A Data-driven approach to softness estimation using tactile sensing

by

Thijs Merkx

Student Name	Student Number
Thijs Merkx	4650425

In partial fulfilment of the requirements for the degree of
Master of Science
at Delft University of Technology,
to be defended publicly on Wednesday August 28, 2024 at 1:45 PM.

Faculty:	Mechanical Engineering
Department:	BioMechanical Engineering
Programme	BioMechanical Design

Mentors / Supervisors:	Dr. M. Wiertlewski G. Vitrani
Graduation committee:	Dr. M. Wiertlewski Dr. C. Della Santina G. Vitrani

TU Delft

A Data-driven approach to softness estimation using tactile sensing

Thijs Merkx

***Abstract*—The accurate prediction of object softness is crucial in many fields, from agriculture to medical care. Vision-based tactile sensors, which capture high-resolution images of contact interactions, have shown great potential in determining this material property. Many existing approaches, particularly those using end-to-end models, suffer from a 'black-box' problem where it is difficult to understand which features the models use to make their predictions. This lack of transparency makes it challenging to determine and correct errors. To overcome this, this paper shows a data-driven method that can decode information from acquired tactile images to extract pressure distributions and assign softness levels to different objects. A novel approach is explored, integrating a Convolutional Neural Network (CNN) to predict the pressure distribution and a Long Short-Term Memory (LSTM) network to assess material softness. It is demonstrated that the CNN model effectively learns necessary features from the tactile images, enabling precise pressure distribution predictions. Concurrently, the LSTM model analyzes temporal sequences of tactile data, accurately predicting material softness and differentiating ripe from overripe fruits. By utilizing the spatiotemporal pressure distribution, this method improves on existing methods by enabling the efficient use of tactile data and providing additional information that can be used to further enhance the model. This paper can be used as a stepping stone to a more complex system in which robotic control can be implemented based on the sensed material properties, allowing for better control loop mechanisms and expanding the applications of tactile sensing technologies.**

***Keywords* —** Tactile sensing, softness estimation, spatiotemporal pressure distribution, CNN, LSTM

I. Introduction

Determining the stiffness of an object is an everyday task for humans, integral to a wide range of activities, from handling delicate items to gauging the firmness of materials. For robots, however, this task is far from trivial. Current methods of estimating stiffness often include destructive methods, making it unsuitable for many applications where delicate handling is required [4]. Other methods used for medical applications frequently introduce unwanted side effects and discomfort for patients [9]. The ability to estimate material stiffness is essential for advanced robotic systems, enabling robots to handle objects with the appropriate force and dexterity and increasing patient comfort. Tactile sensing, which provides detailed information about the environment and objects, is critical. Effective stiffness estimation allows robots to adjust their actions based on the properties of the materials they interact with, thereby enhancing their performance in various applications, including manufacturing, medical care, and agriculture.

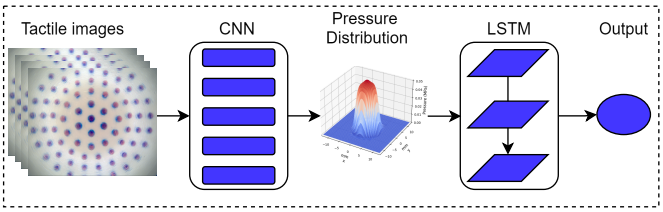

Figure 1. Overview of the tactile sensing pipeline. A CNN processes the acquired tactile images to predict the pressure distributions. These distributions are subsequently analyzed by an LSTM network to produce the final output, representing the estimated material softness.

Understanding human stiffness perception is crucial in developing effective tactile sensors for robots. Humans rely on multiple haptic cues to determine the compliance or stiffness of an object. For objects with sufficient thickness relative to the deformation, the compliance can be expressed by Young's modulus, defined as the applied stress divided by the strain, where the stress is determined by the applied force, F, over the area where the force is applied, A. The strain is defined as the change in length of the material, Δl, divided by its original length, l.

$$Y = \frac{F/A}{\Delta l/l} \tag{1}$$

The stiffness of an object is defined by the ratio of applied force, ΔF, and displacement, Δx:

$$S = \frac{\Delta F}{\Delta x} \tag{2}$$

While stiffness and Young's modulus are related, they are distinct concepts in material science. Stiffness is a measure to determine the resistance of an entire object or structure to deformation when subjected to an external force. This measure depends on the object's material properties and geometry. This makes thicker objects stiffer than thin objects made from the same material. Young's modulus, on the other hand, is a fundamental material property that measures the intrinsic stiffness of a material itself, independent of its shape or size. Young's modulus provides a better way to compare different objects, as it is a constant for a given material under small deformations.

For human perception, cutaneous and kinesthetic cues are used to discriminate objects on perceived hardness. Srinivasan and LaMotte [18] showed that subjects could discriminate deformable objects using only cutaneous cues. This changes when objects do not have a deformable surface. In these cases,

kinesthetic information was required alongside cutaneous cues. Further research into this field showed that the contribution of the cutaneous tactile cues is about 90%, with the kinesthetic cues only accounting for roughly 10% when considering a statistically optimal approach. This finding suggests that surface deformation and pressure distribution are essential when determining the hardness of deformable objects [1].

Pressure and its temporal evolution are crucial for softness estimation because they provide information on dynamic interaction, revealing how an object resists deformation over time as pressure is applied. This temporal pressure profile helps in understanding the material properties, as stiffer objects exhibit less deformation and more rapid changes in pressure distribution compared to softer ones. Next to this, temporal information provides information on viscous objects. Even when the finger remains in contact with the object, the pressure profile changes due to the viscosity of the object. This is further illustrated in figure 2. Humans use both the spatial and temporal information of pressure to accurately perceive the stiffness, with continuous feedback updating the brain on the object's properties. Additionally, the temporal evolution of pressure reflects the object's compliance and resistance, offering a nuanced understanding of its response to touch. Thus, pressure and its temporal evolution are essential for accurately estimating an object's stiffness, providing comprehensive information about its material properties and response to applied forces.

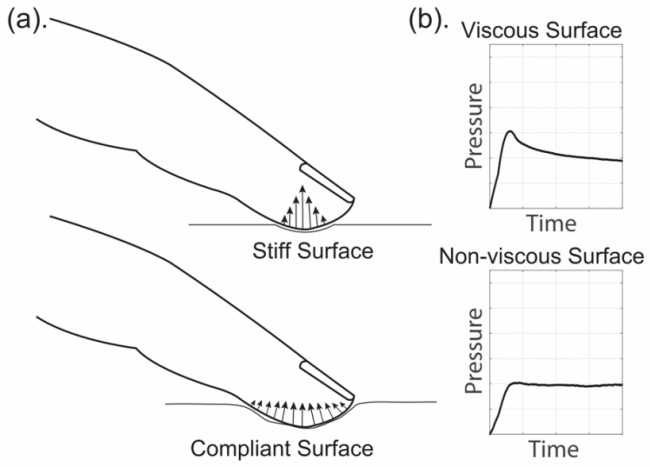

Figure 2. (a) Spatial force distribution comparison showing the difference in the contact area and peak force between surfaces with different levels of compliance. (b) Temporal pressure of viscous and non-viscous surfaces, showing a declining pressure over time while remaining in contact due to viscosity compared to constant pressure for non-viscous surfaces. [15]

It is crucial to delve into the nuances of tactile sensing and its applications in robotics to bridge the gap between human perception and robotic capabilities in stiffness estimation. By understanding how humans utilize both cutaneous and kinesthetic cues for assessing material properties, more advanced and effective tactile sensors for robots can be developed. These sensors can mimic human sensory processes, enabling robots to accurately estimate stiffness and adapt their interactions with various objects. Integrating this knowledge into robotic systems improves performance across numerous fields, from delicate object handling to enhanced patient care in medical applications.

Robotic tactile sensing has garnered significant attention due to advances in sensor technology and computing power. Various tactile sensors have been developed to estimate interaction properties such as deformation, slip detection, and material properties like stiffness. While interaction properties can often be directly inferred from the sensor output, based on the interaction between object and sensor, the tactile cues required to perform accurate stiffness estimation are not as trivial to acquire. Vision-based tactile sensors, which acquire high-resolution images of contact interactions, are particularly promising for determining object stiffness. These sensors can capture detailed information that is crucial for accurate stiffness estimation. Studies have shown that using spatial-temporal pressure data to determine an object's stiffness gives optimal results, similar to how humans estimate an object's stiffness [15].

Several methods can be used to process the acquired data to extract the required information. Mathematical models and image processing techniques can be used to determine linearised models that map the images to specific properties. However, when aiming to capture non-linearities, neural networks have shown to be capable of learning essential features that can also capture non-linearities that otherwise would be lost. These neural networks can be valuable tools for learning complex features embedded in the data. However, setups using such networks often lack the link relating the underlying physical properties to the resulting outcome. This occurs when the data is directly used as the input to a system of neural networks. This so-called end-to-end approach results in difficult-to-understand 'black-box' frameworks. Due to the complex nature of features that must be learned from the tactile image, large datasets and long training times are required. This also increases the risk of overfitting the model, making it unsuitable for dynamic environments where the actual data differs from the training data.

Based on the findings of how humans perceive stiffness and previous studies done in the field of tactile sensing, this thesis will focus on creating a framework to extract spatiotemporal pressure data and relate this to the object's stiffness (figure 1). To accomplish this, the development of a tactile perception framework will be explored, aimed at estimating the softness of objects using high-resolution tactile images obtained by using the latest version of the vision-based tactile sensor called the ChromaTouch [17]. By leveraging this data, a combination of a convolutional neural network (CNN) and a Long short-term memory network (LSTM) is trained to interpret tactile images and estimate the softness of objects. With this approach, the research aims to improve on existing methods, making more efficient use of acquired tactile data and using embedded pressure data to identify specific areas in which the framework makes its errors. This provides a more transparent and interpretable approach to softness estimation that could significantly enhance robotic manipulation capabilities.

This thesis is structured as follows: section II presents related works in this sector, and section III details the experimen-

tal setup and hardware used for this research. Section IV will dive into data collection and research methodologies. Section V presents the results of the tactile perception framework, while sections VI and VII offer a discussion and conclusion, showing the possible improvements and recommendations for future research.

II. Related works

Based on human stiffness perception, researchers were able to determine which aspects play an important part in correctly differentiating between soft and hard objects. From this, it has been identified that contact area and indentation depth are the two most important cues humans rely on when determining an object's stiffness [7, 14]. The relationship between contact area and indentation depth and a finger in contact with an object can be described by the Hertzian contact model [8]. The deformation of the finger and object is proportional to the applied force, Young's modulus, and the shape of the finger and object. This relation is given by the following equation:

$$F = \frac{4E^* a^3}{3R}, \tag{3}$$

with E^* and R the equivalent Young's modulus and radius of the finger and the object and a the contact radius between the two. An important consideration when applying this is the thickness of the material. When looking at thin slabs, this relationship is influenced by the thickness of the slab, with very thin slabs significantly increasing the perceived stiffness [6]. The change in the relation between applied force, Young's modulus, and contact area radius can be seen in the following two equations:

$$F = \frac{4E^* a^3}{3R}(1 + 0.15(\frac{a}{h})^3), \tag{4}$$

where h is the object's thickness, showing that the thickness should be sufficiently large compared to the contact area such that this term effectively becomes zero.

A second relationship can be found between the applied force and the pressure by integrating over the circular contact area giving:

$$F = \int_0^a p(r) 2\pi r dr = \frac{2}{3} p_0 \pi a^2 \tag{5}$$

Using equation 3, the contact area can be related to the applied force and equivalent Young's modules according to:

$$A = \pi a^2 = \pi(\frac{3FR}{4E^*})^{\frac{2}{3}} \tag{6}$$

This shows that objects with varying Young's moduli result in different rates of how the contact area between the two bodies changes over time, also known as the contact area spread rate (CASR). In turn, CASR has been linked to increased precision in human stiffness perception [2]. Apart from the CASR, the measured pressure distribution and its change are also related to the object's Young's modulus. Using this relation, human stiffness perception can be translated to robotic stiffness estimation using tactile sensors capable of measuring this pressure distribution.

To investigate the effectiveness of this approach, research in robotic applications has examined pressure measurements at various levels of complexity, from single-point pressure measurements to comprehensive spatiotemporal pressure distributions, as described previously. Next to this direct measurement method, a different group of researchers focused on using end-to-end machine learning models for stiffness estimations, where tactile information is directly used as the input for various machine learning models that output a stiffness value.

Previous works have shown that simple sensor setups, only capable of measuring a single pressure point can be used to determine initial contact and local properties. However, often additional information is required to determine stiffness correctly and not only the contact [13]. Increasing the complexity by measuring spatial pressure distributions is frequently used in binary tasks. This can be used for manufacturing environments where the internal state of a product needs to be measured [5] or for medical applications where the presence or absence of a lump has to be determined [10]. However, when the environment is less controlled and more precise information is required, spatiotemporal pressure distribution is required. This becomes even more important when handling soft deformable objects, where, due to the deformation, the temporal information can give crucial additional details, as shown by Nonaka et al. [15], using an array of direct pressure sensors to determine an object's stiffness. Such methods show great results but often employ sensors that can directly measure pressure using pneumatic or hydraulic systems and complex systems using sensitive strain-gauges [21]. Thus, different methods should be employed in environments where this is impossible.

As seen from the equations derived from the Hertz contact model, CASR can also be used to measure an object's compliance instead of using the pressure distribution. This can be done by using a vision-based tactile sensor. These sensors typically use a deformable surface and a camera to capture high-resolution tactile images. These images contain a lot of information about the object and grasping conditions. For CASR, a Tactip sensor [19] is used. This sensor uses reflective markers placed on pins to increase the movement of the markers, making it easier to track the motion. Using a model-based approach, this sensor has been used to determine the stiffness of three different objects. Here, the pins were tracked using an optical flow method. From this, they looked at the CASR to calculate how it changed over time to determine the stiffness of the objects. This method shows great promise, however, as it relies solely on the visible changes in CASR, it can only detect changes when large enough forces are applied. This results in the sensor being unable to determine small differences and thus unable to classify very soft objects and objects with comparable softness.

With the rise of vision-based tactile sensors, the use of end-to-end machine learning methods also increased. These sensors are often paired with end-to-end machine learning methods due to the large amount of information embedded in the acquired tactile images. The sensors most often used in combination with such a method are the Gelsight and its variances [11], which uses several RGB-colored LEDs to illuminate a soft deformable surface with a reflective coating,

allowing it to detect changes in the surface based on the alterations in color patterns. This sensor is used by Yuan et al. [20] to determine an object's hardness. Several linked neural networks directly translate the information-rich tactile images to object hardness. Similar results are shown by Chen et al. [4], using a soft, elastic sensing layer with high-stiffness bumps attached to determine an object's overall stiffness and local changes.

While these studies accomplish good results in estimating the stiffness of several objects, the lack of intermediate steps removes the possibility of understanding which properties are essential to the networks. If any errors arise in this case, it is next to impossible to determine the origin of such errors. This black-box phenomenon reduces the understanding of the interactions between sensor and object, requiring large datasets to extract the necessary information. By determining intermediate interaction properties such as spatiotemporal pressure distribution, this can be reduced by focusing on more specific data. Extracting this information to connect the underlying physical relationship between pressure and Young's modulus can result in a more accurate estimate of the object's compliance.

III. Experimental Setup

This section will explain the hardware used for this project and give an overview of the setup used in data acquisition. The setup consists of the ChromaTouch sensor, linear guides for precise motion control, and several objects used to create indentation profiles on the sensor (figure 3). Each of these will be described in further detail below.

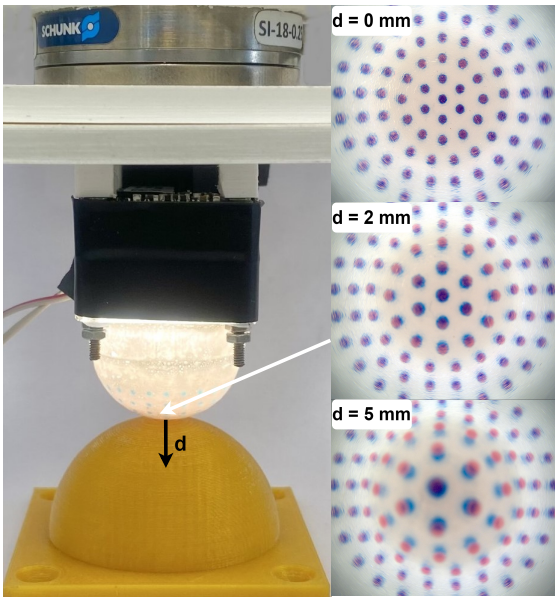

Figure 3. ChromaTouch sensor used for normal indention. Marker size and hue change under increasing force.

A. ChromaTouch

The ChromaTouch sensor used in this experiment is an adapted version of the one presented by Scharff et al. [17]. This version has a similar shape and production method, however, the dome has a radius of 25.4mm and contains 80 trackable markers on its shell. The small silicone dome is placed on top of a Basler Dart daA1920-160uc camera with a fisheye lens attached with a resolution of 1920x1200 pixels. The focal point of the lens is placed at the center of the dome, projecting the markers on the flat camera image. The soft dome of this sensor allows for large deformations under relatively low force, allowing it to detect small changes and softer objects more easily. However, this characteristic also introduces large visco-elastic properties that should be taken into account for stiffness perception. The outside of the dome is covered in white PlastiDip, eliminating the effects of external light sources. To provide sufficient lighting within the dome, a ring of RGB LEDs ensures consistent lighting conditions.

B. Sensing setup

To make data acquisition consistent, a setup was used to attach the sensor to a Thorlabs NRT150/M linear guide with precise position control. The linear guide's precision is critical for consistent data acquisition, allowing for detailed analysis of the sensor's response to various forces and deformations. To record the forces caused by the indentation, an ATI Nano43 force-torque sensor is used. This sensor can measure forces up to 10N with a resolution of 1/512 N. The high-resolution force measurements are synchronized with the tactile images, providing a coupled dataset of images and forces for further processing.

C. Material samples and hardness variations

Several samples with varying hardness levels were created to ensure a diverse dataset. These samples include SORTA-Clear 12 silicone mixed with different percentages of thinner (0%, 10%, 30%, 40%, and 50%), as well as Solaris silicone with 0% and 10% thinner. Adding thinner reduces the silicone's Young's modulus proportionally, creating a range of samples with different stiffness characteristics as shown in figure 4. Since all samples were manually created, minor discrepancies from this ratio can be expected. Apart from these silicone samples, three 3D-printed objects out of PLA were used to simulate the stiffest objects. This consists of a flat plate and two domes with diameters of 20mm and 40mm.

IV. Methodology

This section will explain the methods and requirements for detecting an object's stiffness based on tactile images using several image-processing techniques and multiple neural networks.

A. Contact Mechanics

As described previously, the Hertz contact model can be used to determine the pressure distribution caused by the imposed indentation and the measured force. This model approximates the contact area based on deformation from

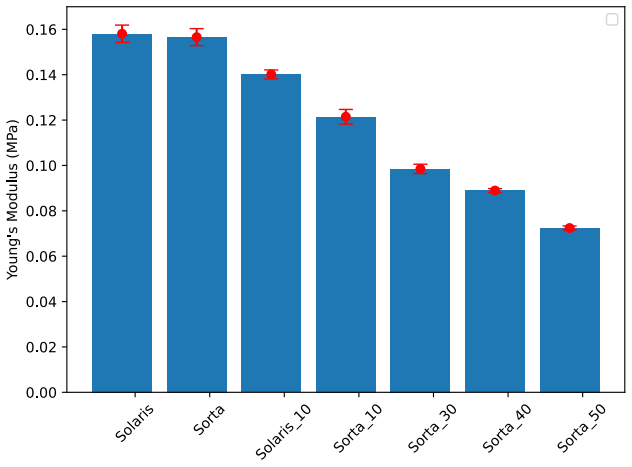

Figure 4. Young's modulus values for different silicone samples used in the experiment.

indentation and radii of both the sensor and object according to the following formula:

$$a = \sqrt{Rd} \tag{7}$$

with a being the contact area, R the equivalent radius, and d the deformation. A pressure distribution profile can be created with the contact area and force readings from the attached force sensor. Since the displacement is centered around the center of the sensor, the local pressure is on the distance from the center with the following ratio:

$$p(r) = p_0(1 - \frac{r^2}{a^2})^{\frac{1}{2}} \tag{8}$$

Where p_0, the maximum contact pressure, can be extracted from rewriting equation 5 resulting in

$$p_0 = \frac{3F}{2\pi a^2} \tag{9}$$

B. Data collection

Using the setup shown in figure 3, data was collected in the following way: A sample was indented into the center of the sensor and a set indentation profile was chosen. This profile was chosen to mimic the human grasping action, indenting and releasing the objects. Two indentation speeds were selected to capture the visco-elastic effects: 0.5 mm/s and 1.0 mm/s. This ensures the relaxation effects that occur with lower speeds are captured as well. Next to this, it also increases data robustness by introducing more varied data and better simulating real-world scenarios where slower or faster movements might be required for specific applications. This cycle was repeated three times for each speed to create data that captures minor variations that might occur due to slight variations in external conditions. To ensure the capture of the complete indentation cycle, the recording was started before indentation and stopped after the sensor returned to the original position, as shown in figure 6. This was done to ensure the linear guide got up to speed before indenting the sensor and that the sensor was not indenting the sample before recording the forces and images. To calibrate when the sensor touched the sample, the force readings were examined, and a threshold was used to determine the moment of initial contact. The speed of the linear guide and force timestamps were used to calculate the final indentation. This resulted in an average indentation of 3.5mm per trial. The images and forces are recorded at approximately 10 Hz. The force recording was done alongside the image acquisition. To ensure stability in the force readings, an average of samples were taken to ensure high-frequency fluctuations were filtered out.

C. Image pre-processing

For the CNN to reconstruct the ground-truth pressure distribution given by the Hertz contact theory, several processing steps must be taken. Firstly, the pressure distribution is discretized at the center of each of the markers. Several image processing techniques have been implemented to determine these centers. A Gaussian filter is applied to smooth the image, followed by background removal to mitigate the effects of uneven illumination and to remove noise. To determine the centers of each marker, the image is first transformed from RGB to HSV color space and subsequently segmented based on the magenta and cyan hue threshold to be able to separate these markers. The center of each marker is determined using a binary representation of the segmented image using the *regionprops* function provided by the *skimage* library in Python. Each marker is assigned the ground-truth value for the local pressure based on equation 8 and its center coordinates. Finally, the whole image is split into sub-images for each marker to enable the CNN to learn the required features. This ensures the entire marker is visible in the image, especially given that the markers change size based on indentation. By focusing on smaller, more localized regions, the CNN can learn the intricate features and variations in pressure distribution embedded in the markers more effectively. This approach also helps reduce the training data's complexity, allowing the CNN to converge more quickly and accurately.

D. CNN Model and training process

Based on previous work done by Dirk-Jan Boonstra and Max Polak, presented in [3] and [16], the decision was made to continue with the Shufflenet V2 [12] model and architecture. This lightweight model offers several benefits, such as the ability to train on relatively low-end hardware, saving time during training, and being able to process small images quickly, decreasing the processing time. This network's last fully connected layer is replaced to ensure a single output value is created by regression. The outputs for each sub-image per timestep are combined to reconstruct the full pressure distribution. This pipeline can be seen in figure 5. The training of the CNN was done on an NVIDIA GeForce RTX 3060 Ti. The dataset was divided into a train and test set, containing 80% and 20% of the data respectively. The model's pre-trained weights were discarded to ensure only the correct features from the new data set were learned. The new weights and biases were tuned during training using backpropagation. The Ranger21 optimizer with its standard parameters: *learning rate*

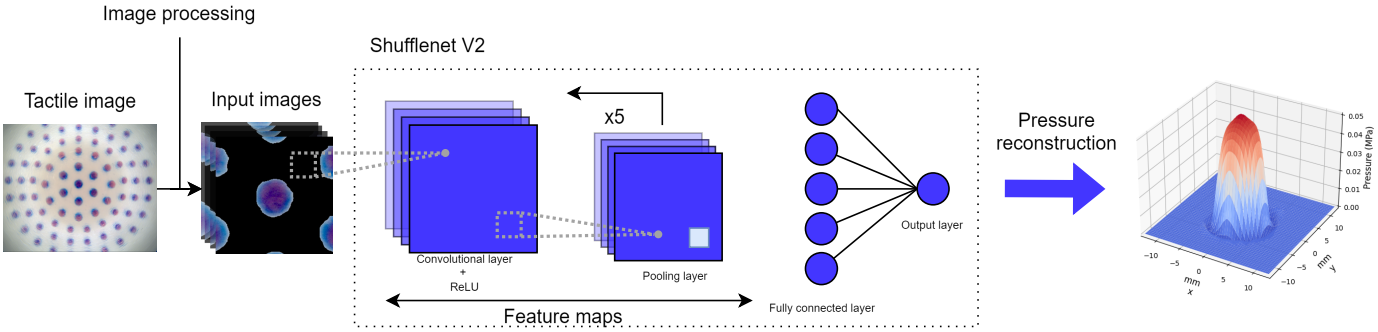

Figure 5. Workflow for pressure distribution detection. Images are pre-processed, split into sub-images, and used as the input for the Shufflenet V2 network. This network consists of 5 stages of convolution and pooling layers. The final layer is a single neuron predicting the local pressure value. These are reconstructed into a 3D pressure distributions.

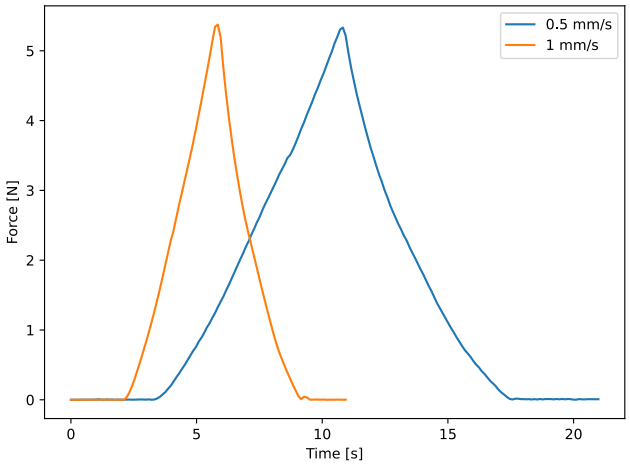

Figure 6. Force profiles during indentation and retraction phases at two different speeds. The profiles show how force is applied and released during the experiment.

= *0.001*, *weight decay* = 1×10^{-10}, β's = (0.9, 0.999) and ϵ = 1×10^{-8}. To ensure training improved over time, a scheduler was used that decreased the learning rate with a factor of 0.1 and a patience of 10. The Mean Square Error (MSE) was used as a metric to evaluate the model's output. For additional validation, a new set of images was collected to ensure its robustness on never-before-seen data.

E. LSTM Network for temporal pressure data analysis

To determine the object's softness. This network takes a temporal sequence of pressure data as input and makes predictions based on this sequence. The main benefit of this network is that it has a 'memory', meaning the outcome depends on the entire sequence and not a single time point. This allows the algorithm to learn the differences in how the pressure distribution changes over time for objects with different Young's modulus. Two distinct approaches were employed: a classification task and a regression task. Both tasks utilized a similar model architecture, consisting of two layers with a hidden dimension size of 128. The primary difference between the two tasks lies in the model's output. Three output neurons were used for the classification task, enabling the model to categorize the objects into three distinct classes. A softmax function was used to transform the output values of each of the neurons into a probabilistic distribution for each class. This ensures stability during training and makes the output more interpretable. A cross-entropy loss was used to evaluate and improve the model during training. This calculates the error between each of the predicted probabilities for each class compared to the given ground-truth class. For the regression task, only one output neuron is required to estimate the objects' Young's modulus. Similar to the CNN, the MSE was used as a performance metric.

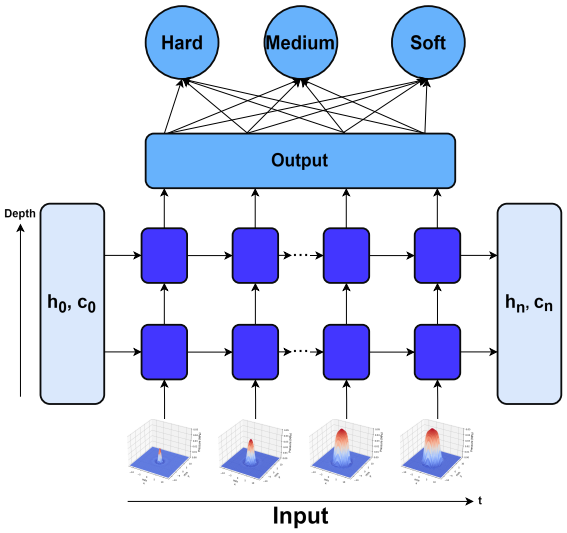

Figure 7. LSTM network used for the classification task. The input to the network is the pressure distribution over time (horizontal axis). The depth of the network is shown vertically, and three output neurons on top are connected to the hidden state of the final layer. Each blue cell represents a single LSTM node containing the forget gate, input gate, candidate memory, and output gate.

V. Results

The developed method demonstrates the capability to accurately reconstruct the pressure distribution, classify objects with varying hardness levels, and determine the Young's modulus of various objects within similar levels of compliance.

This approach was validated using diverse datasets, including synthetic samples and real-world objects like fruits, showcasing its robustness. This section will dive deeper into each of the individual parts of the pipeline, identifying key results.

A. *Physical space calibration*

For the final training of the CNN, data collected from the 10 samples resulted in a total of 170.000 images, split into train and test sets. Four new sets of data were created to ensure the network is validated on new data and at every step of the indentation process to determine where the most significant errors occur.

A comparison test was conducted to determine the effect of surrounding markers on the center marker. Three image sizes have been chosen, ranging from the smallest input size for the Shufflenet V2 network, containing a marker with some partly markers surrounding it, to double this size (224x224, 336x336, and 448x448 pixels) (see Appendix A). It should be noted that only markers (and their surrounding windows) that lie entirely within the image (e.g., not at the edges) are selected to ensure a complete image (without fill) such that the CNN does not learn from its location on the sensor but rather from the marker features. Thus, larger window sizes reduce the amount of data that can be extracted from the image. From the results shown in figure 8, it can be seen that the difference in performance is negligible, thus to improve the detection speed and allow for the highest number of included markers, the smallest image size was chosen. Since this difference is negligibly small, the extra information gained from a larger window does not help the algorithm learn the pressure better and possibly introduces more noise generated from the surrounding markers.

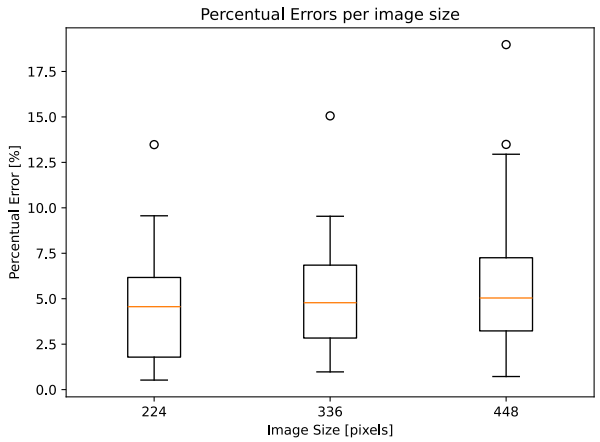

Figure 8. CNN performance comparison on three different window sizes of the percentual error between the predicted value and the ground truth.

To evaluate the overall performance of the CNN, the results for each sample are recorded, and the error score is illustrated in figure 9. The overall average error of the model is 4.45%. The best performance can be seen for the small and large 3D printed dome with mean errors of 1.48% and 1.57% respectively. The highest errors are found in the two softest samples, with mean errors of 8.09% and 8.45%. From these results, it can be seen that the error increases when the object's hardness decreases. This can be attributed to the fact that these softer objects experience higher levels of visco-elasticity, thus causing less visible deformation in the sensor with a similar force.

To see if this is indeed the case, and not only due to the sensor's visco-elastic effects, a separate experiment was conducted that looked at these effects by indenting the flat, 3D printed sample and keeping the sensor in place to allow for the visco-elastic relaxation to take place. For this experiment, the CNN was trained twice, once on the entire set and once on the set that excluded the visco-elastic relaxation. To ensure the CNN had the same amount of training data in both cases, the set with the visco-elastic effects was randomly sampled to match the baseline set. These results can be seen in figure 10, showing the section where the visco-elastic effects can be seen. It can be seen that the CNN is capable of learning the impact of visco-elastic relaxation to some degree. This shows that part of the increased error is due to the internal visco-elastic effects from the silicone samples.

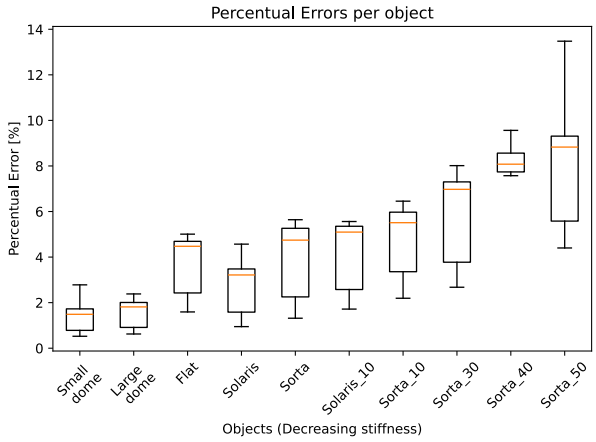

Figure 9. Analysis of the percentual error in pressure predictions across various test objects, organized by material hardness.

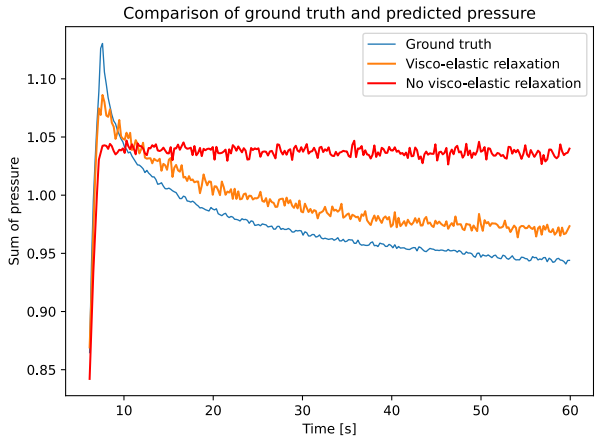

Figure 10. Comparison test of CNN results when trained with and without the effects of visco-elastic relaxation. Displaying the highlighted section where the visco-elastic relaxation takes place.

B. Temporal interpretation of pressure distribution

Given the reconstructed pressure distribution, an LSTM network is used to perform a classification and regression task to analyze its capabilities in handling objects with varying hardness levels.

Three reference samples were chosen to construct the classes containing the flat PLA plate (hard), Solaris silicone without thinner (medium), and Sorta-clear silicone mixed with 50% thinner (soft). All of the 3D-printed samples were used for the hard class. The remaining silicone samples were placed in one of the other two groups based on the reference sample with the closest Young's modulus. An overview of the categories can be seen in table I.

Hard	Medium	Soft
Flat	Solaris	Sorta 30%
Small Dome	Sorta	Sorta 40%
Large Dome	Solaris 10%	Sorta 50%
	Sorta 10%	

Table I
SAMPLES IN THEIR CORRESPONDING CLASSES

From the total of 10 samples, only the three reference samples were used in training. This allows the LSTM to learn the features of each of the classes. The remaining seven samples were used to validate the model, each with six indentation sequences. To reduce the effects of the visco-elastic behavior of both the sensor and the objects and reduce the complexity of the data, only the indentation part of the sequence was taken into account. The results from the CNN show lower errors on average for the indentation phase compared to the retraction phase (see Appendix B). This increases the model's accuracy and reduces the amount of data required, thus speeding up the entire prediction pipeline. A confusion matrix is constructed to show the results of this classification task. From this, we see an overall precision of 82%. These results show that the LSTM is capable of distinguishing objects' hardnesses. A more in-depth look at the results shows that the error for the class with the hardest samples exclusively occurs for the small dome sample, and the sample made from SORTA-clear silicone with 10% thinner accounts for the most errors in the class containing the medium samples. The results from the small dome show it estimates its stiffness to be lower than its actual class. This can be explained by the fact that this pressure distribution differs from that of the other samples due to its smaller contact area when indenting. This results in a pressure distribution that is more concentrated around the center of the sensor. The second set of errors can be explained by the fact that this sample has its Young's modulus almost exactly between that of the two reference samples. Small fluctuations in pressure distribution predictions cause the LSTM to estimate a mix of medium and soft from the different trials, resulting in a higher error rate.

While increasing the number of classes could help improve the accuracy of the model by better classifying objects with slimilar Young's modulus, a larger number of samples is required for the LSTM to learn each of the classes. To overcome this, a regression task is performed where the LSTM can learn the specific features to estimate the Young's modulus precisely,

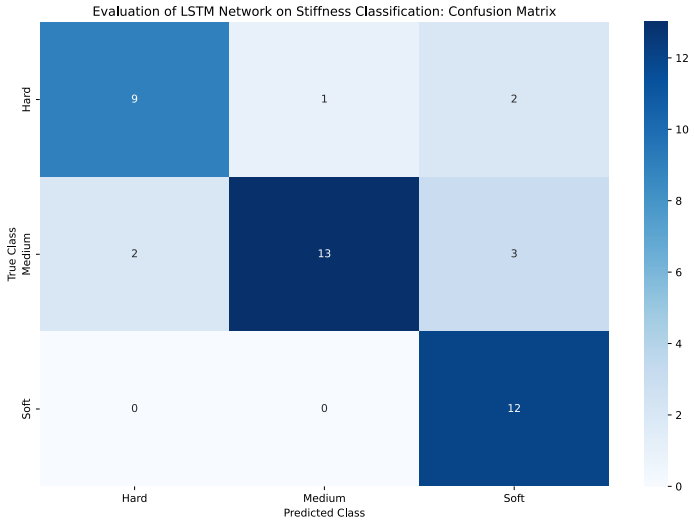

Figure 11. Confusion Matrix of LSTM-Based Stiffness Classification. Containing seven distinct objects with six trials each.

even if the sample has not been seen before. Intermediate results showed that including the PLA samples caused high erros in the estimations for the remaining samples. This error can be attributed to the LSTM estimating the softer objects much harder than their actual value. Thus, these were excluded from the regression task. The results for the remaining silicone samples showed an overall accuracy of 91.9%, with the results for each of the samples shown in figure 12.

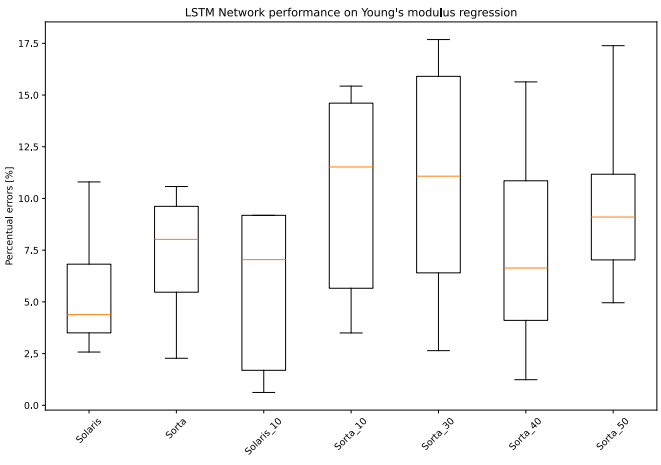

Figure 12. Results from the LSTM network's regression analysis, predicting the Young's modulus for the different silicone samples.

C. Estimation on perishable fruits

Using this trainend regression model, the LSTM was used model was used to evaluate the Young's modulus of ripe and overripe fruits, evaluating it robustness and real-life applications. The selected fruits were kiwis, avocados, and mandarins. These fruits have typical hardnesses during their ripening cycle that fall within the range of the silicone samples, allowing the LSTM model to estimate their Young's modulus. Furthermore,

these fruits show minimal visual cues to indicate their ripeness level, making them exceptionally suited for this application. One of each of the fruits was ripened for a week to ensure sufficient difference. Resulting in ripe and overripe fruits, with the overripe fruits having a lower Young's modulus. Measuring the ground truth value for the fruits proved difficult using a method similar to that used for the silicone samples due to irregularities in the shape and size of the fruits and local differences in softness. Human participants were asked to rank the fruits based on the perceived hardness using the silicone samples as a reference to overcome this issue. The fruits were assigned a number based on this comparison, and an equivalent Young's modulus was assigned to each of the fruits.

From the results of the LSTM, it can be seen that the model can clearly differentiate ripe and overripe fruits from each other, as shown in figure 13. The results from the human experiments show that the LSTM estimations are similar to those of the human participants (figure 14). The most significant differences can be observed with the mandarin samples. This difference can be attributed to it being the most deformable of the fruits, showing irregular deformation patterns. Compared to the other fruits, mandarins exhibit significant and uneven deformations when subjected to indentation. This introduces complexities in accurately capturing and predicting the pressure distribution over time and, consequently, Young's modulus. Due to the exploratory behavior of human palpation, these factors can be averaged out more easily.

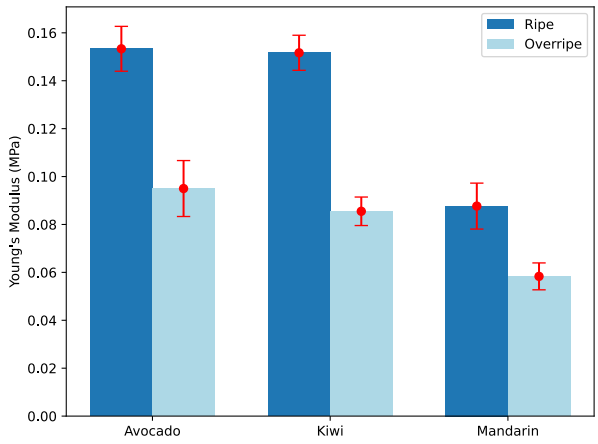

Figure 13. Estimation of Young's modulus for different fruits, comparing predictions for ripe versus overripe samples using the LSTM model..

These results demonstrate the efficacy of the proposed framework in accurately predicting pressure distributions and material softness, as well as determining the ripeness levels of several fruits that are particularly difficult to differentiate based on external properties. With particularly good accuracy for harder materials, while also revealing important challenges with softer, visco-elastic materials, showing the framework's potential and areas for future improvements.

VI. Discussion

The primary objective of this research was to develop a tactile perception framework that leverages high-resolution

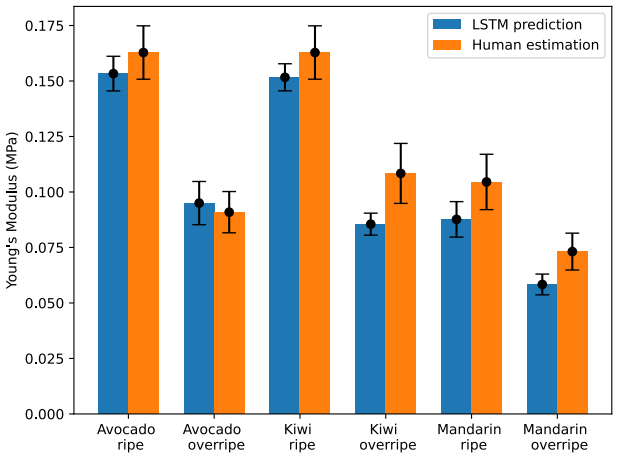

Figure 14. Comparison between LSTM-based predictions and human tactile assessments for estimating the firmness of fruit samples, presented as equivalent Young's modulus values.

tactile images from the ChromaTouch sensor to overcome the limitations of existing methods in object softness estimation. Existing approaches in this field, particularly those using end-to-end machine learning models, often suffer from the "black-box" problem, where no intermediate steps between input and output exist, making it challenging to interpret and evaluate the results. By contrast, this research introduces a framework that not only predicts the spatial pressure distribution using a CNN but also analyzes its temporal change using pressure sequences with an LSTM network to estimate object softness, thus providing greater transparency and interoperability.

To address these issues, the thesis proposes a more interpretable framework by incorporating intermediate steps based on physical principles. Instead of relying solely on an end-to-end approach where the input tactile images are directly mapped to an output softness value, the thesis introduces a hybrid model that combines CNNs and LSTMs with a more structured approach to feature extraction. The CNN is first used to predict the spatial pressure distribution from the tactile images, a step that is directly interpretable in terms of physical pressure patterns. This pressure distribution serves as a meaningful intermediate representation that makes the model's behavior more understandable and allows for easier identification of where the model might be going wrong. The research introduces an additional layer of verification by ensuring that the predicted pressure distributions and the resulting softness estimations are consistent with a well-established physical model. This means that even if the neural network behaves as a black box in some respects, its outputs can still be cross-checked against a transparent, physics-based framework. This hybrid approach significantly mitigates the black-box problem by embedding the neural network's predictions within a structure of known physical behavior, making it easier to diagnose and correct errors when they occur. Inaccuracies in the pressure and softness predictions by the CNN and LSTM can indicate difficult-to-predict indentation patterns and provide information on which input features might need to be refined or where additional training data might be required.

The CNN accurately predicted pressure distributions, particularly for harder materials. However, the framework highlighted a significant challenge with softer materials, where visco-elastic effects introduced difficulties. In contrast to current methods, without any intermediate steps, the early identification of the errors in the pressure distribution predictions makes it possible to mitigate the source of these errors more easily.

Several key findings were made using the LSTM network to analyze the temporal changes in the pressure distributions. Where previous studies typically only focused on either classification or regression, this study employed both, each with its benefits and downsides. It was shown that classification can be used to classify objects with varying softness levels. On the other hand, the regression was shown to be most useful in the case of samples with similar levels of softness, allowing for slight differences to be identified. However, the challenges observed with shape-dependent pressure distributions, as seen with the small 3D-printed samples and mandarins, suggest that further research is needed to generalize this method to a wider variety of objects and shapes.

Overall, this research represents a meaningful improvement over existing methods by integrating human-inspired tactile processing with advanced machine-learning techniques. It addresses the shortcomings of previous models, particularly in terms of interpretability and identifying important causes of errors such as visco-elasticity and shape-dependency, paving the way for more accurate and reliable softness estimation in both controlled and real-world environments.

A. Future research

Building on the findings of this research and addressing its limitations, several recommendations for future research are made. Improving the design of the ChromaTouch sensor by selecting materials that better accommodate visco-elastic effects is crucial. The current sensor material introduces significant visco-elastic properties, affecting the accuracy of reconstructing the pressure distribution estimation. By choosing materials with reduced visco-elastic properties, the sensor's performance can be enhanced.

Several other improvements can be made to increase the sensor's capabilities further. Increasing the number of trackable markers or redesigning the markers to enhance visibility under deformation can further mitigate the impact of visco-elasticity. A greater number of markers or markers that show more pronounced movement under stress would provide richer data for the neural network, allowing it to detect smaller changes with higher accuracy. This would improve the system's overall performance for both visco-elastic and non-visco-elastic objects. Apart from this, by implementing a higher sensing resolution, the pressure distribution can better show small local changes, making it possible to detect non-uniform deformation patterns better.

Extensive testing in real-world applications is essential to further validate the practical uses of the developed framework. This could involve deploying the sensor system in various robotic manipulation tasks across different scenarios. The current framework is optimally suited for ripeness detection during fruit handling, applying capabilities to differentiate ripe from overripe fruits. By increasing the number of samples used for training, thereby introducing a wider range of Young's moduli and shapes, several of the most significant areas where errors occur can be removed. Additional testing would demonstrate the robustness and adaptability of the framework in practical, uncontrolled environments.

While this research has made worthwhile contributions in the field of tactile sensing and object softness estimation, there is still room for further research and innovation. By addressing the identified challenges and pursuing the recommended future directions, the development of highly accurate and reliable tactile sensors can be accelerated, ultimately enhancing the capabilities of robotic systems in handling a diverse range of materials.

VII. Conclusions

This research successfully demonstrated the potential of using a combination of CNN and LSTM networks to interpret high-resolution tactile images for estimating object softness. The results indicate that while the current framework is effective, especially for harder materials, it faces challenges with softer materials due to their complex visco-elastic properties. These findings highlight the need for further advancements in both the hardware and algorithmic aspects of tactile sensing.

The key contributions of this work are as follows: This study validated the integration of a CNN for spatial pressure distribution prediction and an LSTM for temporal sequence analysis in the context of tactile sensing. Furthermore, the research identified the visco-elastic properties of materials as a significant source of error, particularly in softer objects, guiding future research toward incorporating these properties into tactile sensing models and finding methods more robust to these effects.

It was shown that the LSTM network can classify objects, determine an object's Young's modulus, and detect small variations between different objects. However, additional research is required to classify a broader range of objects and overcome the difficulty of shape dependency.

These findings show that this sensor cannot only be used to estimate the frictional safety margin and the reconstruction of deformation patterns as done by previous studies with the ChromaTouch but also allows for the extraction of pressure distribution and stiffness, opening up a wide range of future possibilities using the ChromaTouch. This research not only advances the capabilities of tactile sensing technology but also paves the way for more intuitive, versatile, and robust robotic systems, capable of interacting with their environments in ways that closely mimic human touch, thereby opening up new ways to apply tactile sensing across diverse fields such as healthcare, agriculture, and manufacturing

VIII. Acknowledgement

I would like to express my gratitude to my supervisors, Michaël Wiertlewski and Giuseppe Vitrani, for introducing me to the subject of tactile sensing and creating this project

together with me. Their expertise in the field and continued supervision have been invaluable in shaping this work and allowing me to achieve these results. Furthermore, I would like to give my thanks to the other PhD and Master students from the Human-Robot Interaction department for their feedback and critical thinking during our weekly sessions. Finally, I would like to thank Dirk-Jan Boonstra and Max Polak whose previous work on the ChromaTouch sensor provided invaluable background on the sensor and image processing techniques.

REFERENCES

[1] Wouter M. Bergmann Tiest and Astrid M. L. Kappers. Cues for haptic perception of compliance. *IEEE Transactions on Haptics*, 2(4):189–199, 2009.

[2] A. Bicchi, E.P. Scilingo, and D. De Rossi. Haptic discrimination of softness in teleoperation: the role of the contact area spread rate. *IEEE Transactions on Robotics and Automation*, 16(5):496–504, 2000.

[3] D. J. Boonstra. Learning to estimate the proximity of slip using high-resolution tactile sensing. Master Thesis, Delft University of Technology, 2022.

[4] Yaohui Chen, Jiahao Lin, Xuan Du, Bin Fang, Fuchun Sun, and Shanjun Li. Non-destructive fruit firmness evaluation using vision-based tactile information. pages 2303–2309. Institute of Electrical and Electronics Engineers Inc., 2022.

[5] Sachin Chitta, Matthew Piccoli, and Jürgen Sturm. Tactile object class and internal state recognition for mobile manipulation. In *2010 IEEE International Conference on Robotics and Automation*, pages 2342–2348, 2010.

[6] Charles Dhong, Rachel Miller, Nicholas B. Root, Sumit Gupta, Laure V. Kayser, Cody W. Carpenter, Kenneth J. Loh, Vilayanur S. Ramachandran, and Darren J. Lipomi. Role of indentation depth and contact area on human perception of softness for haptic interfaces. *Science Advances*, 5(8):eaaw8845, 2019.

[7] Steven Conrad Hauser and Gregory John Gerling. Force-rate cues reduce object deformation necessary to discriminate compliances harder than the skin. *IEEE Transactions on Haptics*, 11(2):232–240, 2018.

[8] Hertz. On the contact of elastic solids. *Crelle's Journal*, 92:156–171.

[9] Mohsen Hosseini, Siamak Najarian, Samira Motaghinasab, and Javad Dargahi. Detection of tumours using a computational tactile sensing approach. *International Journal of Medical Robotics and Computer Assisted Surgery*, 2:333–340, 2006.

[10] Xiaodan Jia, Rui Li, Mandayam A. Srinivasan, and Edward H. Adelson. Lump detection with a gelsight sensor. In *2013 World Haptics Conference (WHC)*, pages 175–179, 2013.

[11] Micah K. Johnson, Forrester Cole, Alvin Raj, and Edward H. Adelson. Microgeometry capture using an elastomeric sensor. *ACM Trans. Graph.*, 30(4), jul 2011.

[12] Ningning Ma, Xiangyu Zhang, Hai-Tao Zheng, and Jian Sun. Shufflenet v2: Practical guidelines for efficient cnn architecture design. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 116–131, 2018.

[13] Yulim Min, Yunjeong Kim, Hanbit Jin, and Hye Jin Kim. Intelligent gripper systems using air gap-controlled bimodal tactile sensors for deformable object classification. *Advanced Intelligent Systems*, 2023.

[14] Alessandro Moscatelli, Matteo Bianchi, Alessandro Serio, Alexander Terekhov, Vincent Hayward, Marc O. Ernst, and Antonio Bicchi. The change in fingertip contact area as a novel proprioceptive cue. *Current Biology*, 26(9):1159–1163, 2016.

[15] Tetsushi Nonaka, Arsen Abdulali, Chapa Sirithunge, Kieran Gilday, and Fumiya Iida. Soft robotic tactile perception of softer objects based on learning of spatiotemporal pressure patterns. In *2023 IEEE International Conference on Soft Robotics (RoboSoft)*, pages 1–7, 2023.

[16] M. E. A. Polak. Explainable neural networks for incipient slip sensing in robot tactile learning. Master Thesis, Delft University of Technology, 2023.

[17] Rob B. N. Scharff, Dirk-Jan Boonstra, Laurence Willemet, Xi Lin, and Michaël Wiertlewski. Rapid manufacturing of color-based hemispherical soft tactile fingertips. *CoRR*, abs/2112.05536, 2021.

[18] M. A. Srinivasan and R. H. LaMotte. Tactual discrimination of softness. *Journal of Neurophysiology*, 73(1):88–101, 1995. PMID: 7714593.

[19] Benjamin Ward-Cherrier, Nicholas Pestell, Luke Cramphorn, Benjamin Winstone, Maria Elena Giannaccini, Jonathan Rossiter, and Nathan F. Lepora. The tactip family: Soft optical tactile sensors with 3d-printed biomimetic morphologies. *Soft Robotics*, 5:216–227, 4 2018.

[20] Wenzhen Yuan, Chenzhuo Zhu, Andrew Owens, Mandayam A. Srinivasan, and Edward H. Adelson. Shape-independent hardness estimation using deep learning and a gelsight tactile sensor. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 951–958, 2017.

[21] Yuhua Zhang, Yuka Mukaibo, and Takashi Maeno. A multi-purpose tactile sensor inspired by human finger for texture and tissue stiffness detection. In *2006 IEEE International Conference on Robotics and Biomimetics*, pages 159–164, 2006.

APPENDIX

12 Appendix A: Example sub-image for three different sizes.

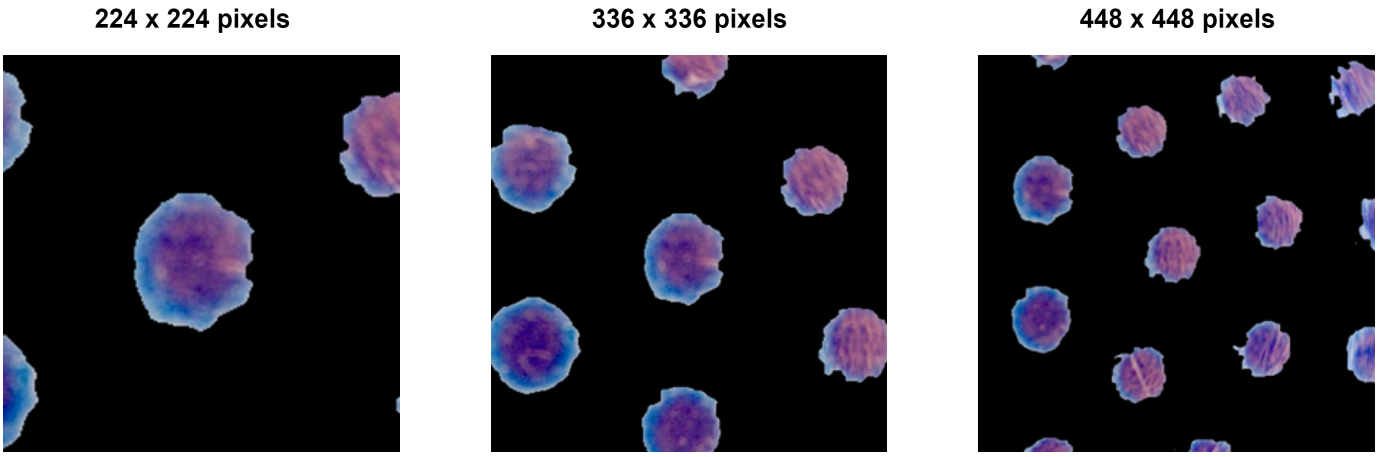

Figure 15. Three sub-images around the same marker for different window sizes

12 Appendix B: Percentual error comparison of indentation and retraction phase

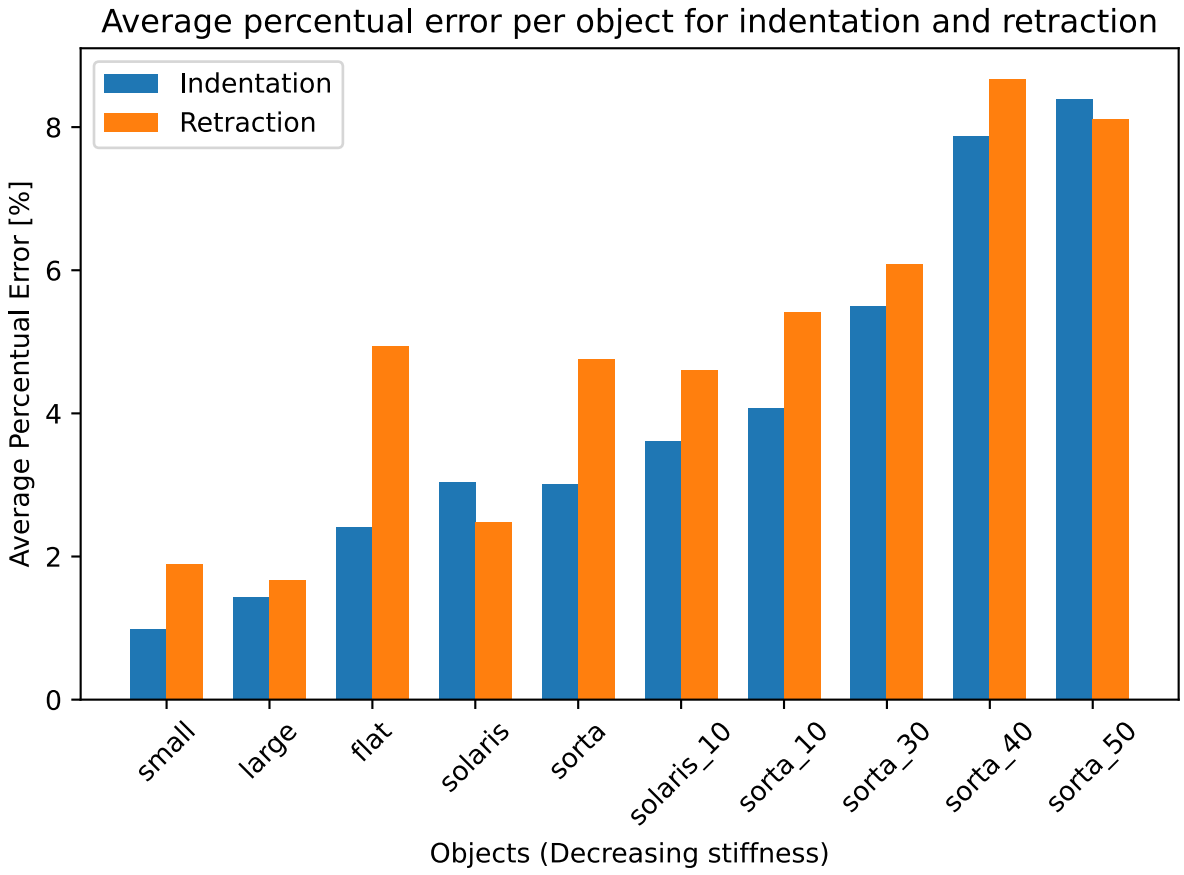

Figure 16. Average percentual error of 6 trials of the indentation and reaction profiles.