anthropomorphic motions pattern. NICO is equipped with two cameras and stereo microphones. Now, with the installation of haptic sensors into its hands another modality for the perception of its environment is available.

The basis for the robot hand is the child-sized three-fingered seed robotics RH4D manipulator [3]. The robotic hand is underactuated and tendon driven: a single motor operates a tendon that coils all three segments of a finger. One motor operates both index fingers, and one motor operated the thumb. With only these two degrees of freedom, the hand can securely grasp objects of different sizes as the finger position adjusts to the shape of the objects. During the closing process, the position of the motors and their currents are registered. This hardware design realizes the human hand synergies [22], i.e., the fact that despite its complexity the human hand can be described by a few parameters because of correlated joint movements.

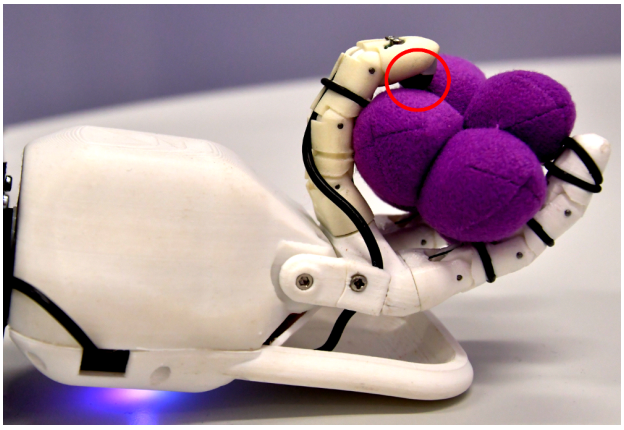

Fig. 2. Newly developed SeedRobotics hand with embedded OptoForce sensor squeezing the object “purple grapes”.

We embedded an OptoForce 3-Axis optical force sensor[4] into the tip of the thumb. Figure 2 shows the three-fingered hand with the embedded sensor. The OptoForce sensor consists of a small rubber dome (11mm diameter) that houses an infrared emitter, reflector, and sensor that registers slightest deformations of the dome to compute the applied force in three dimensions with an accuracy of 2.5 mN. The sensor is installed in a way that the dome acts as a fingertip for the robot hand during grasping. The sensor falls into the category of one-point interaction sensors. Like the hand, the sensor is robust and does not require maintenance; including pretrials, over 2000 grasps were executed without causing wear and tear to the hand or the embedded sensor. Kerzel et al. [15] report good results when using the OptoForce sensor in a robotic setup for material classification with a neural approach.

### B. Dataset Recording

For recording the dataset, NICO is seated at a table; its hand is resting on the table with the palm of the left hand facing upward. The sampling procedure begins with an object being placed into the empty hand. The hand closes in 52 steps, during each step the following data is stored: the position of index fingers, position of the thumb, current in finger motor, current in thumb motor and x-,y-, and z-forces in the tactile sensor embedded into the thumb. The currents of both

[2]Visit http://nico.knowledge-technology.info, for further information and video material.

[3]Visit http://www.seedrobotics.com/rh4d-manipulator.html for more information.

[4]Currently available from *OnRobot* as OMD-3 axis sensor (https://onrobot.com/products/omd-force-sensor/)

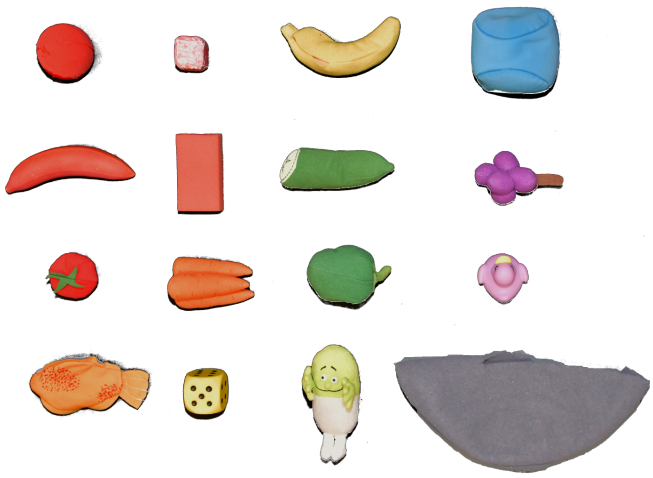

Fig. 3. Set of 16 objects used for recording the dataset. From left to right, top to bottom: red ball, red dice, yellow banana, blue ball, red banana, red sponge, green cucumber, purple grapes, red tomato, orange carrot, green pepper, purple duck, orange fish, yellow dice, green figure and black hat.

motors are monitored to avoid overload and prevent the hand from damaging itself during the closure. If one of the motors exceeds 100 mA, the closing procedure stops. However, later steps are still recorded to ensure equal sample length and to pick up sensory information after closing. This overload protection, together with the hand synergy enabling the design of the tendon operated hands allows the robot to securely grasp objects of unknown size and compliance without risk to the hardware.

The 16 different objects chosen for recording the dataset are depicted in Figure 3. All objects fit well into the child-sized robot hand. The objects are chosen to have a large variety concerning different haptic properties like symmetry (ball-shaped, cylindrical or complex), (minimum) diameter, overall compliance, and material. The properties are selected to be fitting to different sensory channels of our setup.

The collected dataset consists of 1600 samples that are evenly distributed over the 16 objects. Five experimenters were involved in creating the dataset. The procedure for recording was the following: The recording software selects a random object for which the desired number of samples was not yet collected. The object samples were evenly distributed over the experimenters. The experimenter placed the object into the robot hand and started the sampling procedure. The robot closes the hand till a force limit is reached, releases the object and the procedure repeats. All experimenters were instructed to *put the requested items into the robot hand in a way that it does not fall out during closure of the hand*. However, it was not specified how non-symmetrical objects are to be placed. During sampling, pictures were recorded from the robot's cameras; they show that experimenters placed the non-symmetrical objects very differently into the hand leading to a larger variance in measurements.

Figure 4 shows violin plots of the distribution of the seven

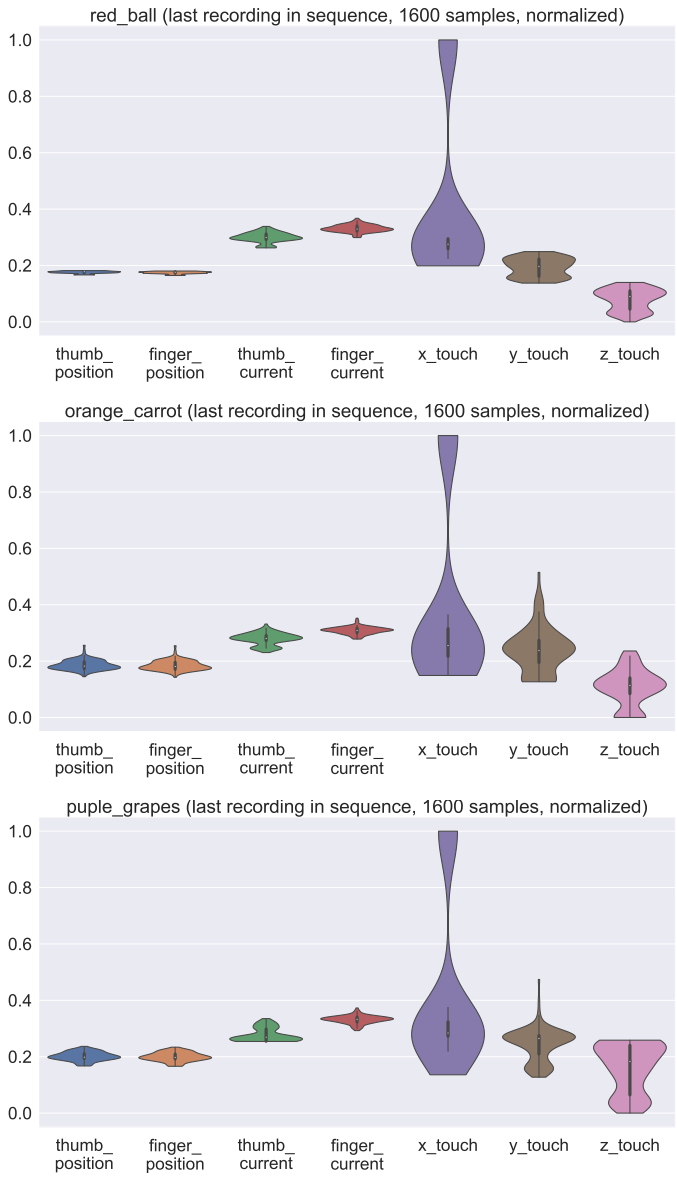


Fig. 4. Violin plots of the distribution of the seven different sensor values for the objects "red ball", "orange carrot", "purple grapes" (last timeframe of the sequence, all 1600 samples). The plots show distinctive distributions for all three objects. I.e., the finger positions show less variance for the spherical "red ball" in contrast to the non-symmetrical "orange carrot" or "purple grapes".

different sensor values for the objects "red ball", "orange carrot", "purple grapes" of the last sample in the sample series. As expected the position of the fingers is mainly influenced by the diameter of the objects, with the variance of the closing angle of the fingers determined by the symmetry. E.g., the finger position variance for the spherical "red ball" is small compared to the other two objects. Also, the forces detected at the fingertip, especially in the z-direction, is on average lower for the soft "red ball". For dynamic grasping, the force of closing and how it ramps up is determined by the objects overall compliance, e.g., when grasping a hard object a sudden, steep increase in force results. Grasping a soft object, in

contrast, leads to a softer increase in force.

### C. Neural architectures

The collected dataset consists of 1600 samples of a fixed length of fifty-two time steps. Each time step contains seven values: the position of the index fingers, the position of the thumb, the current in the index finger motor, the current in the thumb motors and the three x-,y- and z-forces at the point of contact of the tactile sensor. In addition to these values, the Euclidian vector norm of the three x-,y-, and z-forces is calculated and fed into the network as an eighth value. The Euclidean norm is used to compensate for the change of force direction during grasping [15], [23]. All channels of the sample are normalized to the interval $[-1, 1]$ but are not preprocessed in any other way.

Characteristic of the dataset, when compared to other sequences, are the following facts that inform the design of neural architectures: Each sequence has a fixed length, which enables easy application of a convolutional architecture. Each sequence is relatively short with only 52 steps. Finally, the eight channels of the samples are arranged arbitrarily, though the signals in the channels are very likely correlated.

We evaluate three different neural architectures: Based on the previous work [15], we use a convolutional neural network. We compare this architecture with a recurrent LSTM network, that is traditionally known to perform well with sequence modeling tasks. Finally, the third model does not use sequential data, but the single last sensor reading of the squeezing sequence. A simple MLP architecture processes this data. We make this comparison to evaluate the difference between training a neural classifier on the temporal grasping sequence compared to a static grasp. Figure 5 shows a schematic of all three architectures that resulted from the hyperparameter optimization described in the next section.

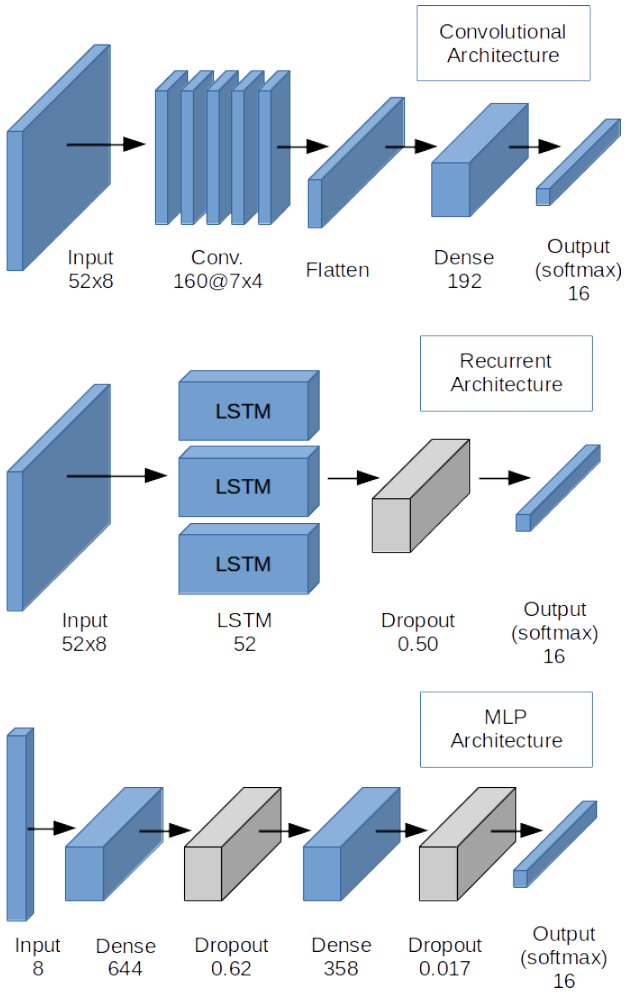


Fig. 5. Three architectures: (top) Convolution over time, (middle) recurrent LSTM, (bottom) MLP. In contrast to the convolutional and recurrent architecture, the MLP's input is only the last of 52 records of the sampling process.

## IV. EXPERIMENTS AND RESULTS

In a first experiment, we evaluate how much each single haptic information channels can discriminate explored objects in isolation. In the main experiment, we develop and optimize a convolutional, a recurrent, and an MLP architecture, for the integration of all three channels of haptic sensory data.

### A. Evaluation of Different Haptic Sensory Channels

To evaluate how much each of the three different haptic channels, the finger position, the finger current and forces at the fingertip, contribute to the achieved classification accuracy we develop neural models for each of these sensory channels in isolation. We split the dataset into three sub-datasets. We use a convolutional architecture with one convolution, and one dense layer informed by preliminary work [15] and perform a 1000-step hyperparameter optimization for each of these sub-datasets using Hyperopt [4]. The search space for the convolutional architecture is described in Table I. For the optimization, the dataset was split in an 80/10/10 ratio into train, validation and test set. Early stopping on the validation

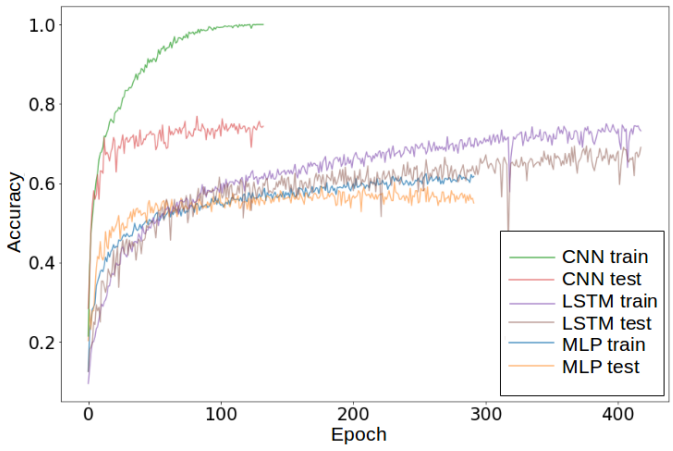


Fig. 6. Training histories for MLP, CNN and LSTM architecture for haptic sensory integration.

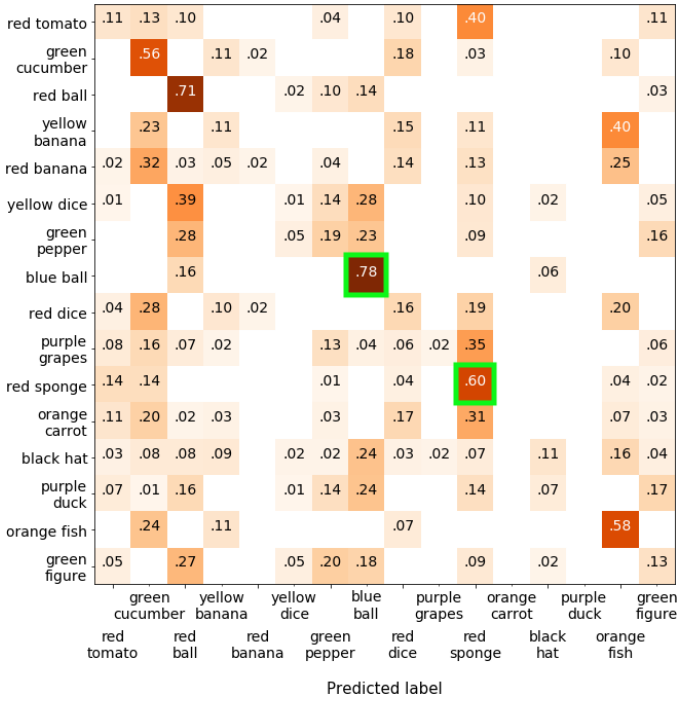


Fig. 7. Confusion matrix for haptic object classification based on finger position only (averaged 10-fold cross-validation). Highly symmetrical objects and objects with a distinctive size like the "blue ball" or "red sponge" are well classified.

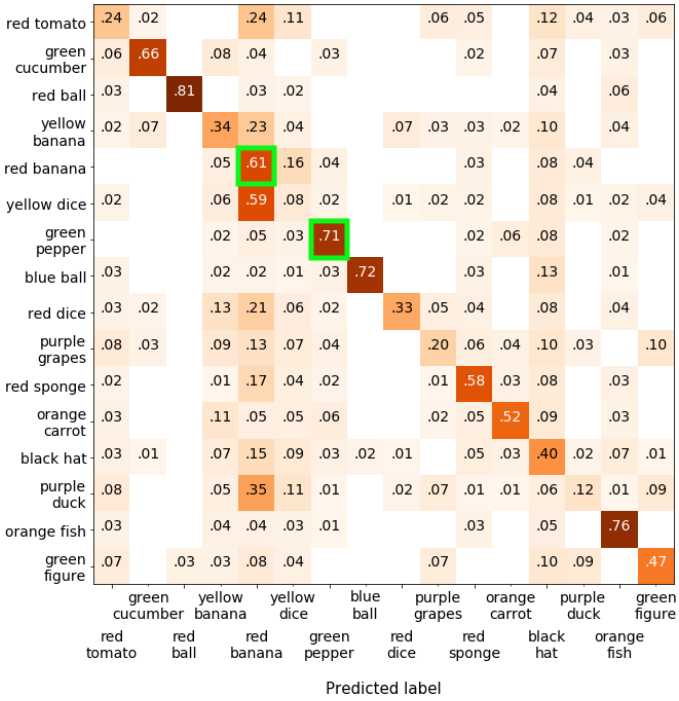


Fig. 8. Confusion matrix for haptic object classification based on tactile fingertip sensor only (averaged 10-fold cross-validation). In comparison to classification based on finger position objects with distinct softness like the soft "green pepper" and the hard "red banana" are classified well.

accuracy with a patience parameter of 50 was used. All models were implemented in Keras [7] and Tensorflow [1].

We perform 10-fold cross-validation for the best performing hyperparameter configurations. We achieve the following accuracies for each of the three sub-datasets: 0.253±0.022 for the position data, 0.525±0.038 for the current data and 0.476±0.022 for the tactile data. Figure 7 shows the confusion matrix for classification based on finger position only. It can be seen that highly symmetrical objects ("red ball") are classified well as they have a constant diameter independent of the exact grasping pose. Also, objects with a distinctively small or large size like the small "red sponge" and the large "blue ball" are recognized well. For comparison, Figure 8 shows the confusion matrix for the classification based on the tactile fingertip sensor. Here, compliance of the object's material contributes strongly to its classification. For instance, the very soft "green pepper" and the hard "red banana" are classified well in contrast to finger position based classification. This result shows that the objects in the dataset are distinguished by different haptic properties.

### *B. Neural Architectures for Integration of Haptic Sensory Channels*

We evaluate three neural architectures for the integration of different haptic sensory channels: convolutional, recurrent and simple MLP. Based on the previous work [15] and preliminary grid-search experiments the number and type of layers for each architecture type was determined and thorough hyperparameter optimization performed. Again, for each architecture, we define a search space for hyperparameters and apply automated optimization using Hyperopt. For all models, the batch size was optimized. For the convolutional architecture, the number of filters and the filter size in x and y dimension and the number of units in the dense layers were optimized. For the LSTM architecture, the number of units in the LSTM layer was optimized. For the MLP architecture, the number of units in the dense layers and the drop out rate were optimized. Table I shows the range of all hyperparameters for all models. We did not optimize the learning rate, as the adaptive learning rate method *Adadelta* [27] was used to train all models. As a loss function for all models, categorical cross entropy was used. All layers use the rectified linear activation function except for the output layer that uses a softmax activation function.

| **Hyperparameter** | **Model type** | **Min** | **Max** |
|---|---|---|---|
| Batch size | All models | 8 | 128 |
| # of filters | Conv. | 16 | 256 |
| kernel sizes x & y | Conv. | 1 | 8 |
| filter size | Conv. | 32 | 256 |
| size of dense layer | Conv. | 16 | 256 |
| # Units in LSTM layer | LSTM | 8 | 64 |
| dropout rates | MLP | 0.001 | 1 |
| # units in dense layers | MLP | 8 | 1024 |

TABLE I
RANGE OF HYPERPARAMETERS FOR OPTIMIZATION.

For the hyperparameter optimization, again, the dataset was split in an 80/10/10 ratio into train, validation and test set.

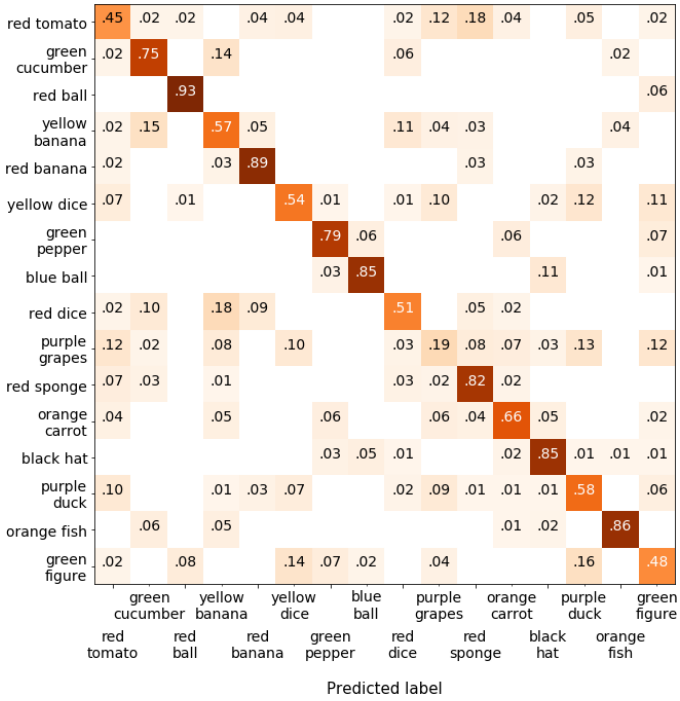


Fig. 9. Confusion matrix for haptic object classification for convolutional architecture (averaged 10-fold cross-validation).

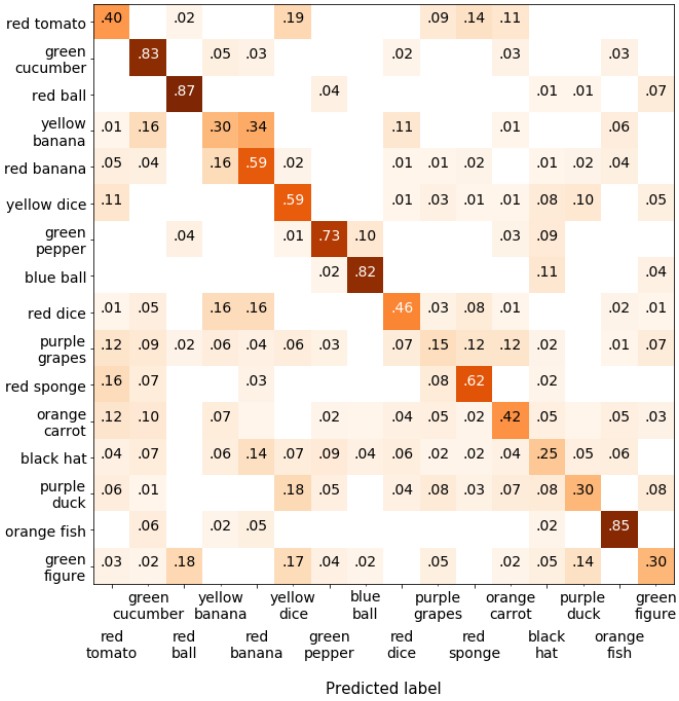


Fig. 10. Confusion matrix for haptic object classification for MLP architecture (averaged 10-fold cross-validation).

The hyperparameter optimization was run for 1000 trials for the convolutional and the MLP architecture and 500 for the LSTM architecture. Again early stopping on the validation accuracy with a patience parameter of 50 was used. As expected, the optimization and training took longest for the LSTM model (159 hours, 38 minutes), significantly shorter for the convolutional model (23 hours, 58 minutes) and shortest for the MLP (8 hours, 58 minutes). The times were measured using an NVIDIA Geforce RTX 2080 TI card, using Cuda Version 10. The optimized architectures were again trained with an 80/10/10 split using the same early stopping criteria as described above. The training and evaluation were repeated in 10-fold cross-validation for each model.

*1) Results: Convolutional Architecture:* The hyperparameter optimization resulted in the architecture shown in Figure 5 (top): One convolutional layer (160 filters of size 7x4) is followed by a flatten layer and a dense layer with 192 units before the output layer with 16 units; batch size is 36. Figure 6 shows the averaged training history of the models. The architecture achieved an average classification accuracy of 0.666±0.029. This result significantly exceeds the classification results of each isolated haptic sensory channel and shows the ability of the neural network for meaningful integration of these sensory channels. Figure 9 shows the averaged confusion matrix for the 16 objects. The convolutional model performed best among all models. It is noteworthy that the optimization process has converged to a broad filter size of seven that almost covers all haptic signal channels. We assume that the additional parameters created in contrast to a 1D-convolution are beneficial for the model.

*2) Results: Recurrent Architecture:* The hyperparameter optimization resulted in the architecture shown in Figure 5 (middle): A recurrent LSTM layer with 52 units is followed by a dropout layer (dropout rate 0.50) before the output layer with 16 units; batch size for training is 28. The architecture achieved 10-fold cross-validation classification accuracy of 0.598±0.058. The performance is slightly lower than that of the convolutional architecture. The averaged confusion matrix does not diverge significantly from the one produced by the convolutional architecture. Training and therefore optimization took significantly longer although the model size (number of trainable parameters) is much lower.

*3) Results: MLP architecture:* The hyperparameter optimization resulted in the architecture shown in Figure 5 (bottom): Two pairs of dense layers (644 and 358 units) followed by a dropout layer (dropout rate 0.62 and 0.017) processes the input before the output layer; the batch size is 76. The architecture achieved a 10-fold cross-validation classification accuracy of 0.534±0.021. Figure 10 shows the averaged confusion matrix for the 16 objects. The performance is lower than the performance of the other two architectures. We see that is evidence that the temporal sequence of haptic sensory data contributes to the haptic classification in comparison to the static data in the state of already having gripped an object. In this experimental setup, objects that are compressed during squeezing to a similar diameter but with different hardness are more likely confused, e.g., the plush “yellow banana” and the hard-plastic “red banana”.

## V. CONCLUSION

We present a large haptic dataset with 83200 haptic measurements, based on 100 samples of each of 16 different objects, every sample containing 52 measurements. The objects are actively explored by a humanoid robot by compressing them in a robotic hand with a tactile sensor. The dataset contains different sensor reading like finger position, finger currents, and forces registered at the fingertip. To address the issue of sparse datasets for robotic haptic perception, we make this dataset publically available.

Based on the dataset we developed different neural models classifying objects based on the haptic information. We achieve an initial baseline classification accuracy of 66.6% in 10-fold cross-validation. We showed that this result is due to the ability of the network to integrate the data both over time domain and over different sensory channels. Neither a single sample from just one-time stamp nor a single sensory channel achieves a comparable result. While some objects differ in their shape and circumference (finger position) others differ in their overall flexibility (finger force) or surface hardness (fingertip sensor). Furthermore, we have shown that these distinguishing features in different haptic properties can be utilized well by neural approaches. These findings underline the ability of artificial neural networks for integration of different sensory (sub-)modalities. We have also shown that convolutional and recurrent architectures achieve comparable performance on the task while showing large differences in training time and number of trained parameters.

Regarding robot design, we have shown that utilizing existing information available in most modern robot motors, like motor position and motor current, significantly adds to the capability of robotic haptic sensing. This source of haptic information does not require extensions of existing hardware.

In future work, we will further extend the robotic hardware by installing one-point haptic sensors in all fingers of the robotic hand. We will evaluate other haptic exploration strategies like lifting, tapping or sliding. We will also include other sensory modalities like sound into our neural models. Finally, we will develop model architectures that combine the compactness of recurrent architectures with the fast training time of convolutional architectures.

## REFERENCES


[1] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, software available from tensorflow.org.

[2] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," *CoRR*, vol. abs/1803.01271, 2018.

[3] P. Barros and S. Wermter, "Developing crossmodal expression recognition based on a deep neural model," *Adaptive Behavior*, vol. 24, no. 5, pp. 373–396, 2016.

[4] J. Bergstra, B. Komer, C. Eliasmith, D. Yamins, and D. D. Cox, "Hyperopt: a python library for model selection and hyperparameter optimization," *Computational Science & Discovery*, vol. 8, no. 1, p. 014008, 2015.

[5] T. Bhattacharjee, H. M. Clever, J. Wade, and C. C. Kemp, "Multimodal tactile perception of objects in a real home," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 2523–2530, 2018.

[6] A. Cangelosi, G. Metta, G. Sagerer, S. Nolfi, C. Nehaniv, K. Fischer, J. Tani, T. Belpaeme, G. Sandini, F. Nori, and Others, "Integration of action and language knowledge: A roadmap for developmental robotics," *IEEE Transactions on Autonomous Mental Development*, vol. 2, no. 3, pp. 167–195, 2010.

[7] F. Chollet *et al.*, "Keras," https://keras.io, 2015.

[8] Y. N. Dauphin, A. Fan, M. Auli, and D. Grangier, "Language modeling with gated convolutional networks," *CoRR*, vol. abs/1612.08083, 2016.

[9] J. Fishel and G. Loeb, "Bayesian exploration for intelligent identification of textures," *Frontiers in Neurorobotics*, vol. 6, pp. 4–23, 2012.

[10] Y. Gao, L. A. Hendricks, K. J. Kuchenbecker, and T. Darrell, "Deep learning for tactile understanding from visual and haptic data," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 536–543.

[11] M. Grunwald, *Human haptic perception: Basics and applications*. Springer Science & Business Media, 2008.

[12] I. Guertel, G. Schillaci, and V. V. Hafner, "Using proprioceptive information for the development of robot body representations," in *Development and Learning and Epigenetic Robotics (ICDL-EpiRob), 2016 Joint IEEE International Conference on*. IEEE, 2016, pp. 172–173.

[13] S. Heinrich, M. Kerzel, E. Strahl, and S. Wermter, "Embodied multimodal interaction in language learning: the emil data collection," in *Proceedings of the ICDL-EpiRob Workshop on Active Vision, Attention, and Learning (ICDL-Epirob 2018 AVAL)*, 2018.

[14] J. H. Kaas, "The functional organization of somatosensory cortex in primates," *Annals of Anatomy-Anatomischer Anzeiger*, vol. 175, no. 6, pp. 509–518, 1993.

[15] M. Kerzel, M. M. M. Ali, H. G. Ng, and S. Wermter, "Haptic material classification with a multi-channel neural network," in *International Joint Conference on Neural Networks (IJCNN)*. Anchorage, Alaska: IEEE, May 2017, pp. 439–446.

[16] M. Kerzel, E. Strahl, S. Magg, N. Navarro-Guerrero, S. Heinrich, and S. Wermter, "NICO – Neuro-Inspired COmpanion: A developmental humanoid robot platform for multimodal interaction," in *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2017, pp. 113–120.

[17] R. L. Klatzky, S. J. Lederman, and D. E. Matula, "Haptic exploration in the presence of vision," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 19, no. 4, pp. 726–743, 1993.

[18] S. J. Lederman and R. L. Klatzky, "Hand movements: A window into haptic object recognition," *Cognitive Psychology*, vol. 19, no. 3, pp. 342–368, 1987.

[19] S. Luo, J. Bimbo, R. Dahiya, and H. Liu, "Robotic tactile perception of object properties: A review," *Mechatronics*, vol. 48, pp. 54–67, 2017.

[20] A. Murali, Y. Li, D. Gandhi, and A. Gupta, "Learning to grasp without seeing," *arXiv preprint arXiv:1805.04201*, 2018.

[21] J. M. Romano and K. J. Kuchenbecker, "Methods for robotic tool-mediated haptic surface recognition," in *2014 IEEE Haptics Symposium (HAPTICS)*. IEEE, 2014, pp. 49–56.

[22] G. Salvietti, "Replicating human hand synergies onto robotic hands: A review on software and hardware strategies," *Frontiers in Neurorobotics*, vol. 12, p. 27, 2018.

[23] J. Sinapov, V. Sukhoy, R. Sahai, and A. Stoytchev, "Vibrotactile recognition and categorization of surfaces by a humanoid robot," *IEEE Transactions on Robotics*, vol. 27, no. 3, pp. 488–497, 2011.

[24] Y. Tada, K. Hosoda, and M. Asada, "Sensing ability of anthropomorphic fingertip with multi-modal sensors," in *IEEE International Conference on Intelligent Robots and Systems*. Citeseer, 2004, pp. 1005–1012.

[25] S. Takamuku, G. Gomez, K. Hosoda, and R. Pfeifer, "Haptic discrimination of material properties by a robotic hand," in *2007 IEEE 6th International Conference on Development and Learning*. IEEE, 2007, pp. 1–6.

[26] N. Wettels, V. J. Santos, R. S. Johansson, and G. E. Loeb, "Biomimetic tactile sensor array," *Advanced Robotics*, vol. 22, no. 8, pp. 829–849, 2008.

[27] M. D. Zeiler, "Adadelta: an adaptive learning rate method," *arXiv preprint arXiv:1212.5701*, 2012.