Tactile Dynamic Behaviour Prediction Based on Robot Action*

Kiyanoush Nazari, Willow Mandill, Marc Hanheide, and Amir Ghalamzan Esfahani

School of Computer Science, University of Lincoln, UK
{KNazariSasikolomi,wimandil, mhanheide, aghalamzanesfahani}@lincoln.ac.uk

Abstract. Tactile sensing provides essential information about the state of the world for the robotic system to perform a successful and robust manipulation task. Integrating and exploiting tactile sensation enables the robotic systems to perform wider variety of manipulation tasks in unstructured environments relative to pure vision based systems. While slip detection and grip force control have been the focus of many research works, investigation of tactile dynamic behaviour based on robot actions is not yet sufficiently explored. This analysis can provide a tactile plant which can be used for both control methods and slip prediction using tactile signals. In this letter, we present a data driven approach to find an efficient tactile dynamic model with different tactile data representations. Having evaluated the performance of the trained models, it is shown that the tactile action conditional behaviour can be predicted in a sufficiently long time horizon in future for doing robot motion control.

Keywords: Tactile sensing · Robotic manipulation · Data dimensionality reduction.

1 Introduction and Related Works

Tactile sensation is an essential tool for intelligent interactions with a surrounding environment. In manipulation tasks in particular, we use tactile sensing to help inform and reinforce our understanding of an objects dynamics and physical properties beyond outputs from visual assessments. Grasping an object whose weight was different than visual assessment suggested, or an object's centre of mass was in an unexpected location are examples where visual information must be reinforced with tactile data to produce reliable manipulation of an object. Visual assessment alone typically falls short due to being physically remote and occlusion by the end effector at the points of contact [1].

As manipulation tasks push from structured environments into more realistic real world states, the ability to utilise tactile information for manipulation control tasks becomes a more critical challenge and still remains an open problem. Humans rely heavily on the use of tactile sensing for grasping and manipulation,

* Supported Cancer Research UK

and the difficulty (and complete failure) of manipulation without tactile feedback was excellently shown in Johansson et al. [2] who anaesthetised peoples fingertips and had them attempt to perform a match striking task. Its importance within the robotics community has also been stated in [1], [3], [4], [5].

Romeo et al. [4] state "control algorithms for force regulation or minimisation and grasp stabilisation" as a key open problems for tactile feedback. However, solving these problems has remained a fiendish problem to solve for the key reasons outlined in [1], (i) tactile sensing technology and fabrication is limited to visual information of soft materials or sparse point-wise force measurements, well behind the resolution of a human skin. (ii) modelling contact forces between objects and fingertip are difficult to create. (iii) specifying a desired tactile signal for use in controlled manipulation is also complex to define.

In this paper, a multi-step tactile predictive model is developed to predict future tactile state vector for a pick and move manipulation task with in hand objects using tactile sensors with point-wise force measurements. The contributions of this paper are as follows: We train a variety of deep neural networks and show that a multi-modal recurrent neural network can accurately predict the future fingertip tactile readings of a robot grasping a slippery object while moving through a non-linear trajectory. We show that this prediction model can be trained with entirely unlabelled data. We exploit data compression methods on tactile data to remove the redundancy in tactile sensory information and reduce the computational complexity of tactile dynamic behaviour prediction. Finally, we prove that accurate predictions can be made with 3D, unconfined trajectories of a 7DOF robotic manipulator with variance in pose, velocity and acceleration with a real object.

We demonstrate and evaluate these contributions with a data set of pseudo random trajectories generated by human kinesthetic motions of a Franka Emika robot arm. This arm has two Xela uSkin sensors attached to a two finger parallel type gripper. The force of the gripper is insufficient to keep the object in a stable location and so the motion of the robot creates slipping and eventually the object falls out of the robots grasp. The purpose of this data set is to produce a rich set of sequences where the motion of the robot has direct effect on the sensation felt at the fingertips.

Research in the field of tactile sensor use for control is in the reactive application of slip detection, surveyed in depth in [4]. However, fewer research works can be found in slip prediction. [1] proved that with accurate tactile predictions of a GelSight tactile sensor by a video prediction model introduced in [6], model predictive control could be used to reach the target tactile reading, in this case rolling a ball on the end of a CNC machine to a desired location on a table. While for doing a manipulation task in our case, specifying the goal tactile signal (or image) is not feasible beforehand.

[7] converted the xela uSkin tactile sensor readings to a visual representation; However, there are significant issues with the proposed representation including reduced resolution of the tactile readings, and taxel objects cross over produces

an impossible problem for the prediction model to interpret. The model presented, uses a simplified convolutional LSTM chain structure presented in [6].

Model-based approaches are also applied for tactile data exploitation from the large variety of tactile sensors available [8]. Spike trains analysis [9], threshold on derivatives of normal and shear tactile forces for slip detection [10], applying PCA and Hidden Markov Models for slip prediction [11], using friction cone for slip detection by estimating friction coefficient [12,13] are among these model-based approaches. Model-based approaches usually suffer from being limited to the type of the sensor and gripper and the known object characteristics are required in advance.

Deep neural networks (DNNs) have also been extensively used on processing tactile information; Such as fusing tactile data with other sensory information for texture recognition[14,15], grasp stability estimation [16], train RL policy in a peg-in-hole task [17]. [18] classifies the direction of slip into seven categories by using ConvLSTM cells on the constructed tactile images from BioTac sensor. Having AEs for dimensionality reduction [19] uses multi-dimensional scaling for tactile servoing. [12] divides each manipulation task to four types of manipulation primitives and friction cone slip detection is used to regenerate robot trajectory.

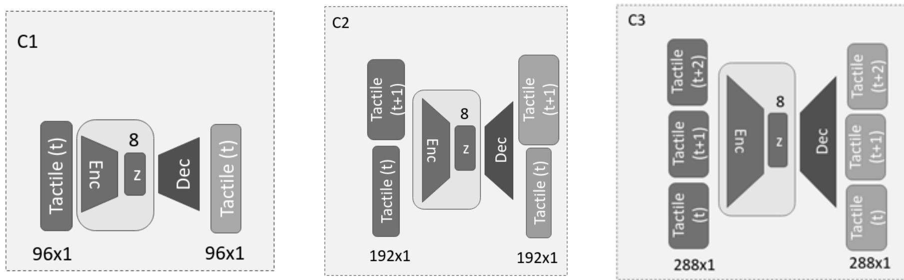
Overall, while tactile signals are exploited for slip detection, grasp stability estimation, stable grasp policy learning, data fusion, grip force control, and robot motion control, only a few [1,18] items try to learn a predictive model to capture the dynamic of a tactile system to predict its behaviour in a sufficiently long time horizon for robot control. We propose an approach in which a model combines recent tactile readings and robot states and based on the future planned robot action, the tactile readings will be predicted in the planned time horizon. The pipeline can be used for different types of tactile data including vision and non-vision tactile sensors and also the trained forward model which learnt the dynamic behaviour of the tactile readings, can be used in different control architectures including tactile Reinforcement Learning controllers.

2 Methodology

2.1 Problem Statement

Let's assume $\vec{S}(t) \in \mathbb{R}^n$ and $\vec{O}(t) \in \mathbb{R}^d$ denote the tactile state and observation space vectors at time t respectively. The external input vector to the system is robot state $\vec{r}(t)$ which effects the tactile interaction with the world. With the curse of dimensionality for $\vec{O}(t)$, there is a highly ambiguous non-linear mapping between the observation to state space. As such, it is desired to map $\vec{O}(t)$ to an abstract lower dimension feature space $\vec{Z}(t)$ which can give us a closer representation to the state space. For the tactile system, $\vec{Z}(t)$ is achieved by applying dimensionality reduction methods on the raw tactile data. The problem of latent tactile state prediction in a time horizon of length τ, can be denoted as follows:

$$\vec{Z}(t+1:t+\tau) = \vec{f}(\vec{Z}(t-\tau:t), \vec{r}(t-\tau:t+\tau)) \tag{1}$$



(a) *AE-C1* (b) *AE-C2* (c) *AE-C3*

Fig. 1: AE structures for tactile data dimensionality reduction

Where $\vec{f}$ is the vector function mapping recent tactile state and robot trajectory to future tactile vector; which is trained by the deep recurrent network. Therefore, the pipeline that we apply for tactile dynamic modelling is to map observation space to lower dimension abstract feature space and then investigate the correlation of tactile features with robot state using the RNN cells.

2.2 Tactile Data Dimensionality Reduction

As the main objective of tactile state signal prediction is for slip prediction, the 96 values of the 32 taxels are having redundant readings which could be more compact by dimensionality reduction methods. In this regard, we use two main approaches for tactile data reduction while trying to avoid any major information loss. First, is the Principal Component Analysis (PCA) which finds orthogonal vectors consisting of linear combination of the original data while preserving as much as variability as possible [20]. In tactile data dimensionality reduction with PCA, keeping at least 80% variance of the data is considered as the threshold value. This means fewer number of principle components which could not preserve the minimum value of the variability are ignored. 20 PCs was the minimum number of PCs which resulted larger than 80% variance.
The second approach is utilising a deep Auto-encoder (AE) [21] for data dimensionality reduction. AEs are self-supervised DNNs which consist of two main components, the encoder which compresses the input data into smaller dimensions, and the decoder which tries to reconstruct the same input data in the output. To include a temporal dependency in encoding the tactile data three types of input data are used for the AEs, I. 96 dimensional tactile vector at each time step *t*, II. 192 dimensional tactile vector including the readings at time *t-1* and *t* stacked row-wise together, and III. 188 dimensional tactile vector including readings at *t-2*, *t-1*, and *t*. We call the three classes of AEs, *AE-C1*, *AE-C2*, and *AE-C3* respectively for easier referencing. Fig.1 presents the structure of the input data for tactile AEs. The depth and hyper-parameters for each AE network are optimised independently.

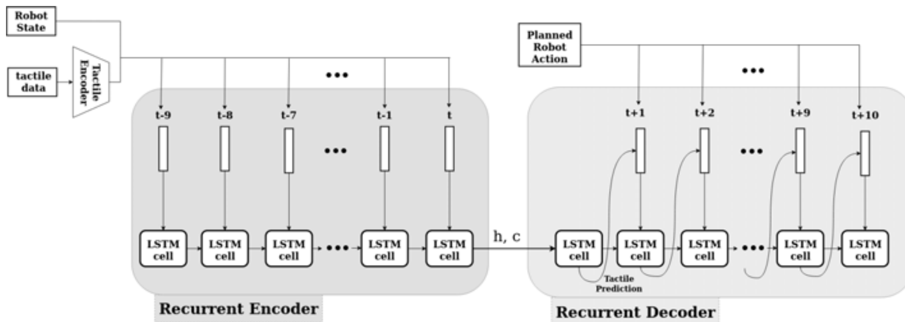


Fig. 2: Recurrent Forward Model

After training, the latent vector $\vec{Z}$ inholds the compact representation of the input data. The dimension of the latent vector is a hyper parameter which is chosen based on a trade-off between how compact the latent vector is and the reconstruction loss. For tactile data compression, the latent vector is a 8 dimensional vector. The size is chosen based on trying various sizes and comparing the loss values.

2.3 Deep Recurrent Model for Prediction

The problem of predicting a system's behaviour in a time horizon with having its recent past data can be categorised in the type of sequence to sequence modelling in the context of deep learning. In the input we are having sequences of tactile readings, robot state, and robot planned action; and in the output future tactile readings are predicted. To process the sequential data we use LSTM cells which are preserving two main components, C_t and h_t which are called cell and hidden state respectively. The update rule for each time step will be defined with the elements called gate values. The *forget*, *input*, and *output* gates are defined by applying sigmoid function over the independent weight matrices multiplied by the input vectors to the cell, scaling them between zero and one to tell the cell how much of the previous state and current data it should forget, take as input or send as output. C_t and h_t can be initialised with either zero or random values after processing each batch of training data.

In tactile prediction model, the hidden state of the LSTM cell should represent the dynamic behaviour of the tactile signals and the correlation between tactile and robot state. Fig.2 shows the architecture of the recurrent model. As Fig.2 shows, the Recurrent Encoder block encodes the tactile and robot state correlation for the past 10 time steps until present time t into h and c. These RNN state vectors are then used as initial state for Recurrent Decoder. From time t onward, at each time step cell's prediction is concatenated with planned robot action at that time step, and then fed back as an input to the next cell. The difference between the input tactile data for Encoder and Decoder blocks is that while in the Encoder ground truth tactile data are combined with past robot states as final input, in the Decoder block, tactile prediction at each time

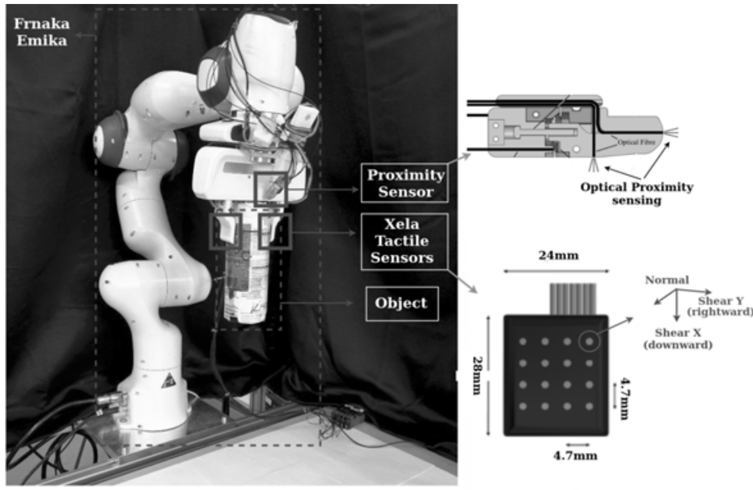


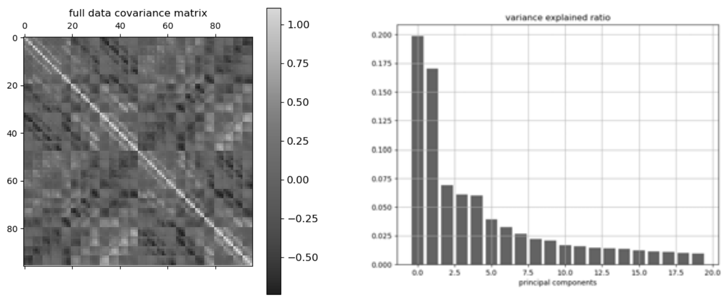
Fig. 3: Experimental setup for in hand manipulation including xela tactile sensors and proximity sensor [22].

step is combined with future robot action as the final input. This helps the Recurrent Forward Model to efficiently use past tactile and robot state and future planned action to deduce future tactile state as equation1 denotes.

3 Experiments

In order to define a manipulation task as a train/test case for tactile state prediction, Franka arm does a pick and move motion consisting of two linear motion trajectory in Z and X directions. The velocity/acceleration profile varies among different pick and move trials. We use a commercially available magnetic based tactile sensor from XELA robotics which is shown in Fig3. The sensor works with average 50hz frequency and has one kg normal force threshold. There are tri-axial readings for each taxel and by having 16 taxels on each sensor the overall 48 readings are available including 16 for normal, shear x and y directions. We have mounted two of the XELA sensors on designed interfaces to connect to Panda Franka normal grippers. Fig3 shows the sensors mounted on the Panda EE.

In order to have labelled data for slip detection we use a proximity sensor attached to the robot wrist which looks directly to the top surface of the object. After grasping the object, as long as the distance measured by the proximity sensor is constant there is no relative motion between the object and the fingers hence no slippage. However, changes in the proximity readings larger than a small specified threshold is indicative of slippage. These proximity readings are then binarized and saved with the data for slip classification labelling purposes. For the object we use a bottle with plastic structure and the weight can change between 5 categories based on the objects inside the bottle. In data acquisition robot, proximity, and tactile data are publishing with 1000, 400, and 50 hz



(a) PCA covariance matrix

(b) variance preserved by each PC

Fig. 4: PCA results with 20 principle components

respectively. All data readings are synchronised in ROS subscriber to have the same frequency for all of the data (50hz).

For data collection for a manipulative task we have applied two approaches to achieve rich enough data for training our tactile prediction recurrent model. The first approach was to do a linear motion pick and place task for the bottle and send Cartesian velocity profiles with trapezoidal shapes and different acceleration/deceleration values to the robot. This is designed to have structure for robot motion for doing control in the next steps of the project once the forward model is learned. As such, for an object with constant weight, while with lower acceleration the task was completed successfully, by increasing the acceleration in the profile the object was dropped in the middle of the task. Therefore, robot motion profile was the only parameter causing the slippage and hence the changes in tactile readings. The second strategy was data collection with robot kinaesthetic teaching. In this mode a human operator does various manipulation motions until the object is dropped. The benefit of this mode is that the zero impedance robot motion lets the user to execute motions with large variety of velocity and acceleration ranges; which might not be possible with robot motion controller in the previous mode. This will provide richer tactile data for training the model.

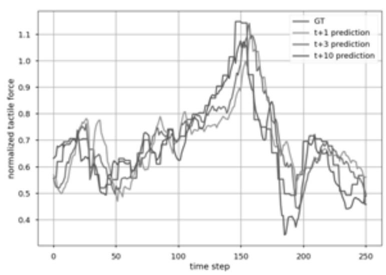
4 Results

In order to achieve the most precise tactile behaviour prediction we have utilised data dimensionality reduction methods to further remove the redundancy of the tactile readings and reduce the computational complexity of the prediction problem. As such, the performance of the recurrent model will be evaluated by three types of tactile inputs: 1. Original raw data, 2. principle components (PCs) resulted from PCA, and 3. latent vectors of three classes of AEs namely *AE-C1*, *AE-C2*, and *AE-C3*. For the implementation, PCA function from python's

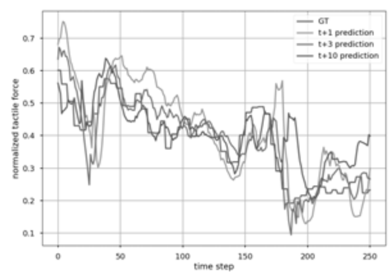
Table 1: Evaluation of recurrent forward models

Tactile Input	Validation Loss†	Test Loss†
raw data	0.0484	0.0552
AE-C1	0.0330	0.0403
AE-C2	0.0351	0.0453
AE-C3	0.0405	0.0510
PCA	0.0298	0.0363

† *MSE* Loss



(a) first principle component

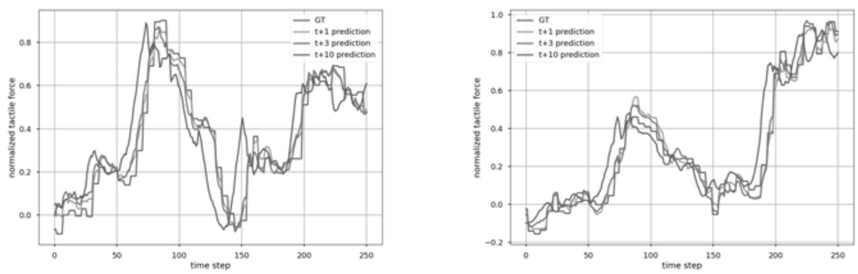


(b) second principle component

Fig. 5: Two main PCA principle components and prediction with recurrent model

sklearn library is used. Fig.4 shows the result of applying PCA on the raw tactile data. Fig.4a presents the covariance matrix between the PCs and Fig.4b illustrates the variance contained in each of the 20 PCs. Overall, 82% of the variance are preserved by all of the PCs.

Having the tactile data analysed by dimensionality reduction methods, we can now evaluate the performance of the recurrent model on the compressed tactile data. As Fig.2 shows we combine tactile data with the robot state as the input for the Recurrent Encoder block; and the predicted tactile vector at each time step and the robot action as the input to the Recurrent Decoder block. As table1 shows the PCA and *AE-C1* tactile input resulted in the best prediction results. Fig.5 shows the ground truth signal and the model prediction for the first two major PCs of the dimensionality reduction resulted from PCA. For visualising model prediction, the ground truth tactile feature vector is plotted alongside the prediction for three different time steps in prediction horizon. The graphs for $t+3$, $t+5$, and $t+10$ are plotted without the shift in time axis, since these data are available at time step t and are compared with the ground truth data available at that time. Fig.6 presents the predictions for two components of the C1 AE latent vector. It can be observed that $t+10$ prediction signal captures the change in ground truth signal before it happens. This is a desired feature to enable the model to be used for slip prediction. Although the loss value for PCA input is smaller than the corresponding value for AE-C1 (table1), the AE-C1 model has the advantage to predict the sudden rise of the latent tactile feature before it happens in the ground truth.



(a) 2nd latent vector components (b) 4th latent vector components

Fig. 6: recurrent model prediction for sample latent vector components

5 Conclusion

We have applied a data driven approach on exploring action conditional tactile dynamic behaviour with deep recurrent neural networks. Data compression could improve the tactile prediction accuracy and diminish the computational complexity of the problem. Having been trained on the collected data from the real robotic setup, the model can predict the tactile behaviour for slippage case in advance of time which is the desired purpose of the model for slip prediction in the next step. With this learnt tactile dynamic, the models can now be used for a closed loop control for object slippage avoidance.

References

1. S. Tian, F. Ebert, D. Jayaraman, M. Mudigonda, C. Finn, R. Calandra, and S. Levine, “Manipulation by feel: Touch-based control with deep predictive models,” in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 818–824.
2. 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.
3. H. Yousef, M. Boukallel, and K. Althoefer, “Tactile sensing for dexterous in-hand manipulation in robotics—a review,” *Sensors and Actuators A: physical*, vol. 167, no. 2, pp. 171–187, 2011.
4. R. A. Romeo and L. Zollo, “Methods and sensors for slip detection in robotics: A survey,” *IEEE Access*, vol. 8, pp. 73 027–73 050, 2020.
5. R. S. Dahiya, G. Metta, M. Valle, and G. Sandini, “Tactile sensing—from humans to humanoids,” *IEEE transactions on robotics*, vol. 26, no. 1, pp. 1–20, 2009.
6. C. Finn, I. Goodfellow, and S. Levine, “Unsupervised learning for physical interaction through video prediction,” *arXiv preprint arXiv:1605.07157*, 2016.
7. X. Zhou, Z. Zhang, X. Zhu, H. Liu, and B. Liang, “Learning to predict friction and classify contact states by tactile sensor,” in *2020 IEEE 16th International Conference on Automation Science and Engineering (CASE)*. IEEE, 2020, pp. 1243–1248.

8. A. Yamaguchi and C. G. Atkeson, "Recent progress in tactile sensing and sensors for robotic manipulation: can we turn tactile sensing into vision?" *Advanced Robotics*, vol. 33, no. 14, pp. 661–673, 2019.
9. Z. Yi, Y. Zhang, and J. Peters, "Biomimetic tactile sensors and signal processing with spike trains: A review," *Sensors and Actuators A: Physical*, vol. 269, pp. 41–52, 2018.
10. M. Saen, K. Ito, and K. Osada, "Action-intention-based grasp control with fine finger-force adjustment using combined optical-mechanical tactile sensor," *IEEE Sensors Journal*, vol. 14, no. 11, pp. 4026–4033, 2014.
11. N. Jamali and C. Sammut, "Slip prediction using hidden markov models: Multidimensional sensor data to symbolic temporal pattern learning," in *2012 IEEE International Conference on Robotics and Automation.* IEEE, 2012, pp. 215–222.
12. F. R. Hogan, J. Ballester, S. Dong, and A. Rodriguez, "Tactile dexterity: Manipulation primitives with tactile feedback," in *2020 IEEE international conference on robotics and automation (ICRA).* IEEE, 2020, pp. 8863–8869.
13. W. Chen, H. Khamis, I. Birznieks, N. F. Lepora, and S. J. Redmond, "Tactile sensors for friction estimation and incipient slip detection—toward dexterous robotic manipulation: A review," *IEEE Sensors Journal*, vol. 18, no. 22, pp. 9049–9064, 2018.
14. K. Takahashi and J. Tan, "Deep visuo-tactile learning: Estimation of tactile properties from images," in *2019 International Conference on Robotics and Automation (ICRA).* IEEE, 2019, pp. 8951–8957.
15. S. Luo, W. Yuan, E. Adelson, A. G. Cohn, and R. Fuentes, "Vitac: Feature sharing between vision and tactile sensing for cloth texture recognition," in *2018 IEEE International Conference on Robotics and Automation (ICRA).* IEEE, 2018, pp. 2722–2727.
16. A. Murali, Y. Li, D. Gandhi, and A. Gupta, "Learning to grasp without seeing," in *International Symposium on Experimental Robotics.* Springer, 2018, pp. 375–386.
17. M. A. Lee, Y. Zhu, K. Srinivasan, P. Shah, S. Savarese, L. Fei-Fei, A. Garg, and J. Bohg, "Making sense of vision and touch: Self-supervised learning of multimodal representations for contact-rich tasks," in *2019 International Conference on Robotics and Automation (ICRA).* IEEE, 2019, pp. 8943–8950.
18. B. S. Zapata-Impata, P. Gil, and F. Torres, "Learning spatio temporal tactile features with a convlstm for the direction of slip detection," *Sensors*, vol. 19, no. 3, p. 523, 2019.
19. G. Sutanto, N. Ratliff, B. Sundaralingam, Y. Chebotar, Z. Su, A. Handa, and D. Fox, "Learning latent space dynamics for tactile servoing," in *2019 International Conference on Robotics and Automation (ICRA).* IEEE, 2019, pp. 3622–3628.
20. I. T. Jolliffe and J. Cadima, "Principal component analysis: a review and recent developments," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, p. 20150202, 2016.
21. A. Ng *et al.*, "Sparse autoencoder," *CS294A Lecture notes*, vol. 72, no. 2011, pp. 1–19, 2011.
22. J. Konstantinova, G. Cotugno, A. Stilli, Y. Noh, and K. Althoefer, "Object classification using hybrid fiber optical force/proximity sensor," in *2017 IEEE SENSORS.* IEEE, 2017, pp. 1–3.