

LEARNING TO DETECT SLIP USING BAROMETRIC TACTILE
SENSORS

by

Abhinav Grover

A thesis submitted in conformity with the requirements
for the degree of Master of Applied Science
Graduate Department of Aerospace Science and Engineering
University of Toronto

© Copyright 2021 by Abhinav Grover

Abstract

Learning to Detect Slip Using Barometric Tactile Sensors

Abhinav Grover

Master of Applied Science

Graduate Department of Aerospace Science and Engineering

University of Toronto

2021

The ability to perceive object slip through tactile feedback allows humans to accomplish complex manipulation tasks. For robots, however, detecting key events such as slip from tactile information is a challenge. This work explores a learning-based method to detect slip using barometric tactile sensors that have many desirable properties; they are durable, highly reliable, and built from inexpensive components. We collect a novel dataset specifically targeted for robustness, and train a TCN to detect slip. The trained detector achieves an accuracy of greater than 91% on test data while displaying robustness to the speed and direction of the slip motion. When tested on two robot manipulation tasks involving a variety of common objects, our detector demonstrates generalization to previously unseen objects. This is the first time that barometric tactile-sensing technology, combined with data-driven learning, has been used for a manipulation task like slip detection.

Acknowledgements

I joined the Master's program at the University of Toronto hoping to delve deeper into a topic in robotics and gain valuable experience conducting academic research. As I am writing this thesis nearing the conclusion of my program, I realize that I have gained so much more. I would like to take this opportunity to thank the individuals who made this work possible and, at the same time, made my Master's experience all the more cherishable. First and foremost, I would like to thank my advisor, Jonathan Kelly, who welcomed me to his lab and supported me throughout the program. Jon, your guidance and feedback were invaluable to complete this work, and the nurturing environment you foster helped me overcome many research-related as well as personal obstacles. Even during the hardest of times in the COVID-19 pandemic, your calmness and playful attitude helped me, and many others, sustain a sense of normalcy and positivity; thank you for being an excellent advisor and an even better mentor. I would also like to thank Dr. Steven Waslander for reviewing this document and providing constructive feedback. Thank you to my lab-mates: Brandon, Emmett, Matthew, Filip, Adam, Oliver, Trevor, Olivier, Chris, Philippe, and Andrej, who were always ready to help in any way possible, and made the experience so much more enjoyable despite the pandemic. Chris, you were a constant source of support and encouragement, and helped me sustain momentum during the lowest of days. Our weekly research meetings and our frequent discussions in the lab were a great source of feedback, which often prompted introspection. Philippe, thank you for being so supportive of my ideas and for all your help with the publications associated with this work. Trevor and Oliver, thank you for always being patient with me and addressing my barrage of questions. To my family: Dadi, Dadu, Mama, Papa, and Kanav, thank you for always being there and supporting me through this process. You helped me get through the hardest of times with your unconditional love and encouragement. I owe everything to you and I am incredibly lucky to have all of you in my life. Finally, I want to thank my girlfriend who has been consistently by my side, figuratively speaking, and has been a huge pillar of support during my Master's. Words are not enough to express my gratitude for your patience, love, and compassion; thank you for everything.

Contents

1	Introduction	1
2	Background	5
2.1	Human and Artificial Tactile Sensing	5
2.2	Phenomenon of Slip	7
2.3	Deep-Learning for Sequence Modeling	9
2.3.1	Supervised Deep-Learning	9
2.3.2	Temporal Convolution Networks	11
3	Related Work	13
3.1	Vibration-based Methods	14
3.2	Motion-Based Methods	15
3.3	Visualizing Deformation	16
3.4	Learning-Based Methods	17
4	Learned Slip Detector	19
4.1	Tactile Sensors and Gripper	20
4.2	Data Acquisition	22
4.2.1	Slip Data Recorder	22
4.2.2	Data Collection and Distribution	25
4.3	Slip Detection Network	28
4.3.1	Network Design	28
4.3.2	Data Augmentation	30
4.3.3	Training Details	30
4.3.4	Parameter Tuning	31
4.3.5	Training Result	33

5	Experiments and Analysis	35
5.1	Classification Performance Comparison	35
5.2	Detection Performance Analysis	37
5.2.1	Sensitivity Analysis	37
5.2.2	Latency Analysis	38
5.3	Robot Experiments	40
5.3.1	Mallet Tap Test	40
5.3.2	Object Lift Test	44
6	Conclusion	49
6.1	Summary and Contributions	49
6.2	Potential Improvements	50
6.3	Future Work	50
	Bibliography	52

List of Tables

2.1	Properties of human mechanoreceptors [1]	6
4.1	The distribution of training data by slip type, slip speed, slip direction, and surface curvature.	27
4.2	Slip detection performance of the TCN on the test dataset.	34
5.1	Performance comparison between TCN and frequency-based methods on test data. The weighted average of the classes is used to compute each metric.	36
5.2	TCN performance, measured by F1-score, with variation in slip type, slip speed, slip direction, and surface curvature.	38
5.3	TCN detection latency distribution.	39
5.4	Surface properties of test objects	40
5.5	Slip-detection results for two real-world manipulation experiments. . .	44

List of Figures

1.1	The TakkTile sensing units mounted on the fingertips of a Robotiq 3-finger adaptive gripper.	2
2.1	Classification of the artificial sense of touch	6
2.2	Changes in the co-efficient of friction for a general sliding object under constant force [2].	8
2.3	Visualization of a stack of dilated causal convolution layers [3].	11
2.4	Residual block used in [4] and the building block of TCNs in our work.	12
3.1	Components of a basic vision-based tactile sensor.	17
4.1	Robotiq three-finger adaptive gripper, retrofitted with the TakkTile sensor kit.	20
4.2	Power grasps of the Robotiq gripper.	21
4.3	Perspective view and front view of the TakkTile fingertip.	22
4.4	Functional flow block diagram of the data recording setup emphasizing the exchange of messages between the constituent components.	24
4.5	Surfaces used for slip data acquisition.	27
4.6	Data collection trial.	28
4.7	The TCN architecture used for slip detection.	29
4.8	F1-score of slip detection TCN on the test data versus the temporal window size.	32
4.9	Precision versus Recall curve for changing binary classification threshold, calculated on the test data.	32
4.10	Validation accuracy and loss curve during training of the TCN for 800 epochs over the entire training set.	33

5.1	Slip probability response of two detectors on test data. The fingertip was executing a translational slip maneuver on a spherical surface. . .	36
5.2	Test objects for real-world experiments.	41
5.3	Mallet-tap test setup with plastic ball.	42
5.4	Mallet-tap test for one trial where the plastic ball was tapped from above and from the side.	43
5.5	Object-lift test setup.	45
5.6	Object-lift test with the cardboard can, showcasing a successful trial.	45
5.7	Object-lift test with the foam sleeve, showcasing a failure trial.	46

Chapter 1

Introduction

For humans, the sense of touch plays a crucial role in perceiving the outside world. We are able to perform complex dexterous manipulation tasks with our hands in a large part because of their rich tactile sensing capabilities. The skin, which is our primary cutaneous tactile modality, is equipped with numerous mechanoreceptors that transduce mechanical stimuli, like pressure and vibration, into electrical signals. For decades, the prosthetics and robotics community has sought to match the human tactile-sensing capabilities artificially. Despite significant recent progress, artificial tactile sensors are yet to achieve the fidelity and accuracy of human tactile perception.

The sense of touch is essential for controlling the gripping force required to hold an object without slipping, that is, without allowing relative motion at the contact interface between the hand and an object [5]. As part of humans' sensory nervous system, a fast somatosensory feedback loop enables grasp adjustments to be performed within about 100 milliseconds, allowing for seamless and automatic grasp adaptation when handling a wide variety of objects in our daily lives [6]. Tactile signals provide vital information about slip faster than any exteroceptive perception method such as vision. Slip can be both disastrous (e.g., when transporting a fragile object) and advantageous (e.g., when moving an object without lifting it) depending on the context and the task [7]. In robotics, the well-studied “handover” task—in which a robot hand passes an object to a human hand—requires control of the gripping force with accuracy and speed to avoid significant slip [8]. The requisite feedback can only be provided through tactile sensing [9] and, consequently, the detection and control of slip events is fundamental to the completion of handovers and many other relevant robot and human-robot tasks.

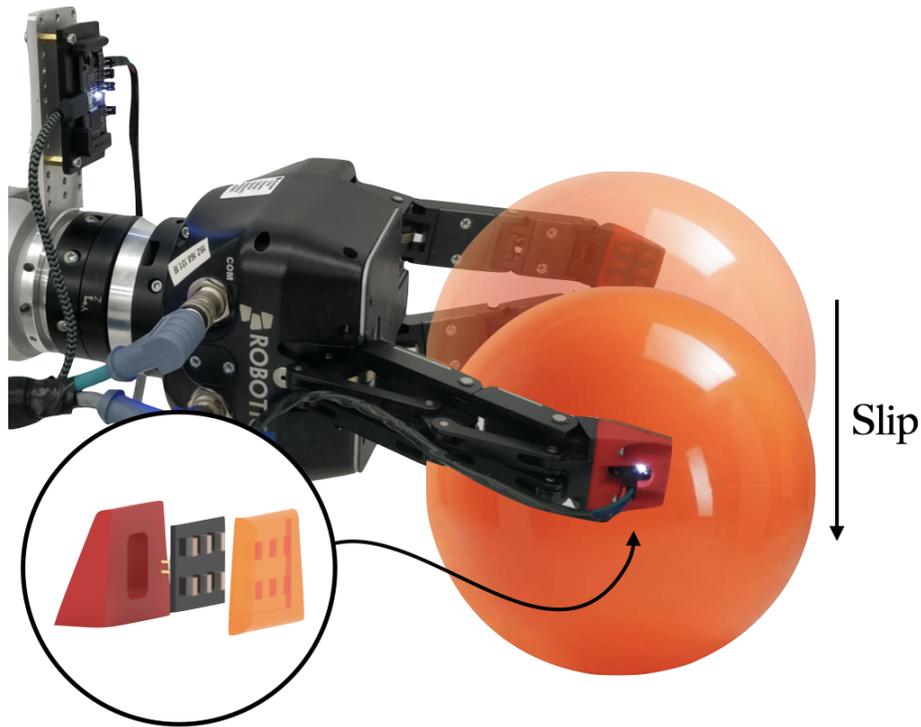


Figure 1.1: The TakkTile sensing units are mounted on the fingertips of a Robotiq 3-finger adaptive gripper. Inset: exploded view of each RightHand Labs TakkTile sensor (plastic support frame, barometer circuit board, and rubber matrix, from left to right).

In the last decade, a wide range of new tactile sensors have become available. These sensors measure various physical properties at the sensing interface including capacitance [10], impedance [11], or optical changes [12–15]. The BioTac fingertip, for example, is equipped with impedance-based tactile receptors, hydrophones, and thermistors to provide rich multimodal tactile information [11]. Each new tactile sensor has inherent characteristics such as fragility, bulk, resolution, nonlinearity, hysteresis, and production cost. The best tactile sensor for a specific job is one that balances these factors with how they relate to varying task constraints.

Inspired by human biology, many early researchers designed custom pressure sensors to analyze the vibrations induced by slip (see Section 3.1), in which the presence of slip was determined based on heuristics applied to properties like spectral power or vibration energy. These approaches and their heuristics fell short in their ability to generalize to previously unseen tasks as a consequence of the sensitive heuristics. Currently, a popular approach is to infer slip by visually detecting ‘skin’ deformation (see

Section 3.3). The top layer of a vision-based tactile sensor produces visible features when deformed, which can be used to infer tactile events like slip. The vision-based tactile sensors, however, are bulky and using them to detect slip is computationally intensive [1].

This thesis investigates the potential of combining low-cost tactile sensors, assembled from off-the-shelf components, with a state-of-the-art neural network to detect object slip. The sensors—the TakkTile model from RightHand Labs—are built as an array of commercial MEMS barometers fixed to a PCB backplane, with a thin rubber matrix forming the contact surface. Although inexpensive, these sensors have a low profile, are mechanically robust, exhibit a consistent linear pressure response, and integrate easily with existing end-effectors.¹ The complex spatiotemporal signature of pressure changes during slip is difficult to model analytically—instead, a data-driven approach is taken by training a temporal convolution neural network (TCN) to classify the time-series data produced by the tactile sensor as either static or slipping.

To the best of the author’s knowledge, this work is the first to use either barometric tactile sensors or a TCN-based network for slip detection. It provides the following research contributions:

- an algorithm for slip detection using very low-cost barometric sensors that achieves an average accuracy of over 91%;
- a comparison of the TCN approach with two prior slip-detection methods that rely on vibration information;
- a preliminary analysis of the sensitivity and robustness of the TCN detector to factors related to surface properties and slipping motion;
- extensive experimental results for in-hand slip detection of objects with various curvatures, hardnesses, and surface properties.

¹The MEMS barometer used in this work (NXP MPL115A2) costs US\$2, approximately, per unit in quantities of 1,000 or more.

Chapter 2

Background

The chapter provides an overview of methods for artificial tactile sensing, with a focus on their relationship to the human sense of touch (in terms of fidelity); a definition and characterization of the phenomenon of slip, the primary topic of the thesis; and a review of deep learning and sequence modeling, which highlights appropriate models for slip detection. The following chapter, Chapter 3, examines recent developments in artificial tactile sensing and slip detection for various use cases.

2.1 Human and Artificial Tactile Sensing

The human hand is a symbolic representation of the evolution of primates. With twenty-one degrees of freedom (DOFs), our hand is an incredibly dexterous machine with a wide range of capabilities—to have an idea, one may consider the variety of tasks, such as writing, hammering, playing music, and so on, that humans do on a daily basis. Considering any of the mentioned tasks, it is obvious that the sense of touch is a crucial component of this machinery. Though usually unconsciously, humans use this sense to recognize many properties of a touched object: shape, motion, temperature, roughness, hardness. Such capabilities have historically played a key role in the evolutionary trajectory of humans as a species.

The human sense of touch comprises of two main sub-modalities: cutaneous (in which inputs come from the receptors in the skin), and kinesthetic (in which inputs come from the receptors in muscles and tendons) [16]. Mechanical forces and changes of muscle lengths are transduced into electrical signals by receptors in our skeletal muscles and tendons; *muscle spindles*, for example, are sensory fibers that spiral

Table 2.1: Properties of human mechanoreceptors [1]

Name	Receptor Type	Field Size (mm^2)	Encoded Quantity
Meissner Corpuscles	I (fast)	12.6	High frequency vibrations (<50Hz) and accelerations
Pacinian Corpuscles	II (fast)	101	High frequency vibrations (>50Hz)
Merkel Disks	I (slow)	11	Static load, skin indentation
Ruffini Endings	II (slow)	59	Skin stretch, stretch direction

around muscle fibers and sense flexion or contraction of muscles [17]. Similarly, to transduce cutaneous stimuli into electrical signals, our skin is embedded with multiple types of *mechanoreceptors* (see Table 2.1), for example, information about static loads on human skin is encoded by mechanoreceptors called *Merkel Disks* [1]. In the context of this thesis, the definition of tactile sensing is restricted to the contact-level transduction of cutaneous stimuli, which are mainly comprised of contact forces.

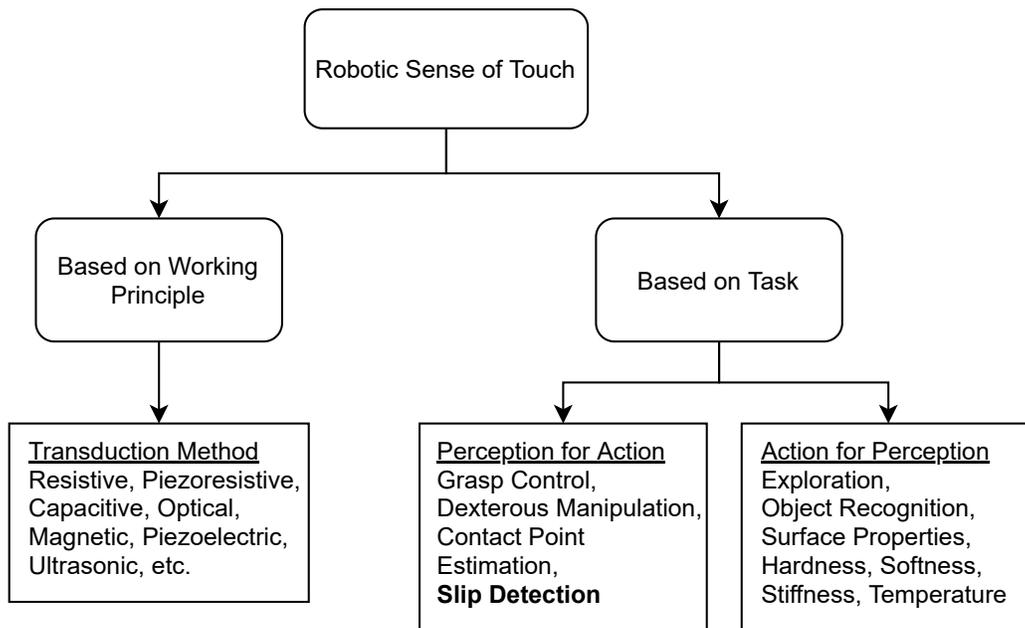


Figure 2.1: Classification of the artificial sense of touch bifurcated based on task and working principle [18].

Over the last four decades, the prosthetics and the robotics research communities have tried to mimic the human sense of touch through artificial tactile sensing. One

line of this research focuses on the development of appropriate transduction technologies (see Figure 2.1), several types of which have been developed over the years: including, but not limited to, resistive, capacitive, piezoelectric, optical, and magnetic [18]. In the early years, researchers designed and fabricated custom electronic components capable of transducing tactile inputs but, more recently, they have favored the use of off-the-shelf electronic components, as they are more reliable and cost-effective. The solid-state barometer, manufactured as a micro-electromechanical system (MEMS), is one such off-the-shelf electronic device capable of transducing atmospheric pressure into electrical signals using a piezoresistive membrane. Tenzer et al. [19] designed, fabricated, and tested a tactile sensing unit that uses off-the-shelf MEMS barometers to transduce contact forces. The barometers are embedded in a layer of urethane rubber, which helps to transmit surface level forces and causes the barometer membrane to flex. The tactile sensing technology created by Tenzer et al. forms the basis for this thesis.

Another line of the research in artificial tactile sensing focuses on inferring useful information from the transduced electrical signals. Based on the task to be accomplished, there are two broad categories of artificial tactile sensing (see Figure 2.1): action for perception—inferring properties of objects through active movement—and perception for action—grasp control, dexterous manipulation, slip detection, and so on. Of these well-studied tasks, slip detection, a subsection of the perception for action category, has remained a challenge (see the survey paper on slip detection by Romeo et al. [1]). In this work, we propose a novel method for solving the task of object slip detection using the recently developed barometric tactile sensing technology.

2.2 Phenomenon of Slip

Object manipulation is a complex task that requires perception of contact details such as surface properties, object motion, and contact forces. Detecting object slip in manipulation scenarios requires understanding the physical phenomenon of slip.

For a given tactile surface F and an object surface S , the portion of S in contact with F is called the contact surface C . Slip occurs when there is relative motion between a point $p \in F$ and a point $q \in S$, where $p, q \in C$, such that the motion is tangential to the contact surface C [20]. There are two types of slip that can occur: *incipient slip*, in which the displacement is localized to a narrow region on C , and *total*

slip, in which displacement occurs throughout C . For this work, we focus our efforts on the task of detecting total slip of an in-hand object during robotic manipulation.

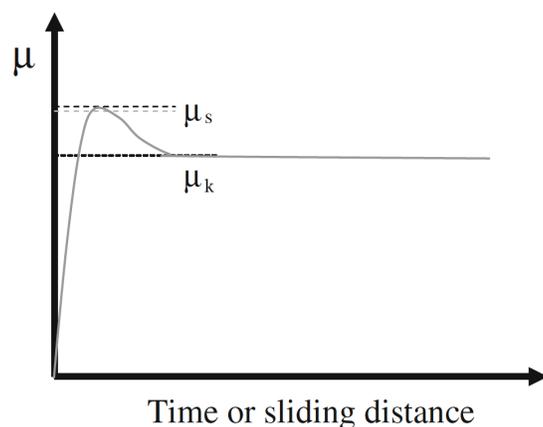


Figure 2.2: Changes in the co-efficient of friction for a general sliding object under constant force [2].

From a dynamics perspective, object slip is a consequence of an imbalance between the contact force and other body forces acting tangentially in the object’s frame of reference. Since friction is the only contact force that acts tangentially, the occurrence of slip is caused by insufficient friction leading to relative motion. According to classical mechanics, friction is a product of the coefficient of friction (which depends on the surfaces in contact) and the normal contact force. The coefficient of friction μ changes throughout the slip motion—at the inception of slip, the coefficient rises to its the static value (μ_s) but soon settles to its the kinetic value (μ_k), as shown in Figure 2.2. Slip can also occur due to an abrupt reduction of the normal contact force. For slip to occur, at least one of the following conditions needs to be met:

1. Insufficient normal gripping force (due to incorrect estimation of object properties);
2. Sudden increase in tangential force (due to an unexpected perturbation or collision);
3. Reduction of frictional coefficient (such as wet or slippery surfaces) [20].

2.3 Deep-Learning for Sequence Modeling

Perceiving the world through the sense of touch is a non-trivial task learned in early childhood. To accomplish this, human skin is embedded with two types of fast-adapting mechanoreceptors (see Table 2.1) that respond to accelerations and high frequency vibrations, that is, dynamic stimuli. During a slip event, each mechanoreceptor encodes the spatiotemporal tactile information as neural stimuli, which elicits a response from the central nervous system in an estimated ~ 100 ms [21]. Though the physiology of the transduction mechanisms embedded in human skin is well understood, the mechanism of tactile perception is still an active topic of research. In particular, deep learning methods, which were developed to mimic the learning process of the human nervous system, have a strong potential to reliably detect slip using artificial tactile sensing.

Recently, as computation costs have diminished while highly-supported open-source libraries have been developed, the field of deep-learning has seen a resurgence. There are many hard tasks where deep learning-based approaches boast the state-of-the-art performance, the *ImageNet* large-scale visual-recognition challenge being one of the earliest. The applications of deep learning are far-reaching, especially in the field of autonomous robotics in which there is an abundance of sensor data. An extensive overview of the field is available in [22].

As described above in Section 2.2, object slip is a non-trivial spatiotemporal event. During slip, artificial tactile sensors provide information about the contact locations and forces, making the data rich in spatial and temporal features. If an appropriate representation of this time-series data is learned, it may be possible to distinguish between slipping and stable scenarios. Therefore, slip detection can be considered as a sequence modeling problem, and there are many deep learning models specialized for this task, including recurrent networks, transformers, and TCNs. This section covers the fundamentals of supervised deep learning and concludes with an explanation of sequence modeling using TCNs.

2.3.1 Supervised Deep-Learning

Deep neural networks (DNNs) are a class of machine-learning models that represent a non-linear mapping $\hat{\mathbf{y}} = f(\mathbf{x}; \boldsymbol{\theta})$, where $\mathbf{x} \in \mathbf{X}$ is the input and the mapping f is parameterized by $\boldsymbol{\theta}$. The basic building block of a DNN is a neuron, which is simply

a parameterized function that maps an input of a given size onto an output of size one. Multiple neuron units are stacked together to form a network layer, which can be represented by a single vector function (see Equation (2.1)) Each layer takes in an input \mathbf{z}_n , applies an affine transformation, followed by a non-linear function σ (commonly referred to as the *activation function*) to produce the output \mathbf{z}_{n+1} (here, $\mathbf{z}_0 = \mathbf{x}$). Adding connections between many such layers forms a DNN. DNNs are classified into various model types based on how these neuron layers are connected.

$$\mathbf{z}_{n+1} = \sigma(\mathbf{W}_n \cdot \mathbf{z}_n + \mathbf{b}_n) \quad (2.1)$$

The activation functions σ define the non-linear mapping from $\mathbf{R}^m \rightarrow \mathbf{R}^n$ that add the required non-linearity to DNNs. Currently, there are many popular activations functions, namely: rectified linear unit (ReLU), sigmoid function, hyperbolic tangent function (tanh), and many more.

$$\text{ReLU}(x) = \max(0, x). \quad (2.2)$$

While the ReLU is the most popular activation function, it suffers from the dying ReLU problem [23].¹ Many variations of ReLU have been developed specifically to solve this problem, such as the exponential linear unit (ELU), the leaky ReLU, and the scaled exponential linear unit (SELU) [24]. One way to learn from data is to use the *supervised learning* paradigm; the goal is to learn a functional mapping from inputs to known pre-collected outputs. In the case of supervised learning with DNNs, the goal is to learn the desired mapping $\mathbf{y} = f^*(\mathbf{x}; \boldsymbol{\theta}^*)$ that maps any input $\mathbf{x} \in \mathbf{X}$ to its corresponding expected output \mathbf{y} . A loss function \mathcal{L} is the measure of the difference between the network output and the true output. The learning process is a minimization of the loss function \mathcal{L} over the network parameters $\boldsymbol{\theta}$,

$$\boldsymbol{\theta}^* = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \mathcal{L}(f(\mathbf{x}; \boldsymbol{\theta}), \mathbf{y}). \quad (2.3)$$

The function $f(\mathbf{x}; \boldsymbol{\theta})$, which represents a DNN, is differentiable with respect to the parameters $\boldsymbol{\theta}$. As a consequence, a first-order optimization algorithms like stochastic gradient descent (SGD) can be used to minimize the loss function \mathcal{L} . SGD is an

¹The dying ReLU problem occurs during network training when a large gradient causes the neuron weights to change in such a way that the the input become less than 0 and the neuron never activates again, causing it to “die”.

iterative optimization approach; in each iteration, the gradient of the loss function is calculated with respect to the network parameters θ , and each parameter is updated in a direction opposite to the corresponding gradient. The user is required to pick the learning rate as a hyperparameter and certain variations of SGD need additional hyperparameters to be set. The Adam optimization algorithm [25], which is an extension of SGD, is used for training the DNNs for this thesis.

2.3.2 Temporal Convolution Networks

Convolution is a widely used mathematical operation in the field of signal and image processing. Analog signals can be convolved with various kernel functions to identify specific signal features. Similarly, digital images can undergo discrete convolution with two-dimensional kernels to identify features in image space; this operation forms the basis for Convolutional Neural Networks (CNNs). Both these operations—continuous and discrete—are differentiable.

As demonstrated by Oord et al. [3], digital audio signals can be generated with a DNN that uses causal convolution (when convolution is performed along the time dimension) in its discrete form. More recently, in 2018, Bai et al. [4] brought dilated causal convolution to prominence in the world of deep-learning when they empirically showed that neural networks with causal convolution outperformed Recurrent Neural Networks (RNNs) on sequence modeling tasks.²

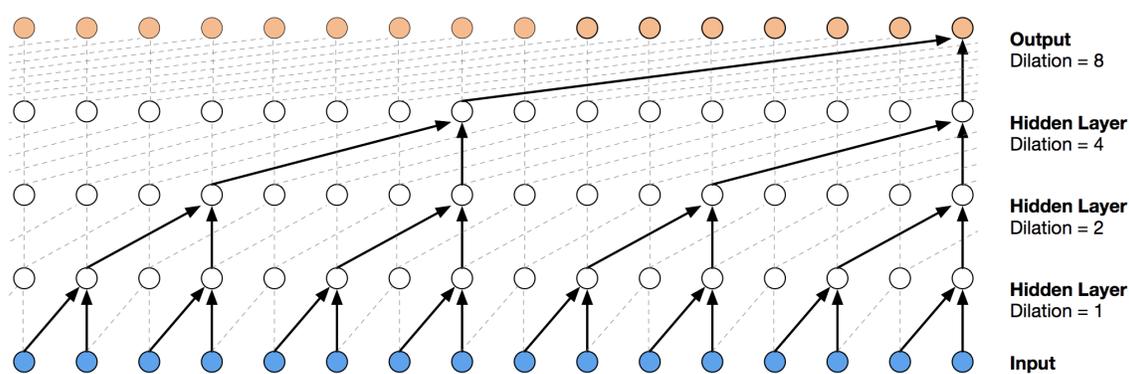


Figure 2.3: Visualization of a stack of dilated causal convolution layers [3].

Causal convolution, represented visually in Figure 2.3, leads to a non-linear op-

²Dilated causal convolution was first introduced by Oord et al. [3]

eration along the temporal dimension, whereas the dilation at each layers leads to a widening of the receptive field as the layers go deeper. To improve the training and inference speed, Bai et al. pushed the dilated causal convolution idea further and introduced the temporal convolution network (TCNs), which is comprised of multiple residual blocks, stacked like network layers, and where the causal dilation of each block increases with network depth. The residual block includes dropout and weight normalization along with residual connections in order to facilitate regularization. Figure 2.4 shows the diagram of a residual block with dilated causal convolutions. In this research, a TCN is utilized to extract temporal features from the barometric tactile data for the purpose of detecting slip events.

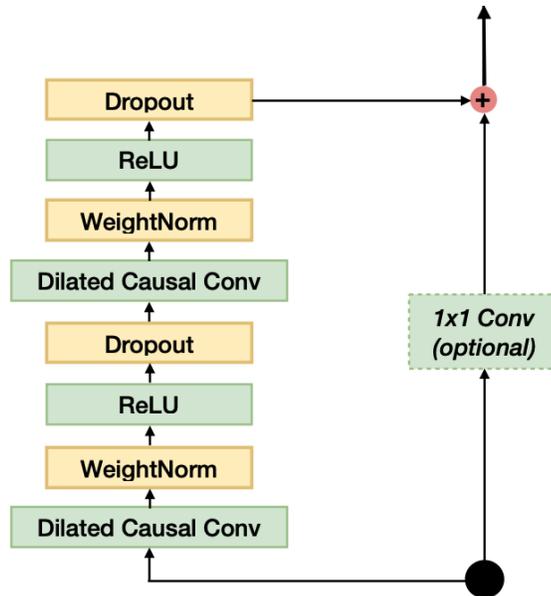


Figure 2.4: Residual block used in [4] and the building block of TCNs in our work.

Chapter 3

Related Work

The study of tactile sensing for robotic systems has an extensive history, stretching back more than five decades (see, e.g., [26] and [27] for pioneering work). This research effort has been driven, in large part, by an evolving understanding of the essential role played by the sense of touch in human dexterous manipulation [16], and slip detection is a rudimentary part of the perception-for-action paradigm. This chapter discusses prior research in the field and contrasts prior approaches with the one adopted in this thesis. A comprehensive review of slip detection is provided by [1] and [20]. For a comprehensive review of existing sensors types, see [28].

A basic slip detector simply outputs a binary value indicating whether there is slippage or not. Some methods detect the inception of slip (characterized by movement in small regions), while others detect total slip (movement across the whole contact surface). Methods of greater complexity can also provide information about the direction and speed of slip. The large set of slip detection methods can be classified in terms of: (1) the contact parameter to be monitored (displacement, micro-vibration, vision etc.) or (2) the transduction mechanism used (capacitive, piezoelectric, magnetic, optical, thermal, etc.) [20]. In this chapter the taxonomy is based on the monitored parameter, while close attention is paid to the transduction mechanism.

When a contact parameter is selected to be monitored for slip detection, an unintended restriction is imposed on the information available to the detector. Recent research has sought to resolve this limitation by applying learning-based methods directly to the sensor data. The last section in this chapter provides an overview of learning-based approaches that employ popular machine-learning tools for slip detection.

3.1 Vibration-based Methods

Early methods for slip detection often relied on an analysis of the vibration pattern induced by object slip, where the frequency of the vibration (caused by material resonance) depends on the composition of the surfaces in contact. A common transduction element used for this purpose is a piezoelectric film, which translates material deformation directly into electric signals. As one of the first attempts to detect slip, Salisbury et al. [29] and Baits et al. [30] implemented slip compensation on a prosthetic hand using thumb-mounted piezoelectric crystals and a vibration analysis circuit. In a later paper by Mingrino et al. [31], the authors used a piezoelectric film in conjunction with a three-axial force sensor to detect slip and showed that the amplitude of the piezoelectric signal grows with slip speed. In a more recent paper by Cotton et al. [32], a thick-film piezoelectric sensor observed useful signal in the frequency band of 200–1000 Hz during slip.

Vibrations due to slip usually occur at high frequencies (>100 Hz) [1]. Many slip detection approaches, therefore, analyzed the relevant signals in terms of spectral features. Signal transforms such as Fast Fourier transform (FFT), short-term Fourier transform (STFT), and power spectral density (PSD) transform, are commonly used for spectral analysis. Holweg et al. [33] developed two approaches for slip detection using piezoresistive pressure sensors. For one of the two approaches, a 16×16 piezoresistive matrix was used to collect pressure data during slip events; a threshold for the PSD of the signal was characterized to distinguish slip events. Similarly, Fernandez et al. [34] used FFT and PSD transforms to detect micro-vibrations in the tangential force signal during slip; the contact forces were measured using a tactile sensor based on strain gauges.

Instead of manually searching for features in the frequency spectrum, some have attempted to learn such spectral features. For instance, Holweg et al. [33] compared their PSD-threshold method to a learning-based approach that used an artificial neural network and found the network-based approach to be significantly better. Another technique, by Meier et al. [35], utilized the STFT of the pressure signal from a piezoresistive tactile matrix to create frequency images, and used a CNN to learn to detect slip events; they achieved an accuracy of over 97% on a real-world slip detection task and showed that translational slip could be distinguished from rotational slip. The work of Goeger et al. [36] combined two tactile sensing modalities (piezoelectric

and conductive polymer resistance changes) to generate ‘tactile features’, which were extracted from the Fast Fourier transform (FFT) of the resulting signals; a trained k-nearest neighbour classifier served as a slip-state discriminator. Methods based on the analysis of the frequency spectrum of tactile signals can be sensitive to vibrations coming from the environment, such as those produced by nearby machinery; these vibrations have the potential to lead to the confounding of object-environment and object-gripper slip events. In [37], a dictionary of spectral tactile features was built to encode tactile readings and an SVM was trained to differentiate between object-environment and object-gripper slip.

Slip between contact surfaces induces vibrations which can be measured using a variety of transductions mechanism (e.g., piezoelectric or strain gauges). In certain situations, these vibrations generate obvious spectral features for which manually designed approaches tend to work. In other situations, the spectral features are less obvious and require the use of machine learning tools for extraction. The surveyed methods all point to the superiority of learning-based approaches for slip detection. Methods that use vibration analysis often fall short in their ability to generalize to different slip scenarios, since they are highly sensitive to roughness of the surface. A majority of sensors used for vibration analysis are custom-made and cannot guarantee signal/mechanical robustness without significant testing. This research addresses these concerns by training a DNN with a vast and diverse dataset that spans many aspects of slip.

3.2 Motion-Based Methods

Since slip is defined as the relative motion between the gripper and an object at the surface of contact (see Section 2.2 for a complete definition), the detection of motion at the surface can be associated with the detection of slip. One way to detect motion is to utilize a rubber-embedded accelerometer to capture small perturbations in the contact skin [38]. A recent method used a laser doppler velocimeter mounted on the fingertips to detect object motion [39].

Another popular approach to detect motion is to create a distributed tactile sensor with multiple force transducers stacked to form a matrix; by knowing the location of each transducer as well as the association between the transducer and the signal, the location of the center of (normal) force distribution (COF) can be computed.

Holweg et al. [33] used the pressure outputs of a 16×16 piezoelectric sensor matrix to calculate the COF. They found it difficult to detect events faster than 60 ms due to computational hardware limitations. Similarly, Zhang et al. [40] inserted an array of capacitive pressure sensors into the skin of a robot hand and estimated the center of (normal) pressure (COP); the power spectrum at the COP was used to distinguish object/world and object/hand slip.

The tactile sensors used for this thesis are based on barometers embedded in rubber to form a distributed matrix. The sensors are capable of transducing normal force into barometric pressure, which can be utilized to infer the COP during a contact event. The flexibility of the rubber membrane also permits deformation due to the tangential contact force, which is captured by the distributed pressure array.

3.3 Visualizing Deformation

In recent years, vision-based tactile sensors have advanced significantly, due in part to the boom in the fields of deep learning and computer vision. A basic vision-based tactile sensor requires a light source illuminating a photo-elastic material and a camera that can capture the material deformation, as shown in Figure 3.1. Vision is used to infer tactile information from the sensed deformation of the sensor contact surface. Some sensors like GelSight [14], FingerVision [41], and DIGIT [12], use an opaque contact layer, while others, like GelSlim [13] and TacTip [42], have visible patterns embedded in a transparent contact layer (translating deformation into pattern distortion). The methods that rely on visual patterns generally require fewer computational resources to infer tactile information.

Vision-based tactile sensors are commonly used for slip detection. For the GelSight [43] and GelSlim [44] tactile sensors, slip can be detected by monitoring changes to shear force at the contact surface. However, as noted in [45], vision-based sensors can struggle to detect slip in certain situations due to a limited camera frame rate.

In [46], the TacTip sensor is modified to use a camera that operates at up to 120 frames per second to better identify rapid slip movements. The authors of [47] extract deformation speed using optical flow and train a support-vector-machine (SVM) classifier for slip detection; the classifier performs with a high accuracy ($>95\%$) for a wide range of slipping speeds. However, like most other vision-based sensors, the TacTip is bulky, preventing it from being used as anything other than fingertips. In

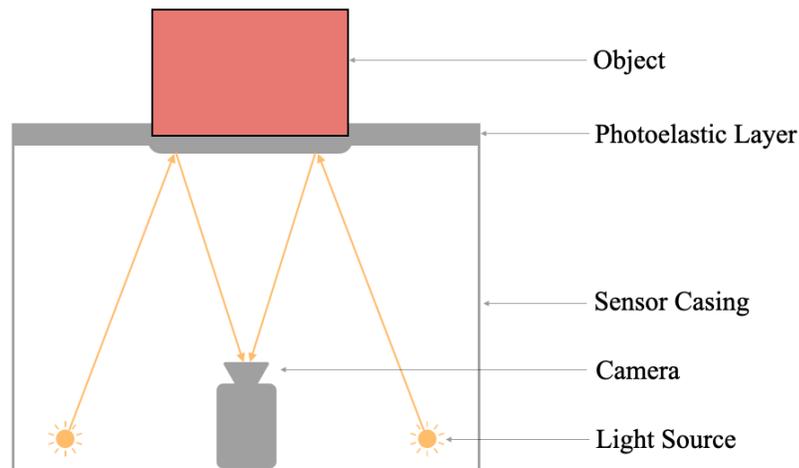


Figure 3.1: Components of a basic vision-based tactile sensor. The image shows deformation of the photoelastic layer due to contact with the object.

contrast, barometric sensors are compact and can be easily distributed on the palm of a hand [48].

There has been a significant improvement in slip detector performance with the introduction of learning-based methods. For example, James et al. [47] achieve a much higher detector performance when using an SVM classifier as compared to a heuristics-based method. Given the nature of the data collected by vision-based tactile sensors, CNNs are also commonly used with the visual data for slip detection [44, 45].

3.4 Learning-Based Methods

Due to the complex nature of tactile signals, data-driven approaches are increasingly being used to assess grasp stability and to detect slip. While machine-learning tools are used in many cases in conjunction with other methods, such as vibration analysis [33, 35, 36] and visuo-tactile inference [43–45, 47], some methods are capable of extracting slip information directly from the raw sensor inputs [49–52].

Classifiers including hidden Markov models, random forest models, and SVMs are often used to detect slip. Veiga et al. [52] use a random forest and an SVM classifier to infer slip, where the input comprises data from the commercially available Biotac tactile sensors [11]. Jamali et al. [49] use hidden Markov models to transform piezoelectric-sensor data into a sequence of symbols for slip prediction.

Deep learning is another powerful machine learning tool that has seen abundant development in the last decade. CNNs are the most commonly-used deep learning networks for slip detection. Methods that utilize a matrix of transducers as tactile sensors often encode the sensor data as tactile images, which are also compatible with CNNs. Methods like [35,51] are illustrative examples of this data transformation; the information from an array of pressure sensors is converted to pressure images.

Tactile signals also constitute time-series data that can be fed as inputs to RNNs (see Section 2.3.2). Zapata-Impata et al. [51] use the commercially available Biotac [11] to infer the direction of slip using a convolutional LSTM network. Van Wyk et al. [50] examine the use of three different commercially available tactile sensors and infer slip for each sensor using an LSTM network. The authors also analyze how variables such as temporal window size, sampling rate, object material, slip speed, and sensor calibration affect the performance of the slip detection LSTM network.

Recently-published experimental results in [4] suggest that networks based on generic convolution, such as TCNs, can outperform conventional RNNs in a diverse set of sequence modeling tasks. Generic convolution has also been successfully used for audio synthesis [3] and for inertial measurement processing [53]. This thesis explores the utility of TCNs for the complex task of slip detection.

Chapter 4

Learned Slip Detector

Slip between two surfaces is a complex spatiotemporal event that involves highly non-linear contact forces (see Section 2.2). Many of the heuristic-based techniques mentioned in Chapter 3 are limited by their ability to generalize. Given the complexity of slip, data-driven approaches offer greater potential to generalize.

For contact-level perception, we use recently-developed barometric tactile sensors that employ MEMS barometers as the transduction units. These sensors are mechanically robust, have high signal reliability, and are cost-effective. This work creates a learning-based slip detection method that operates reliably with barometric tactile sensors and generalizes well.

The spatial and temporal distribution of contact forces during in-hand object slip depends on the motion of the slipping object, the applied grasping force, and the properties of the object, such as roughness, deformability, material type, curvature, and elasticity, among others. To have any hope of generalization, the data used to teach the task of slip detection must span these aspects of slip.

As mentioned in section Section 2.3.2, TCNs show better performance on sequence modeling tasks than RNNs. They also enable parallel computation in the temporal domain, making them comparatively faster for inference. Consequently, slip detection in this thesis is achieved by extracting learned temporal features from raw barometer data using a TCN; these features are then used to classify a slip event.

This chapter explains how these aspects are identified and how a large dataset containing synthesized slip scenarios is collected. The hardware used and the tools created to collect and process the sensor data are also described. Finally, the chap-

ter reviews details of the TCN architecture, the training procedure, and parameter tuning, followed by the training results.

4.1 Tactile Sensors and Gripper



Figure 4.1: Robotiq three-finger adaptive gripper, retrofitted with the TakkTile sensor kit (image from [54]). The single finger on the right is referred to as the ‘thumb’ of the gripper.

We use an off-the-shelf barometric tactile sensor package (called the TakkTile kit and manufactured by Right Hand Labs¹). Each tactile sensor is a matrix of commercial MEMS barometers (NXP MPL115A2) assembled on a PCB and coated with urethane rubber, which creates a flexible medium to transduce contact forces into barometric pressure signals. This transduction setup was first created and tested by Tenzer et al. [19] who are also the co-founders of Right Hand Labs. The kit includes three fingertip mounts and a palm mount designed to be retrofit to the *Robotiq* three-finger gripper (Figure 4.1), such that the original fingertips and palm are replaced by the TakkTile sensors. Each sensing unit provides pressure and temperature data at a sampling rate of 100 Hz, which is higher than that of most vision-based tactile sensors. The sensors are compact, robust, and inexpensive relative to, for example, the

¹Purchased from <https://www.labs.righthandrobotics.com/robotiq-kit>



Figure 4.2: Power grasps of the Robotiq gripper (sourced from [54]). An object is held firmly between the fingers and the palm for the sphere grasp (left), whereas for the pinch grasp (right) the object is held solely by the fingertips while making no contact with the palm region.

competing BioTac sensors [11]. They exhibit greater linearity ($< 1\%$), no noticeable hysteresis, and a low signal-to-noise ratio (< 0.01 N).² Barometric tactile sensors are well-suited for industrial applications due to their high mechanical robustness and low cost.

The *Robotiq* three-finger gripper used for this study is capable of adapting to the size and shape of an in-hand object.³ The gripper is attached to a six degrees-of-freedom (DOF) UR10 robotic arm as an end-effector; it is possible to pick up and place objects with this setup. The gripper can execute two power grasps, shown in Figure 4.2, and for both grasping scenarios, the fingertips (especially the thumb) are always in contact with the in-hand object due to the adaptive nature of the gripper. We hypothesize that the barometric data solely from the fingertip units would be sufficient to infer in-hand object slip for the above two grasps, although the palm sensor might also be useful.

The design of the fingertip unit is straightforward; the layer of rubber is 10 mm

²Performance characterization of barometric tactile sensors is available in [19]

³An adaptive gripper is capable of passively adjusting the flexion of the fingers to envelope an in-hand object.

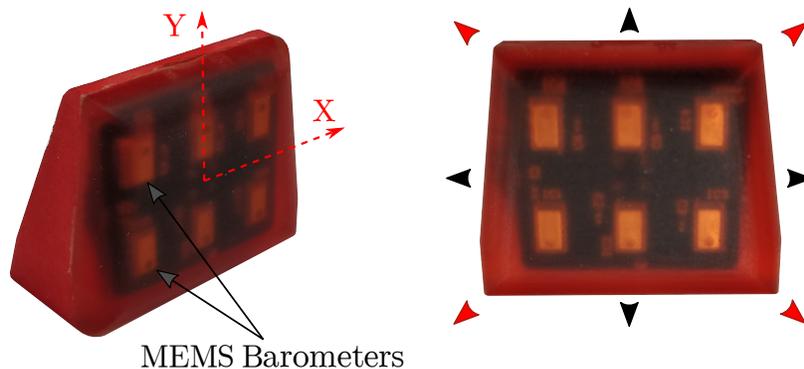


Figure 4.3: Perspective view (left) and front view (right) of the TakkTile fingertip. In the front view, black arrows represent the primary sensor axes and red arrows represent the oblique axes.

thick; six MEMS barometers form a 2×3 sensing array that spans the contact region of each fingertip, as shown in Figure 4.3. We define a planar coordinate system on the contact surface of the sensor, which is entirely flat: the origin lies at the center of the PCB, while the X and Y axes align with the length and width of the PCB, respectively (see Figure 4.3).

The TakkTile sensors and the gripper, along with the robotic arm, constitute the manipulation setup used in this study to collect a diverse slip dataset and to conduct the real-world experiments described in Chapter 5.

4.2 Data Acquisition

This work is motivated, in part, by a desire to learn a general model of slip, which requires collecting a large dataset spanning all facets of slip. To facilitate the data collection, the available manipulation setup is used to slide the tactile sensors across static surfaces, in order to emulate in-hand object slip. A slip data recorder is implemented using the Robot Operating System (ROS) framework. Furthermore, data collection is automated using the manipulation setup to facilitate diversity.

4.2.1 Slip Data Recorder

The tactile-sensing units (in Figure 4.3) used for this work were commercially manufactured by Right Hand Labs. Each kit comes with a signal processing integrated circuit board that is responsible for receiving and time-synchronizing the barometer

data for all sensing units. The manufacturers have also provided supporting software for their barometric tactile-sensing kit: a USB driver for the processing board and a ROS node to transmit synchronized barometer data at 100 Hz.⁴ Additional ROS nodes are also available for the UR10 arm and the Robotiq gripper, which provide control over the entire manipulation setup. It is possible to calculate the 3D velocity of the end-effector of the robot arm using proprioceptive encoder feedback and forward kinematics. This information can be readily accessed and transformed to calculate fingertip velocity along a plane during motion. The velocity feedback of the manipulation setup provides the necessary information to label our dataset (explained in Section 4.2). Ultimately, a custom data recorder is created as a ROS node to record the TakkTile sensor data along with the fingertip velocity.

The scope of this study is limited to the detection of slip using the barometric signal from a single fingertip tactile unit. To this end, the same TakkTile fingertip unit is used during data collection and for the robot experiments in Chapter 5. The reason for this choice is to eliminate the manufacturing variability of each unit as an influencing factor for the learned slip detector. The raw pressure values received from the sensing unit may have an offset from their originally-determined zero levels, possibly because of rubber deformation or temperature fluctuations. For this reason, the manufacturer of the sensing kit recommends using a calibration script, which is included in the supporting software as a ROS service. The calibration procedure includes both temperature compensation for each barometer as well as sensor zeroing. Once calibrated, the TakkTile node publishes raw and calibrated pressure values as well as the temperature measurements for each of the connected MEMS barometers; only the calibrated pressure data was recorded to remove the complexity introduced by temperature dependence and zero bias.

A learned slip detector is only as good as the quality of the recorded data, hence an intricate data recording system is created to collect high-quality measurements. Figure 4.4 shows a block diagram of the various components of the data recording system. The pressure data and the slip velocity data, transmitted by the hardware components, are published on their respective ROS topics at different frequencies. It is important that the two data streams be recorded in a time-synchronized manner in order to minimize the delay between the occurrence and the detection of slip. Each

⁴The USB driver is available at <https://github.com/TakkTile/TakkTile-usb> and the ROS node is available at https://github.com/harvardbiorobotics/takktile_ros

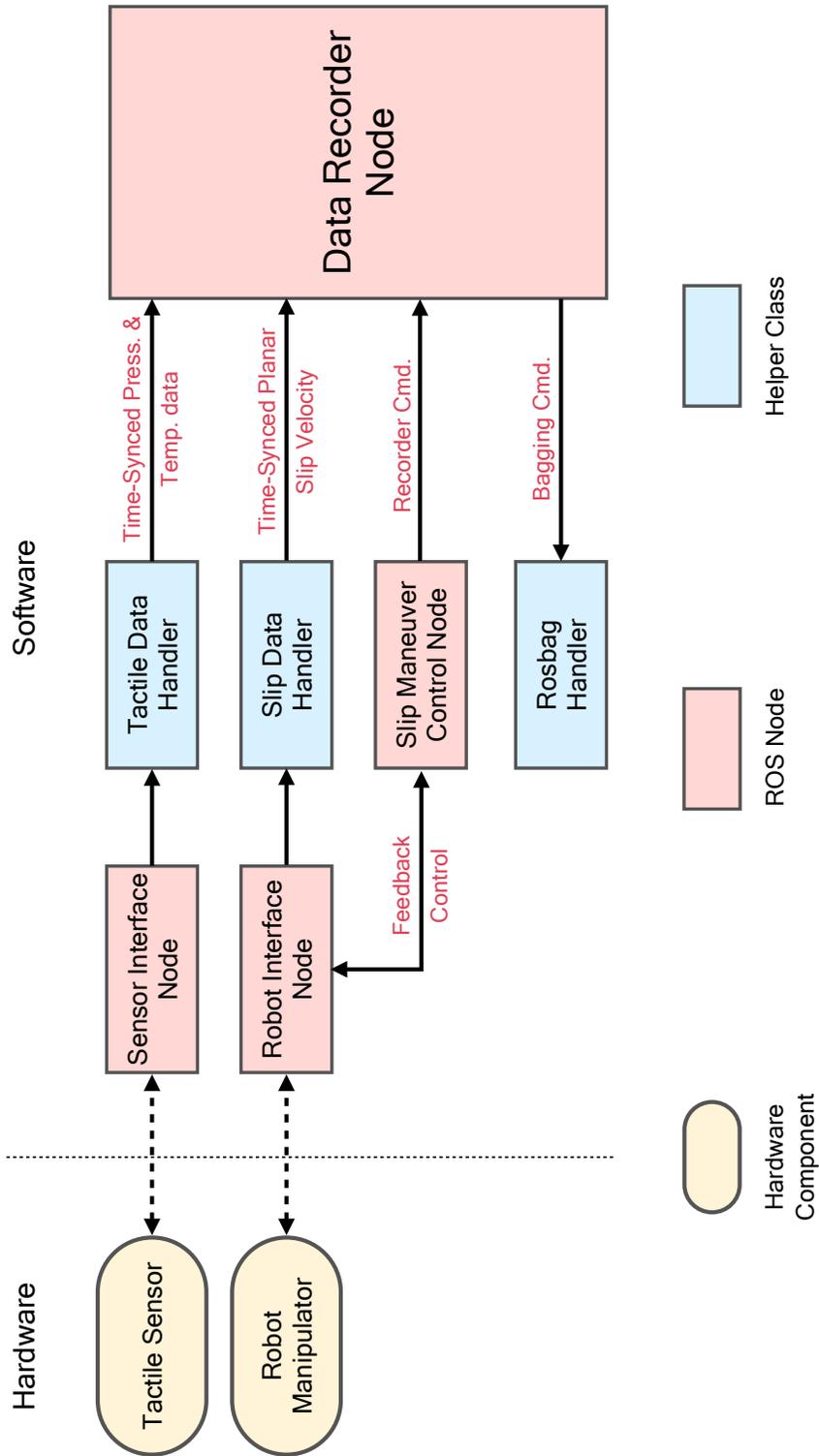


Figure 4.4: Functional flow block diagram of the data recording setup emphasizing the exchange of messages between the constituent components.

datapoint is recorded at the frequency of the TakkTile sensor (i.e., 100Hz), which is triggered every time new TakkTile data is available—at the same time, the last available end-effector velocity is recorded. This synchronization is handled by the ‘data handler’ classes (shown in Figure 4.4) that store the data of each sensor in local buffers. A main recording loop runs at a frequency higher than either sensor and polls the data handlers. While the data is being recorded, a ‘rosv bag handler’ class records the data communication during the entire collection trial as a ‘rosv bag’ (defined in [55]). The recorded rosv bags can be ‘played’ to simulate a collection trial without the need for actual hardware; rosv bags are also utilized to evaluate the performance of the detector on previously recorded data in real time. Another important component of the data recording setup is the ‘Slip Maneuver Node’ (Figure 4.4), which is responsible for generating a path for the tactile fingertip during the slip maneuvers mentioned in Section 4.2. The node also provides a user interface window to manually control all aspects of a slip maneuver.

By using the manipulation setup with the custom-made slip data recording system, it is possible to record a high-quality dataset with barometric pressure information as input and a velocity-derived slip label as output.

4.2.2 Data Collection and Distribution

A general slip detector must be trained on data that account for the many variables related to slip. Six such variables are identified: namely, object material properties (e.g., roughness and rigidity), object surface curvature, slip speed, slip direction, slip type (translational and rotational), and normal contact force. Differences in these variables stimulate the barometric units in different ways.

The huge variety of materials and surfaces in our world make it nearly impossible to cover all such object materials as part of the data collection effort. Instead, this research focuses on the latter five factors and a commonly-used material with a low coefficient of static friction: smooth rigid plastic. Even though the detector is trained with data collected on one material, the experiments in Chapter 5 have been conducted with objects that have many different material properties, and the promising results indicate material independence in many cases.

Instead of generating in-hand object slip events for training, it is easier to emulate slip scenarios by sliding the fingertip unit on a fixed, static surface. This sliding mo-

tion stimulates the tactile units in a way that is very similar to that when an in-hand object undergoes slip. The surface in contact with the fingertip is kept stationary, which makes it possible to calculate the relative velocity between contact surfaces simply based on the proprioceptive feedback of the motor encoders. In order to include both translational and rotational slip scenarios in the collected dataset, the UR10 arm is programmed to execute two types of maneuvers: an ‘asterisk’ maneuver, in which the fingertip moves along a straight line sequentially in eight different translation directions at a chosen speed; and a ‘pendulum’ maneuver, in which the gripper oscillates like a pendulum with the fingertip acting as the pivot point. These maneuvers are programmed to have periods with no motion in order to include examples without slip in the dataset. The translation slip data is collected at three speeds: 0.05 m/s, 0.075 m/s, and 0.1 m/s; whereas the rotational slip is collected at 1 rad/s only. The fingertip slides on ABS plastic surfaces with three different curvatures: spherical (radius of 1.5 in), cylindrical (radius of 2 in), and planar. The curved surfaces were 3D printed and the irregular edges were sanded off to obtain the smooth curves shown in Figure 4.5. For the planar surface, the lid of a smooth plastic box is used; the plastic box is also one of the objects used for the robot experiments in Chapter 5. Figure 4.6 shows a part of the asterisk sliding maneuver executed on the spherical ABS surface.

The available manipulation setup includes a wrist-mounted three axis force-torque (F/T) sensor (Robotiq FT300), visible in Figure 1.1. In the situation where the fingertip is the only part of the gripper in contact with the surface, the normal force measured by the F/T sensor is equal to the normal contact force at the fingertip, with the assumption that the weight of the gripper has already been accounted for. In order to consistently maintain a chosen normal contact force during a slip maneuver, a PID controller regulates the height of the fingertip based on the feedback signal of the F/T sensor. We attempted to collect data at specific levels of normal contact force (2.5 N, 5 N, 7.5 N), but due to the limited precision of the PID controller, the contact force varies by a small amount throughout the slip maneuvers, leading to a dataset with varying, yet bounded (≤ 10 N) normal contact force. This setup allows for automation of the data collection process and ensures that the ground-truth slip labels are accurate. A slip label is derived using the recorded planar velocity of the fingertip, such that translational and/or rotational velocities greater than 3 mm/s and 0.2 rad/s, respectively, are labelled as ‘slipping.’ These thresholds are determined

Table 4.1: The distribution of training data by slip type, slip speed, slip direction, and surface curvature.

	Max. Speed	Planar	Sph.	Cyl. (y -axis aligned)	Cyl. (x -axis aligned)	
Translation (primary axes)	5 cm/s	5.3%	3.7%	3.5%	3.6%	16.1%
	7.5 cm/s	4.5%	4.7%	3.7%	3.7%	16.7%
	10 cm/s	4.8%	4.7%	3.3%	3.2%	15.9%
Translation (oblique axes)	5 cm/s	4.9%	3.3%	3.3%	3.3%	14.8%
	7.5 cm/s	3.5%	4.2%	3.5%	3.4%	14.5%
	10 cm/s	3.9%	4.2%	3.0%	3.1%	14.3%
Rotation	1 rad/s	3.8%	1.2%	1.5%	1.2%	7.8%
Total		30.7%	26.1%	21.7%	21.5%	100%

empirically based on recordings of the TakkTile sensor under static conditions (i.e., resting on a surface without moving). The total data collected consists of over 45 minutes of TakkTile pressure samples at a rate of 100 Hz, with 143,584 data points belonging to the static class and 122,918 data points belonging to the slip class. The data in the slip class is evenly distributed across slip speed, slip direction, and surface curvature, and includes both translational and rotational motions, as shown in Table 4.1. To prevent abrasive damage to the rubber surface due to repetitive sliding, a layer of cellophane tape is applied over the fingertip during all collection trials and experiments.



Figure 4.5: Surfaces used for slip data acquisition; from left to right the images show a planar box lid, a spherical plastic surface (radius 1.5 in), and a cylindrical plastic surface (radius 2 in).

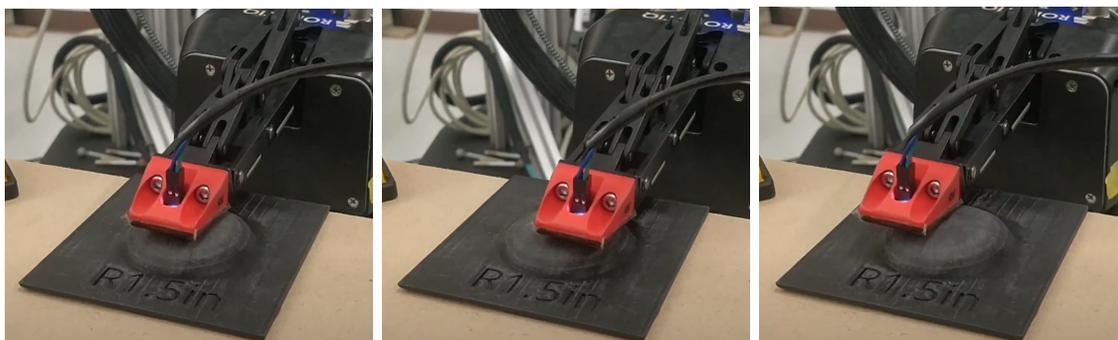


Figure 4.6: TakkTile sensor mounted on the fingertip of our Robotiq gripper [19] and actuated automatically to slide over the 3D printed spherical surface during training data collection.

4.3 Slip Detection Network

Prior research has successfully employed deep-learning models to extract spatiotemporal features from tactile sensor data for slip detection [35, 51]. Given the superior performance of TCNs over recurrent architectures for sequence modeling tasks [4], in this work a TCN-based architecture is used for slip detection. This section describes the architecture of the network and provides details regarding the training procedure. The section concludes with a discussion of the selection of the best TCN and showcases the detection performance of the chosen network.

4.3.1 Network Design

Since this is the first time a TCN has been used to learn a tactile perception task, there is no prior work to provide guidance as to the design of the network. We experimented with TCNs of various depths and with different numbers of trainable parameters in order to arrive at a network that achieves high accuracy with the smallest number of parameters. The goal is to have the shortest inference time (which is crudely indicated by its depth and number of parameters) without compromising the detection performance. As shown in Figure 4.7, the final network consists of four residual blocks (see Section 2.3.2), where each block contains two dilated causal convolution (DCC) layers and a residual connection. For each DCC layer, d represents the dilation and k represents the kernel size. The subsequent two layers are fully connected, followed by a softmax output for classification between slip and static. The constituent residual blocks are similar to the ones described in [4], except that layer normalization [56] is

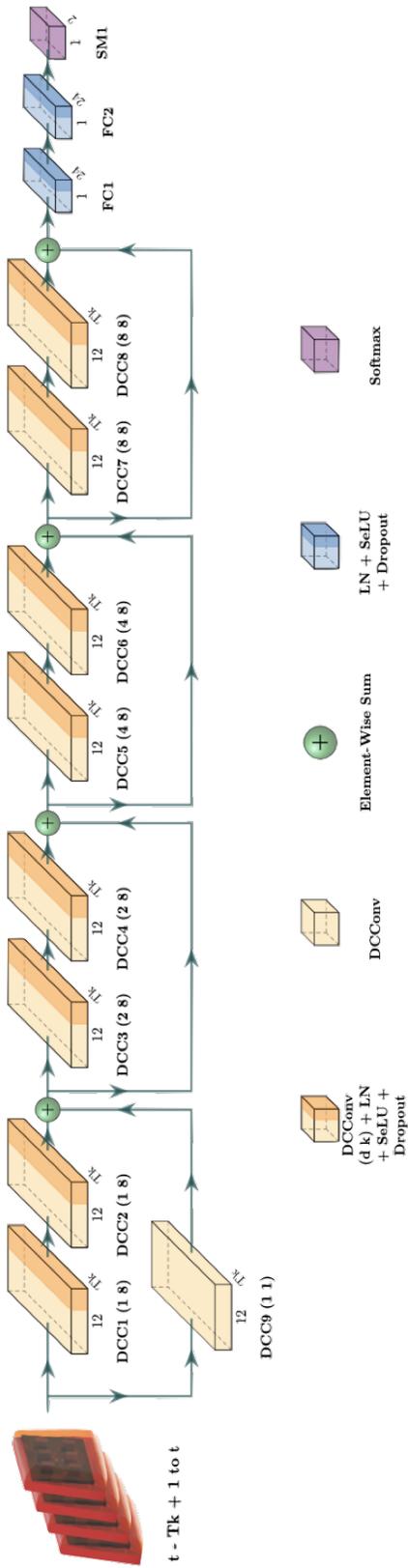


Figure 4.7: The TCN architecture used for slip detection.

used instead of weight normalization [57] and SELU activations [24] are used instead of the ReLU activations. These architectural alterations are motivated by experimental tests, which indicated better performance on the tactile dataset. The input to the TCN is a temporal stack of six barometer values that constitute the last T_k readings (choice of window size is explained in Section 4.3.4) of tactile data, and the output is a probability of slip. A binary classification probability threshold of 0.5 (the choice of threshold is explained in Section 4.3.4) is used to generate the binary slip and static labels. An open-source implementation of the TCN [58] is used to build this detection network.

4.3.2 Data Augmentation

The MEMS barometers form a 2×3 array on the fingertip module. To reduce overfitting, the axial symmetry of the array (visible in Figure 4.3) is exploited to augment the data. Before each training epoch, one of three transformations— an x -axis flip, a y -axis flip, or a 180° rotation—is randomly applied to every data point with a probability of 25% each. At the same time, small amount of random gaussian noise is added to the network inputs to ensure the network’s robustness to sensor noise. The result of data augmentation is an accuracy improvement of more than 10% that is primarily due to synthetic variations in the input.

4.3.3 Training Details

The network is trained using the Adam optimizer [25] with a learning rate of 0.002 and a binary cross-entropy loss function. Several techniques are used for network regularization: layer normalization; a 20% dropout (for each layer including the fully connected layers); a mini-batch of size 256 and *He normal* [59] kernel initialization. Moreover, before training, the class distribution of the entire training dataset is equalized through random under-sampling, in order to remove class bias. The dataset is split into training (80%), validation (10%), and test (10%) subsets, and the network is trained until the performance improvement plateaus, which happens after 800 epochs.

4.3.4 Parameter Tuning

Before the final slip detection network is chosen, two parameters related to the model require tuning: the input window size, which is the temporal length of the input data samples; and the binary classification threshold, which is the probability threshold that differentiates the slip and static predictions. Choosing these parameters is not always intuitive and since their values are problem-specific they are therefore determined experimentally.

In [50], a DNN is trained to detect slip, and experiments are performed to understand the effects of changing sensor parameters such as window size and sampling rate on the network’s performance. It is found that the performance of the network improves as the sampling rate and the temporal window size increases. A similar experiment focused on the window size is conducted for the proposed TCN where 10 different slip detection TCNs, with window sizes evenly distributed between 10 and 100, are trained on the entire dataset. These networks are trained for 400 epochs while following the training procedure mentioned in Section 4.3. Figure 4.8 shows a graph of the test performance of the 10 TCNs. The performance of the TCN improves with an increasing window size, similar to the results in [50]. Moreover since the convolution operation across a temporal dimension of a TCN layer is parallelized on the GPU, the inference time for each of the networks is identical on average. The results indicate that a large window size should be applied and hence a window size of 100 is used in all the subsequent experiments.

The choice of input window size affects the network structure, which means that a change in window size requires network retraining. On the other hand, it is possible to alter the classification probability threshold without such requirement. In order to optimize this threshold, a common approach is to use the precision-recall curve (PR curve), such that the two parameters are maximized together. Figure 4.9 shows the PR curve of the proposed TCN with an input window size of 100. The PR curve intersects with the red line at a value close to 0.5, which indicates that the precision and recall are balanced and that the classifier is unbiased towards the class type. Therefore, a probability threshold of 0.5 is used for binary classification with the proposed TCN.

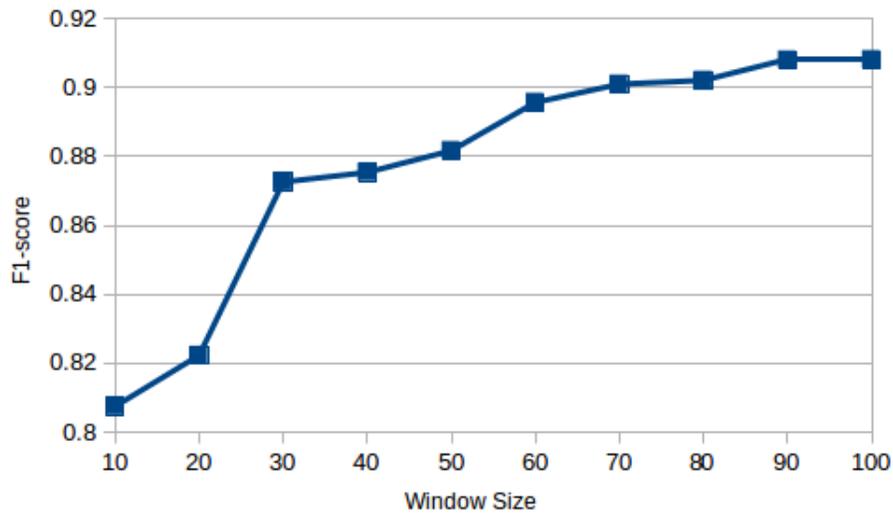


Figure 4.8: F1-score of slip detection TCN on the test data versus the temporal window size.

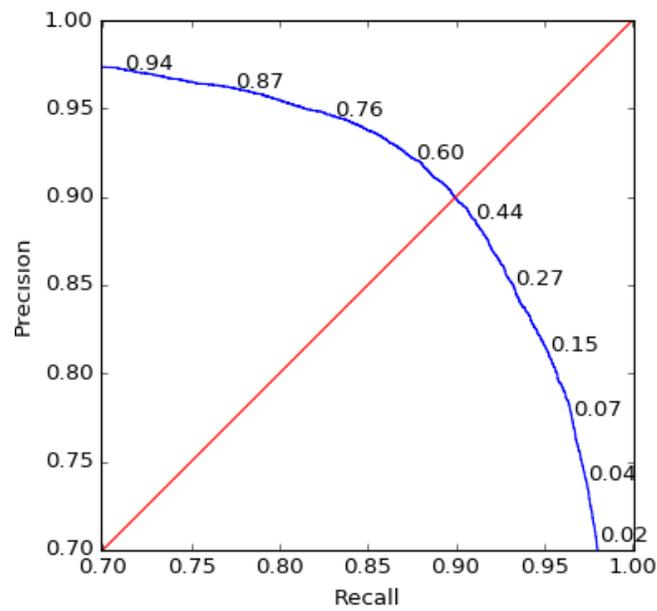


Figure 4.9: Precision versus Recall curve for changing binary classification threshold, calculated on the test data. The red line represents the function $f(x) = x$.

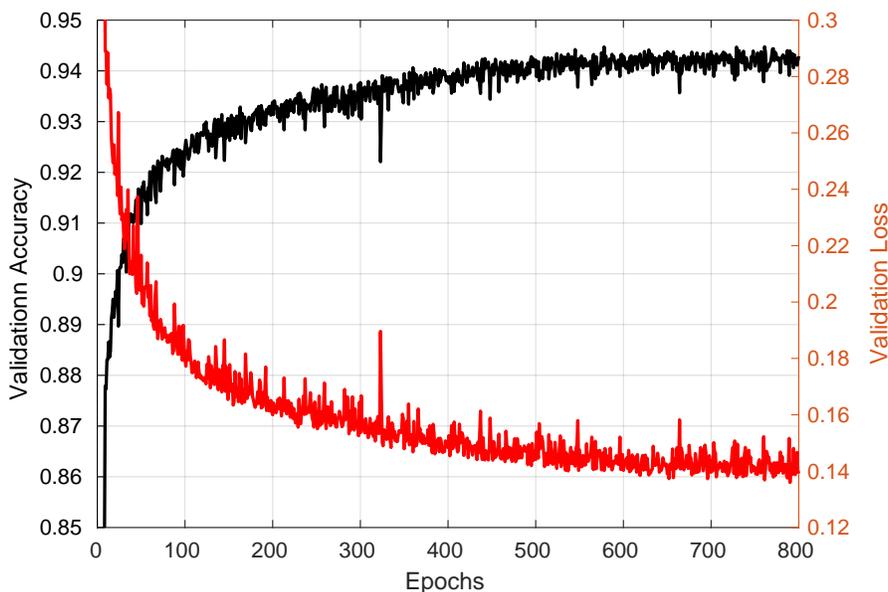


Figure 4.10: Validation accuracy and loss curve during training of the TCN for 800 epochs over the entire training set.

4.3.5 Training Result

Once the best parameters are determined, the network is trained and the best intermediate network is chosen based on the validation accuracy metric. Figure 4.10 shows the validation accuracy and loss curves throughout the training process. Classification accuracy of more than 90% is achieved within 50 epochs of training. Moreover, the accuracy curve increases consistently, implying that the network does not overfit to the training data.

The performance of the TCN on the complete test dataset is listed in Table 4.2. The performance metrics of precision and recall are almost identical for both classes, which is a consequence of data balancing before training. The detector performs better for static-label data; this is likely due to the larger number of training examples available for this class. Overall, the TCN-based slip detector demonstrates high accuracy ($> 90\%$) and good generalization (no over-fitting) with minimal-to-no class bias.

Table 4.2: Slip detection performance of the TCN on the test dataset.

	Contribution	Precision	Recall	F1-Score	Accuracy
Static Class	57.7%	92.3%	92.5%	92.4%	—
Slip Class	42.3%	90.1%	90.0%	90.0%	—
Overall	100%	91.4%	91.4%	91.4%	91.4%

Chapter 5

Experiments and Analysis

The proposed TCN has a classification performance of more than 90% on the test dataset for each metric (as described in Chapter 4). While there are many prior methods that use pressure-sensitive tactile matrices for slip detection (see Chapter 3), to showcase the advantage of our TCN-based approach, we compared the classification performance of our method with prior comparable approaches trained on an identical dataset. In addition to classification performance, there are other characteristics that determine the utility of a slip detector in the real world, such as sensitivity and latency. We analyzed the sensitivity of the slip detector to the variables of slip mentioned in Section 4.2. Additionally, we determined the influence of these variables on the real-time detection latency of the TCN. Finally, we carried out additional real-world experiments to evaluate the generalizability of the detector to different slip detection tasks.

5.1 Classification Performance Comparison

To demonstrate the effectiveness of the novel slip-detection method, the approach must be compared against prior techniques in the literature. We compared the performance of our proposed TCN with methods that utilize pressure information to detect slip. While slip detection using barometric tactile sensors has not previously been attempted, there is prior work that employs other pressure transduction technologies. Two methods that are comparable to our approach are: (1) [33] by Holweg et al., which uses a heuristic for classification; and (2) [35] by Meier et al., which utilizes deep learning for classification. For comparison purposes, we implemented

Table 5.1: Performance comparison between TCN and frequency-based methods on test data. The weighted average of the classes is used to compute each metric.

Method	Precision	Recall	F1-Score	Accuracy
PSD Thresh. [33]	57.9%	57.4%	57.5%	57.4%
Freq. CNN [35]	86.0%	86.0%	86.0%	86.0%
TCN (ours)	91.4%	91.4%	91.4%	91.4%

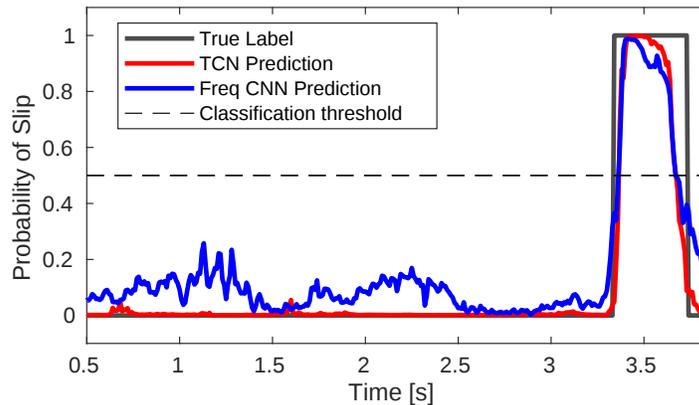


Figure 5.1: Slip probability response of two detectors on test data. The fingertip was executing a translational slip maneuver on a spherical surface.

our own versions of these methods and tested them on our dataset. For the approach in Holweg et al., we found the optimal PSD threshold by sweeping across various threshold values and evaluating the detector performance on the training dataset. For the latter, we built a CNN with three times as many parameters as the TCN, using the last one second of sensor data as input (identical to the method proposed in this work) to calculate the frequency components, and generated input images of size 2×3 . We used the Adam optimizer, batch normalization, 20% dropout, class balancing, and data augmentation while training on the entire training set. Table 5.1 shows a performance comparison of the above two methods with our TCN-based approach. While both of the learning approaches demonstrated promising results ($> 85\%$ accuracy), slip detection using the PSD threshold exhibited poor performance (as noted in [33] as well). The TCN outperformed the frequency CNN, with more than 5% improvement on each metric as shown in Table 5.1. Moreover, the TCN predicted the class label with significantly lower variance, according to the time-series plot shown in Figure 5.1. We believe the difference in the network performance can be attributed

to the fact that Meier et al. [35] enforce a specific transformation for the temporal features in the sensor data, making it harder to learn relevant features, whereas the TCN learns to extract temporal features from the raw sensor outputs.

5.2 Detection Performance Analysis

The slip detector proposed in this work is a small TCN with $\sim 10,000$ trainable parameters. The network was chosen with the intention to train a fast and accurate detector that can be used for slip compensation in manipulation scenarios. To gauge its utility for real-world tasks, we analyzed the sensitivity and detection latency of the proposed slip detector.

5.2.1 Sensitivity Analysis

The primary objective of this work is to create a highly accurate slip detector for barometric tactile sensors that can generalize to many forms of slip. This detection accuracy may change based on many variables, that is, the detector may be sensitive to some while being independent of others. To characterize this sensitivity, the classification performance of the detector was evaluated for various combinations of slip type, slip speed, slip direction, and surface curvature. The proposed TCN was trained on the entire dataset and tested on portions of the test data divided based on the slip variables. Table 5.2 summarizes the results of this experiment, where the F1-score is used as a metric for comparison, according to which the classification performance of the detector was sensitive to all of the variables of slip.¹

The detector performed significantly better on planar surfaces than for other surface curvatures; it also performed better on cylindrical curvatures when the longitudinal axis of the cylinder was aligned with the x -axis of the fingertip, rather than the y -axis. This difference may occur because the TakkTile fingertip only has two rows of MEMS barometers in the y direction but three rows in the x direction. With the cylindrical axis aligned with the x -axis of the sensor, the number of barometers stimulated during a slip event were more than the other cylindrical case. When the sensor slipped on a planar surface, all of the barometers were activated, which was not the case for other surface curvatures. Thus, there was a correlation between the

¹F1-score is the harmonic mean of precision and recall of a classifier for a given class.

Table 5.2: TCN performance, measured by F1-score, with variation in slip type, slip speed, slip direction, and surface curvature. The table also includes classification results for cases independent of slip and motion type.

	Max. Speed	Planar	Sph.	Cyl. (<i>y</i> -axis aligned)	Cyl. (<i>x</i> -axis aligned)	Surface Ind.
Translation (primary axes)	5 cm/s	92.5%	88.8%	78.7%	94.0%	89.7%
	7.5 cm/s	94.9%	88.0%	90.0%	90.3%	91.6%
	10 cm/s	95.7%	84.5%	87.8%	92.1%	93.1%
Translation (oblique axes)	5 cm/s	94.0%	88.9%	87.1%	93.2%	91.5%
	7.5 cm/s	96.6%	89.3%	93.7%	92.2%	93.1%
	10 cm/s	96.4%	88.6%	93.5%	93.4%	93.8%
Rotation	1 rad/s	84.6%	74.5%	84.0%	86.1%	81.5%
Motion Ind.	—	94.0%	88.9%	88.0%	91.9%	91.4%

number of activated barometers and slip-detection accuracy of the TCN.

In Table 5.2, for almost all curvatures and slip directions, the classification performance improves with increasing slip speed; higher speeds induce larger surface deformations, leading to more prominent temporal features within the input window. Performance was also affected by slip direction: slip along the oblique axes of the sensor was better-detected than along the primary axes, although the difference was not significant for most slip speeds and curvatures. The detector yielded poor results (<80% F1-score) for rotational slip, with exceptionally bad performance on spherical surfaces. A scarcity of training data containing rotational slip examples could be the reason for this discrepancy, while a lower number of stimulated barometers during rotational slip may explain the poor performance on spherical surfaces.

5.2.2 Latency Analysis

The time difference between the occurrence and the detection of an event is referred to as the detection latency. Despite a lower detection latency being essential for the functionality of a slip detector, much of the previous slip-detection research has failed to mention latency, typically characterizing the inference time of their detectors only (i.e., the time a detector takes to produce an output). For example, in [47], the authors only mention the sub-millisecond inference time of their slip detector, but

do not comment on its real-time detection latency. Detection latency is, however, a better indicator than inference time of the expected slip response time during a robotic manipulation task. This section provides an overview of our methodology for characterizing the detection latency of the proposed slip detector along with the results.

The slip data recorder, discussed in Section 4.2, includes the functionality to record rosbags [55] during data collection. With a rosbag, it is possible to play the recorded barometric data and simulate a TakkTile unit running in real-time. To characterize the latency, the slip detector was run on the rosbags of test dataset recordings. The latency was calculated by measuring the time difference between the static-to-slip transition in a recording and in the detector outputs. For example, in Figure 5.1, the time difference between the peaks of the true label curve and the TCN curve is the detection latency for that recording. Latency was measured over all test data recordings and calculated for various combinations of slip variables. Table 5.3 shows the results of this experiment. For the case of translational slip, the results agree with the intuition that slip was detected earlier if the relative motion was faster—most likely because the temporal features related to slip appear earlier in the data stream. The detection latency appears to follow an identical trend to that of classification performance, seen in Section 5.2.1, when comparing for different combinations of slip variable. On average, the proposed method had a network inference time of 21 ms and a detection latency of 134 ms, which is comparable to the 100 ms latency of human response to slip events [21].

Table 5.3: TCN detection latency (in seconds) with variation in slip type, slip speed, and surface curvature. The latency is shown for spherical, cylindrical, and planar surface curvatures. The table also includes surface- and motion-independent latency.

	Max. Speed	Planar	Sph.	Cyl. (y -axis aligned)	Cyl. (x -axis aligned)	Surface Ind.
Translation	5 cm/s	0.126	0.161	0.187	0.149	0.154
	7.5 cm/s	0.123	0.134	0.150	0.126	0.134
	10 cm/s	0.054	0.118	0.145	0.111	0.106
Rotation	1 rad/s	0.049	0.169	0.260	0.099	0.158
Motion Ind.	—	0.093	0.140	0.169	0.125	0.134

Table 5.4: Surface properties of test objects

	Smooth	Rough
Deformable	Foam sleeve (cylindrical)	Football sleeve (spherical)
Rigid	Plastic ball (spherical), Cardboard can (cylindrical), Plastic box (planar)	Metal can (cylindrical)

5.3 Robot Experiments

The previous sections established a performance baseline for the slip-detection TCN on pre-recorded emulated slip data. To gauge the utility of the method, it must be tested on real-world scenarios involving robot manipulation. In this section, we studied the performance of our proposed slip-detection method on two robotic manipulation tasks and evaluate its ability to generalize.

The slippage of an object out of a robot hand is either triggered by an external force or by an error in the control. To be able to reliably repeat a slip-detection experiment on a robot, slip was artificially induced. We designed two manipulation tasks to test the ability of our detector: (1) slip detection for an in-hand object under an externally-applied impulsive force; and (2) slip detection while lifting an object with insufficient grasping force. The latter task was inspired by the work of James et al. [46], where slip was detected and compensated for while lifting. The test objects for the experiments, shown in Figure 5.2, were selected with the intention to vary the properties of the contact surfaces, such as curvature, roughness, and deformability. Table 5.4 shows the distribution of surface properties for each selected object. Only the fingertip unit that was used for data collection was applied for these tests to avoid the effects of manufacturing variabilities. The robot experiments demonstrate an ability to generalize from single-material training to real-world, multi-material slip detection involving different objects.

5.3.1 Mallet Tap Test

The goal of this experiment was to evaluate the performance of the slip detector for in-hand objects held with a constant grasping force. For the test, we programmed the



Figure 5.2: Test objects used for robot experiments. From left to right, the objects are: *plastic box* , *plastic ball* , *football sleeve* , *foam sleeve* , *metal can* , and *cardboard can* .

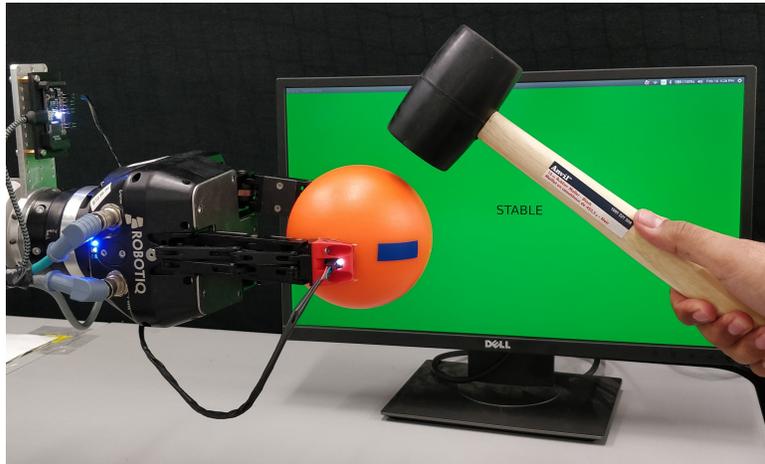


Figure 5.3: Mallet-tap test setup with plastic ball.

robot arm and hand (fitted with the TakkTile fingertip unit at the thumb location) to grab an object with sufficient force to lift the object above the test table. The object was then manually tapped using a 16-ounce rubber mallet with enough force to induce slip without causing grasp failure. The setup for this test is shown in Figure 5.3; the LCD monitor in the background displayed slip events in real time. A slip state was registered only if the detection network produced a slip label for two consecutive outputs. A trial was deemed successful if slip was registered when the object was tapped and if the detector output returned to the nominal value (static) when the object stopped moving. Figure 5.4 contains two series of images showcasing the mallet-tap test for the plastic ball. In sequence, the figure shows images and timestamps of the object being tapped; slip being detected by the system; and the detector output returning to stable. The detection latency in Figure 5.4 was approximately 200 milliseconds, which is higher than the average detection latency of 134 milliseconds in Section 5.2.2. This was because the experimental latency included the time it took to capture the second consecutive output from the detector, which increased the detection latency for this experiment.

We conducted 20 trials per object, with an equal number of taps on the top and on the side of the object; the results are shown in Table 5.5. In this experiment, the slip detector achieved a success rate of 80% or greater for each test object, with an average success rate of 87.5%. Slip detection for the foam sleeve, a very challenging object, had the worst performance, which was likely due to the smooth and

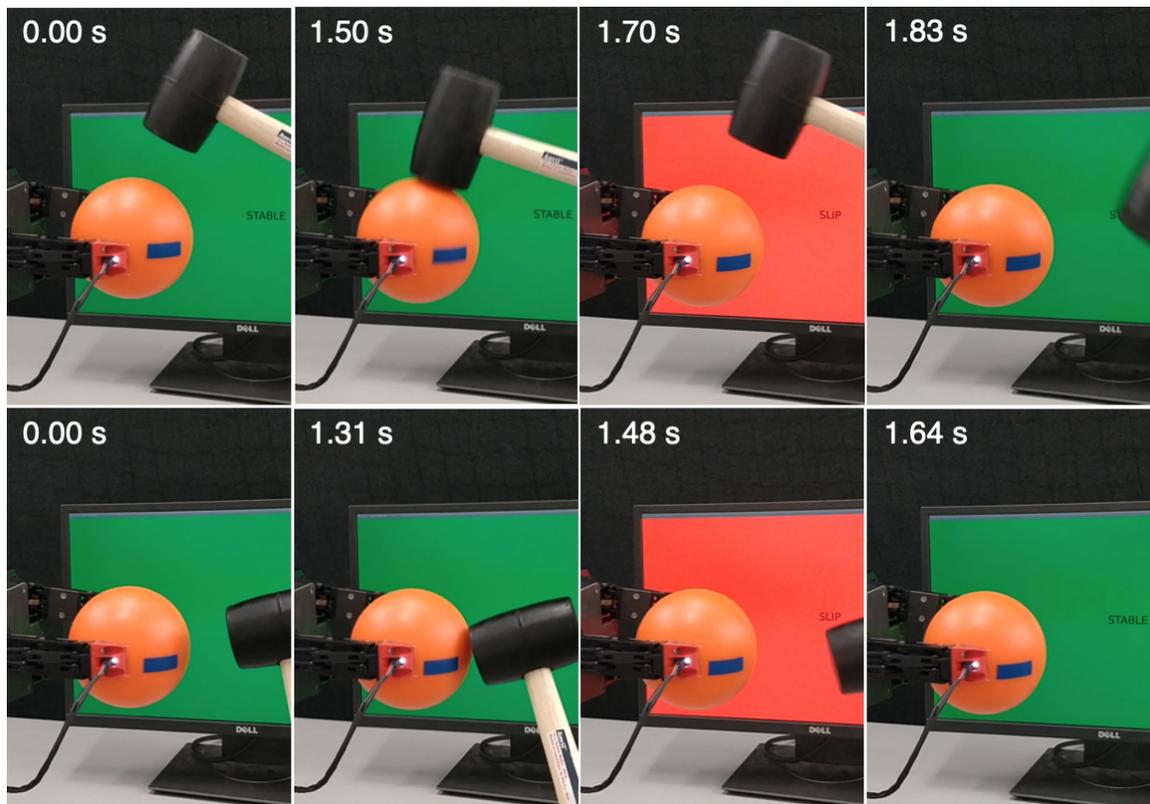


Figure 5.4: Mallet-tap test for one trial where the plastic ball was tapped from above (top) and from the side (bottom). The time stamps were extracted from the video and may not be accurate.

Table 5.5: Slip-detection results for two real-world manipulation experiments.

Object	Mallet-Tap	Object-Lift
Plastic Ball	90%	35%
Plastic Box	85%	100% ³
Cardboard Can	90%	85% (45% ³ + 40% ⁴)
Football Sleeve	90%	85%
Foam Sleeve	80%	35%
Metal Can	90%	95% (75% ³ + 20% ⁴)
Avg. Success	87.5%	72.5%

³Final grasp within 6 cm of the initial grasp position.

⁴Final grasp between 6-12 cm of the initial grasp position.

deformable surface of the sleeve. Failure modes for this experiment included a nearly equal number of false negatives and false positives, implying that the TCN did not exhibit any significant detection bias. The slip detector performed well under constant load conditions despite variations in the material, curvature, deformability, and smoothness of the object surfaces. Since contact was maintained continuously, this test detected the true object slip, rather than simply detecting a uniform reduction in force across the tactile sensor array (i.e., a sudden ‘air gap’ upon grasping force reduction). This experiment showcased the capability of the slip detector to perform in basic manipulation scenarios where external disturbances induce slip.

5.3.2 Object Lift Test

Lifting an object is a complicated task that involves precise hand-eye coordination and continuous grip adjustment to maintain a stable grasp. When lifting an object, humans intuitively change their grasping force based on the weight of the object. If a mismatch exists between the perceived and the actual weight, slip will occur due to insufficient grasping force. As an experiment, we replicated this scenario with our gripper and arm, using the same set of objects (shown in Figure 5.2). Since the gripper is not built for fine finger control, it was challenging to find gripper configurations that maintain contact with the object but generate insufficient grasping force. To solve this problem, we added weights to the lighter objects (i.e., the cardboard can, the football sleeve, and the foam sleeve) to increase the likelihood of slip. Figure 5.5 shows the setup used for this experiment. For the purpose of repeatability, we used



Figure 5.5: Object-lift test setup.



Figure 5.6: Object-lift test with the cardboard can, showcasing a successful trial. The time stamps were extracted from the video and may not be accurate.

a predetermined gripper configuration for each object and kept the initial grasping position consistent throughout the experiments. The blue tape on the objects marked the initial grasp location; additional markings were added to the longer objects (i.e., the cardboard can, the metal can, and the plastic box) for reference. Once the gripper fingers reached a predetermined position, the slip detector was initialized—this was followed by a lifting motion. If slip was detected for two consecutive time steps, a slip event was registered and the gripper was programmed to tighten its grasp immediately in order to compensate. As the normal contact force remained low throughout the lifting phase (despite the added weights), this task was particularly challenging for the slip detector.



Figure 5.7: Object-lift test with the foam sleeve, showcasing a failure trial. The time stamps were extracted from the video and may not be accurate.

An experimental trial was considered successful if slip was correctly detected and compensated for, that is, if the object was stably grasped and lifted up during the course of the hand motion. For the longer objects (namely the cardboard can, the plastic box, and the metal can), successful trials were split on the basis of the distance between the initial and the final grasp position—distance thresholds of 6 cm and 12 cm were used to define these two categories. Figure 5.6 shows one trial with snapshots displayed in succession. The images after the first depict (from left to right): the gripper closing; the slip while lifting the object; the slip detection with grasp compensation; and the stable grasp (static) detection. Compared to the mallet tap test, the detection latency for this test (which was ~ 400 ms) was higher due to the limited normal contact force throughout the slip motion, which lead to low barometer stimulation and low detection accuracy. We conducted 20 trials for each object; the experimental results are shown in Table 5.5. As a general trend, the success rate was greater for longer objects, which was the result of a longer reaction time: there was more time to ‘catch’ longer objects. Similarly, the tests with the football sleeve were very successful, which we attribute to its longer body and rough surface texture. Two of the objects had prominent failure modes: (1) tests with the plastic ball failed due to the fast reaction time required; and (2) tests with the foam sleeve frequently resulted in false negatives. Given the curvature of the plastic ball, it could only be stably grasp at its center—after a small amount of slip over a brief interval, the ball was off-centre relative to the gripper fingers, making a slower reaction time the dominant reason for failure. The smooth and compressible surface of the foam sleeve made it

especially difficult to detect slip, leading to a large number of false negative results. Figure 5.7 displays a trial with the foam sleeve where the detector fails to detect slip. We note, however, that deformable objects are generally very difficult to manipulate and that our detector was not trained on the collected slip data from any of these objects. Overall, this test was an indicator of the applicability of a slip detector on a more difficult manipulation task—for example, when lifting a fragile object with low gripping force, the object may slip out of the gripper. The test also shed light on the generalization capability of the proposed method as well as its limitations.

Chapter 6

Conclusion

The human sense of touch is complex and difficult to replicate artificially. Slip detection, one of the many tasks tactile sensing is applied to, has proven to be a challenge to reproduce in robots, despite the plethora of research on this topic. Prior work includes methods that solve slip detection by using one of several transduction mechanisms, or by focusing on specific parameters related to slip. In this work, we presented a learning-based method for slip detection using inexpensive barometric tactile sensors, addressing many of the shortcomings of prior work and attempting to rejuvenate interest in barometric tactile-sensing technology.

6.1 Summary and Contributions

We used a temporal convolution network [4] that learned to detect slip events by extracting spatiotemporal features directly from the tactile data. To ensure that the learned model was able to generalize, it was trained on a diverse dataset that included variations in surface curvature, slip speed, slip direction, and slip type. The accuracy and performance of our slip detection algorithm was demonstrated by comparing its performance with an existing classical method and a learning-based method. We also assessed the sensitivity and robustness of our method to the same variations, and characterized its real-time detection latency. Finally, we evaluated the performance of the proposed slip detector on two real-world robotic manipulation tasks, using objects with different surface properties such as curvature, roughness, and deformability. Compared to [33,35], our method achieved the best performance, with an accuracy of greater than 91% on a diverse dataset. The detector displayed

high sensitivity to slip type and surface curvature, while being relatively robust to slip speed and direction. On average, the proposed detector was able to detect a slip event within 134 milliseconds of its occurrence. We also demonstrated that our learned model is transferrable to real-world tasks and to different materials without retraining.

6.2 Potential Improvements

Although better classification performance can sometime be achieved using other tactile sensors, barometric tactile sensors offer a unique combination of reliability, mechanical robustness, and price point, making them suitable for many industrial applications. The slip-detection performance of barometric tactile sensors would be further improved by increasing the spatial density of the MEMS barometers and improving measurement fidelity. Similarly, a faster sampling rate for the tactile sensors would greatly improve the latency of the detector. The detection latency would also be greatly improved through targeted software and hardware optimization—we did not aim to optimize detection latency in this work.

The training data collected for this work included variations in the type of slip, that is, translational and rotational slip. However, a scarcity of data containing the latter slip type limited the current investigation. A similar limitation was due to the inability to precisely control the normal contact force of the tactile sensor during data collection. A revised data collection setup, where the UR10 arm is replaced by a torque-controlled robotic arm with higher-resolution motor encoders, would improve the quality of training data and the performance of the slip detector.

6.3 Future Work

As a continuation of this work, it would be useful to investigate the benefits of employing multiple TakkTile fingertips and a palm sensor for slip detection during manipulation. Similarly, the ability to estimate the type (translational or rotational) and direction of object slip would be valuable. The use of newly-introduced transformer networks [60] as a model for slip detection would be interesting to evaluate, especially with regards to the trade-off between classification accuracy and inference speed. Also, it would be interesting to explore control strategies for dexterous manip-

ulation tasks that rely on controlled slip. Finally, we would like to examine the use of different barometric sensor configurations for tactile sensing, including as flexible ‘skin,’ similar to [48].

Bibliography

- [1] R. A. Romeo and L. Zollo, “Methods and sensors for slip detection in robotics: a survey,” *IEEE Access*, April 2020.
- [2] G. Straffelini, *Friction and Wear: Methodologies for Design and Control*, pp. 21–60. Cham: Springer International Publishing, 2015.
- [3] A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, “Wavenet: A generative model for raw audio,” in *ISCA Speech Synthesis Workshop*, pp. 125–125, 2016.
- [4] S. Bai, J. Z. Kolter, and V. Koltun, “An empirical evaluation of generic convolutional and recurrent networks for sequence modeling,” *arXiv preprint arXiv:1803.01271*, 2018.
- [5] G. Westling and R. Johansson, “Factors influencing the force control during precision grip,” *Experimental Brain Research*, vol. 53, no. 2, pp. 277–284, 1984.
- [6] R. S. Johansson and G. Westling, “Roles of glabrous skin receptors and sensorimotor memory in automatic control of precision grip when lifting rougher or more slippery objects,” *Experimental Brain Research*, vol. 56, no. 3, pp. 550–564, 1984.
- [7] C. Wang, S. Wang, B. Romero, F. Veiga, and E. Adelson, “Swingbot: Learning physical features from in-hand tactile exploration for dynamic swing-up manipulation,” in *IEEE/RSJ Intl. Conf. Intelligent Robots and Systems (IROS)*, pp. 5633–5640, 2020.
- [8] V. Ortenzi, A. Cosgun, T. Pardi, W. Chan, E. Croft, and D. Kulić, “Object handovers: a review for robotics,” *arXiv preprint arXiv:2007.12952*, 2020.

- [9] W. P. Chan, C. A. Parker, H. M. Van der Loos, and E. A. Croft, “Grip forces and load forces in handovers: implications for designing human-robot handover controllers,” in *Proc. ACM/IEEE Intl. Conf. Human-Robot Interaction*, pp. 9–16, 2012.
- [10] A. Maslyczyk, J.-P. Roberge, V. Duchaine, *et al.*, “A highly sensitive multimodal capacitive tactile sensor,” in *IEEE Intl. Conf. Robotics and Automation (ICRA)*, pp. 407–412, 2017.
- [11] R. S. Johansson, G. E. Loeb, N. Wettels, D. Popovic, and V. J. Santos, “Biomimetic tactile sensor for control of grip,” Feb 2011. US Patent 7,878,075.
- [12] 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 Robot. Autom. Lett.*, vol. 5, no. 3, pp. 3838–3845, 2020.
- [13] E. Donlon, S. Dong, M. Liu, J. Li, E. Adelson, and A. Rodriguez, “Gelslim: A high-resolution, compact, robust, and calibrated tactile-sensing finger,” in *IEEE/RSJ Intl. Conf. Intelligent Robots and Systems (IROS)*, pp. 1927–1934, 2018.
- [14] M. K. Johnson, F. Cole, A. Raj, and E. H. Adelson, “Microgeometry capture using an elastomeric sensor,” *ACM Trans. Graphics (TOG)*, vol. 30, no. 4, pp. 1–8, 2011.
- [15] C. Chorley, C. Melhuish, T. Pipe, and J. Rossiter, “Development of a tactile sensor based on biologically inspired edge encoding,” in *2009 Intl. Conf. on Advanced Robotics*, pp. 1–6, 2009.
- [16] R. S. Dahiya, G. Metta, M. Valle, and G. Sandini, “Tactile sensing—from humans to humanoids,” *IEEE Trans. Robotics*, vol. 26, pp. 1–20, February 2010.
- [17] U. Proske and S. Gandevia, *Kinesthetic Senses*, vol. 8, pp. 1157–1183. 07 2018.
- [18] R. Dahiya and M. Valle, *Robotic Tactile Sensing – Technologies and System*, pp. 79–129. Springer International Publishing, January 2011.

- [19] Y. Tenzer, L. P. Jentoft, and R. D. Howe, “The feel of MEMS barometers: Inexpensive and easily customized tactile array sensors,” *IEEE Robot. Autom. Mag.*, vol. 21, no. 3, pp. 89–95, 2014.
- [20] M. T. Francomano, D. Accoto, and E. Guglielmelli, “Artificial sense of slip—a review,” *IEEE Sensors Journal*, vol. 13, no. 7, pp. 2489–2498, 2013.
- [21] 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.
- [22] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
- [23] L. Lu, “Dying relu and initialization: Theory and numerical examples,” *Communications in Computational Physics*, vol. 28, p. 1671–1706, Jun 2020.
- [24] A. M. Günter Klambauer, Thomas Unterthiner and S. Hochreiter, “Self-normalizing neural networks,” in *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 972–981, 2017.
- [25] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.
- [26] R. S. Fearing and J. M. Hollerbach, “Basic solid mechanics for tactile sensing,” in *IEEE Intl. Conf. Robotics and Automation (ICRA)*, vol. 1, (Atlanta, Georgia, USA), pp. 266–275, March 1984.
- [27] M. H. Raibert, “An all digital VLSI tactile array sensor,” in *IEEE Intl. Conf. Robotics and Automation (ICRA)*, vol. 1, (Atlanta, Georgia, USA), pp. 314–319, March 1984.
- [28] M. R. Cutkosky and W. Provancher, “Force and tactile sensing,” in *Springer Handbook of Robotics*, pp. 717–736, Springer, 2016.
- [29] L. L. Salisbury and A. B. Colman, “A mechanical hand with automatic proportional control of prehension,” *Medical and biological engineering*, vol. 5, no. 5, pp. 505–511, 1967.

- [30] J. C. Baits, R. W. Todd, and J. M. Nightingale, “The feasibility of an adaptive control scheme for artificial prehension,” *Proceedings of the Institution of Mechanical Engineers, Conference Proceedings*, vol. 183, no. 10, pp. 54–59, 1968.
- [31] A. Mingrino, A. Bucci, R. Magni, and P. Dario, “Slippage control in hand prostheses by sensing grasping forces and sliding motion,” in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS’94)*, vol. 3, pp. 1803–1809 vol.3, 1994.
- [32] D. P. J. Cotton, P. H. Chappell, A. Cranny, N. M. White, and S. P. Beeby, “A novel thick-film piezoelectric slip sensor for a prosthetic hand,” *IEEE Sensors Journal*, vol. 7, no. 5, pp. 752–761, 2007.
- [33] E. Holweg, H. Hoeve, W. Jongkind, L. Marconi, C. Melchiorri, and C. Bonivento, “Slip detection by tactile sensors: algorithms and experimental results,” in *IEEE Intl. Conf. Robotics and Automation (ICRA)*, vol. 4, pp. 3234–3239, 1996.
- [34] R. Fernandez, I. Payo, A. S. Vazquez, and J. Becedas, “Micro-vibration-based slip detection in tactile force sensors,” *Sensors*, vol. 14, no. 1, pp. 709–730, 2014.
- [35] M. Meier, F. Patzelt, R. Haschke, and H. J. Ritter, “Tactile convolutional networks for online slip and rotation detection,” in *Intl. Conf. Artificial Neural Networks (ICANN)*, pp. 12–19, 2016.
- [36] D. Goeger, N. Ecker, and H. Woern, “Tactile sensor and algorithm to detect slip in robot grasping processes,” in *IEEE Intl. Conf. Robotics and Biomimetics*, pp. 1480–1485, 2009.
- [37] J.-P. Roberge, S. Rispal, T. Wong, and V. Duchaine, “Unsupervised feature learning for classifying dynamic tactile events using sparse coding,” in *IEEE Intl. Conf. Robotics and Automation (ICRA)*, pp. 2675–2681, 2016.
- [38] M. R. Tremblay, W. Packard, and M. R. Cutkosky, “Utilizing sensed incipient slip signals for grasp force control,” tech. rep., Stanford Univ. CA, Center for Design Research, 1992.
- [39] N. Morita, H. Nogami, Y. Hayashida, E. Higurashi, T. Ito, and R. Sawada, “Development of a miniaturized laser doppler velocimeter for use as a slip sensor

- for robot hand control,” in *2015 28th IEEE International Conference on Micro Electro Mechanical Systems (MEMS)*, pp. 748–751, 2015.
- [40] X. Zhang and R. Liu, “Slip detection by array-type pressure sensor for a grasp task,” in *2012 IEEE International Conference on Mechatronics and Automation*, pp. 2198–2202, 2012.
- [41] A. Yamaguchi and C. G. Atkeson, “Combining finger vision and optical tactile sensing: Reducing and handling errors while cutting vegetables,” in *2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids)*, pp. 1045–1051, IEEE, 2016.
- [42] B. Ward-Cherrier, N. Pestell, L. Cramphorn, B. Winstone, M. E. Giannaccini, J. Rossiter, and N. F. Lepora, “The tactip family: Soft optical tactile sensors with 3d-printed biomimetic morphologies,” *Soft robotics*, vol. 5, no. 2, pp. 216–227, 2018.
- [43] W. Yuan, R. Li, M. A. Srinivasan, and E. H. Adelson, “Measurement of shear and slip with a gelsight tactile sensor,” in *IEEE Intl. Conf. Robotics and Automation (ICRA)*, pp. 304–311, 2015.
- [44] S. Dong, D. Ma, E. Donlon, and A. Rodriguez, “Maintaining grasps within slipping bounds by monitoring incipient slip,” in *IEEE Intl. Conf. Robotics and Automation (ICRA)*, pp. 3818–3824, 2019.
- [45] Y. Zhang, W. Yuan, Z. Kan, and M. Y. Wang, “Towards learning to detect and predict contact events on vision-based tactile sensors,” in *Conf. on Robot Learning (CORL)*, vol. 100, pp. 1395–1404, Proceedings of Machine Learning Research, 2020.
- [46] J. W. James, N. Pestell, and N. F. Lepora, “Slip detection with a biomimetic tactile sensor,” *IEEE Robot. Autom. Lett.*, vol. 3, no. 4, pp. 3340–3346, 2018.
- [47] J. W. James and N. F. Lepora, “Slip detection for grasp stabilization with a multifingered tactile robot hand,” *IEEE Trans. Robotics*, 2020.
- [48] R. Koiva, T. Schwank, G. Walck, M. Meier, R. Haschke, and H. Ritter, “Barometer-based tactile skin for anthropomorphic robot hand,” in *IEEE/RSJ Intl. Conf. Intelligent Robots and Systems (IROS)*, pp. 9821–9826, 2020.

- [49] N. Jamali and C. Sammut, “Slip prediction using hidden Markov models: Multidimensional sensor data to symbolic temporal pattern learning,” in *IEEE Intl. Conf. Robotics and Automation (ICRA)*, pp. 215–222, 2012.
- [50] K. Van Wyk and J. Falco, “Calibration and analysis of tactile sensors as slip detectors,” *IEEE Intl. Conf. Robotics and Automation (ICRA)*, pp. 2744–2751, 2018.
- [51] B. S. Zapata-Impata, P. Gil, and F. Torres, “Tactile-driven grasp stability and slip prediction,” *Robotics*, vol. 8, p. 85, Dec 2019.
- [52] F. Veiga, J. Peters, and T. Hermans, “Grip Stabilization of Novel Objects Using Slip Prediction,” *IEEE Trans. on Haptics*, vol. 11, pp. 531–542, Oct 2018.
- [53] E. Kaufmann, A. Loquercio, R. Ranftl, M. Mueller, V. Koltun, and D. Scaramuzza, “Deep drone acrobatics,” *Robotics: Science and Systems (RSS)*, 2020.
- [54] “Robotiq 3-finger adaptive gripper instruction manual.” https://assets.robotiq.com/website-assets/support_documents/document/3-Finger_PDF_20210617.pdf?_ga=2.72042256.2001988252.1629727060-1973336742.1627501348. Accessed: 2021-07-30.
- [55] “Rosbag.” <http://wiki.ros.org/rosbag>. Accessed: 2021-09-01.
- [56] J. L. Ba, J. R. Kiros, and G. E. Hinton, “Layer normalization,” 2016. arXiv preprint arXiv:1607.06450.
- [57] T. Salimans and D. P. Kingma, “Weight normalization: A simple reparameterization to accelerate training of deep neural networks,” 2016. arXiv preprint arXiv:1602.07868.
- [58] P. Remy, “Temporal convolutional networks for keras.” Github, 2020. [Online]. Available: <https://github.com/philipperemy/keras-ten>.
- [59] K. He, X. Zhang, S. Ren, and J. Sun, “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification,” in *Intl. Conf. Computer Vision (ICCV)*, pp. 1026–1034, 2015.

- [60] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin, “Attention is all you need,” in *Advances in Neural Information Processing Systems* (I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds.), vol. 30, Curran Associates, Inc., 2017.