Learning Interpretable Representations for Visuotactile Sensors

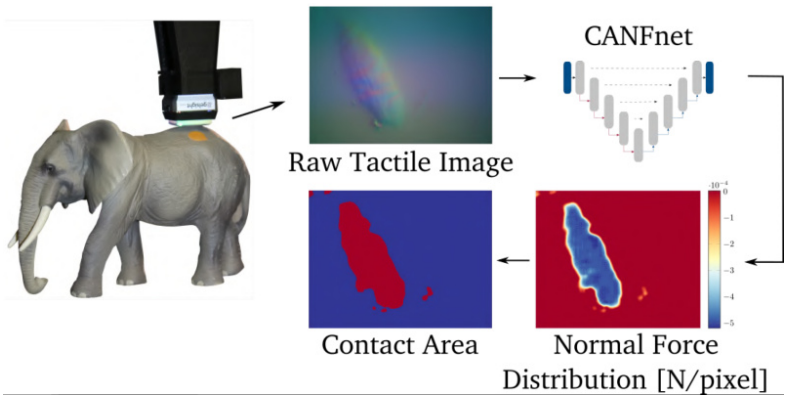


Figure 1: Overview of the proposed method. A raw tactile image is passed through the CANFnet, which outputs a normal force distribution. Subsequently, the contact area is estimated via straightforward thresholding.

Project Manager
Paul-Otto Müller

Researchers
Niklas Funk, Dr. Boris Belousov and Dr.-Ing. Anton Savchenko

Principal Investigator
Prof. Jan Peters (PhD)

Project Term
2022 - 2023

Clusters
Lichtenberg Cluster Darmstadt

Software
PyTorch

Additional Software
ROS Noetic, Motive (OptiTrack)

Institute
Intelligent Autonomous Systems

University
Technische Universität Darmstadt

Introduction

Visuotactile sensors are gaining momentum in robotics because they provide high-resolution contact measurements at a fraction of the price of conventional force/torque sensors. It is, however, not straightforward to extract useful signals from their raw camera stream, which captures the deformation of an elastic surface upon contact. To utilize visuotactile sensors more effectively, powerful approaches are required, capable of extracting meaningful contact-related representations. This work proposes a neural network architecture called CANFnet (Contact Area and Normal Force) that provides a high-resolution pixelwise estimation of the contact area and normal force given the raw sensor images. The CANFnet is trained on a labeled experimental dataset collected using a conventional force/torque sensor, thereby circumventing material identification and complex modeling for label generation. We test CANFnet using commercially available DIGIT and GelSight Mini sensors and showcase its performance on real-time force control and marble rolling tasks. We are also able to report generalization of the CANFnets across different sensors of the same type. Thus, the trained CANFnet provides a plug-and-play solution for pixelwise contact area and normal force estimation for visuotactile sensors.

Methods

We propose the following procedure to collect pixelwise labeled experimental data with standard, established hardware. We mount an object with known geometry, that we denote as indenter, on top of a Force/Torque (F/T) sensor and attach the

visuotactile sensor at the end-effector of a robot manipulator. The underlying idea is to bring the sensor and indenter into a contact configuration where the normal force acts uniformly upon the gel. This allows obtaining pixelwise labels by dividing the total measured normal force from the F/T sensor by the contact area. Specifically, we can obtain pixelwise normal force labels as follows. We represent the geometry of the indenters' flat contact area through characteristic vectors. By applying the pinhole camera model, we can project these vectors subsequently onto the rectified 2D sensor image which indicate the contact area. Finally, each pixel that is inside of the projected contact area is assigned the force value given by dividing the measurement of the F/T sensor over the contact area in pixels. After collecting the data, the CANFnet, based on an U-Net, can be trained supervised. Due to the data size of about 200000-300000 samples and model capacity, the Lichtenberg High Performance Computer was used for training the models.

Results

Real-world experiments establish the effectiveness of our proposed CANFnet architecture in combination with the corresponding training data generation procedure. We validated the real-time capability of our network architecture and the quality of the pixelwise normal force and contact area estimation in various scenarios, ranging from statically interacting with novel indenters to more dynamic force control and a marble roll task requiring both normal force and spatial information. Finally, we observe that the CANFnets can generalize to scenarios that are substantially different from the training data distribution, i.e., interacting with more realistic, non-flat objects, thereby creating non-planar contact configurations. Throughout all the experiments, the CANFnet trained for the GelSight sensor outperforms the DIGIT version. It provides good accuracy in estimating the pixelwise normal force acting onto the sensor, with mean absolute errors for the total force below 1N in all our experiments except one case of large indenters and high force range. What is even more encouraging is that this model generalizes seamlessly to the test sensor.

Discussion

Evaluating the DIGIT CANFnet using the training sensor only yields slight losses in performance compared to the GelSight CANFnet. However, the difference in performance could also originate from the fact that DIGIT slightly violates one of our assumptions for labeled data generation: namely, that the sensors have a flat gel. While the DIGIT CANFnet can be transferred to the test sensor and handle a marble roll task, which requires both force and spatial information to roll a marble to a desired position on the tactile gel, we still observe a decrease in performance. Future work should investigate whether additional data augmentation or training data collection using an ensemble of sensors could mitigate this issue. Furthermore, a more thorough analysis of CANFnet for non-parallel contact configurations, possibilities of incorporating the

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN TECHNISCHE UNIVERSITÄT DARMSTADT UNIKASSEL VERSITÄT Philipps Universität Marburg

estimation of tangential forces across the sensor surface, and the use of high-resolution tactile representations in more complex downstream tasks provide potential directions for future work.

Last Update: 2023-04-04 15:33