The University of Hong Kong

Department of Computer Science

Final Year Project

Cloth Quality Checking Using Tactile Sensor

Final Report

Name:	Lau Kwai Lam
UID:	3035572856
Course Code:	COMP4801
Supervisor:	Professor Pan Jia
Date:	April 21, 2022

Abstract

In view of the fall behind technology in the cloth quality checking process nowadays, this project aims to develop a Tactile Vision combined model to achieve an objective unmanned cloth quality checking technology. Fabric type classification (task A) and rapid defect localization (task B) are the two objectives of this project. To achieve it, the LSTM model with a magnetic tactile sensor is applied in task A, and Vision Tactile combined CNN model with an optical Gelsight sensor is applied in task B. The result of both tasks is satisfactory and as expected. The testing accuracy of 30 types of fabric is 100% in task A. For task B, 6 common defects suggested by experienced tailors are mimicked. More than 6000 data are used to train the Vision CNN model. As a result, by combining the prediction of the Tactile CNN model, the testing accuracy is increased by 1.4% in task B. The testing accuracy of the combined model is 95.8% and 14.4 times faster than the Vision CNN model. Furthermore, the factors affecting the testing accuracy are studied in this paper by repeating the data collection process by hand.

Acknowledgement

This project is supervised by professor Pan and his PHD student Ruixing. Thank you professor Pan for guiding the direction of this project. Thank you Ruixing for passing on the mechanics knowledge and many technical advice.

Table of Content

List of Figure and Table

1 Introduction

The introduction of this project consists of three parts. First, the technical problem which introduces the current situation of quality checking in the textile industry. Second, the two main objectives of this project. Third, the introduction of the magnetic tactile sensor, the LSTM model, the Gelsight sensor and The CNN model and how they relate to this project.

1.1 Technical Problem

The technology of cloth quality checking is falling behind nowadays. Tailors adopt the “Uniformity test” and “Light Test” to check the quality of the fabric. For the “Uniformity test”, the tailor determines the quality of fabric by stroking the fabric with his palm. For the “Light test”, the tailor put the sample of fabric under the light to find the defect of the fabric. However, for both methods, the results are subjective and tailor experience-dependent. Furthermore, only part of the sample fabric is checked due to the limitation of human resources.

On the other hand, the magnetic tactile sensor with the LSTM model and the Gelsight sensor with the CNN model provides a digital, objective criterion for the quality of the fabric. Moreover, the design of the magnetic tactile sensor allows it to be enlarged easily. Thus, checking all the fabric manufactured and not limited by the fabric size. This is the reason why this project is applying the tactile sensor and machine learning model to enhance the cloth quality checking process.

1.2 Objective

The objective of this project is to develop various machine learning models to provide a fast and accurate prediction of defective fabric. Such that the tactile sensor with the ML model is capable of replacing the existing fall behind cloth quality checking technology in the textile industry. There are two sub-objectives in this project. The two sub-objectives and their corresponding assessment criteria are as follows.

1.2.1 Fabric Type Classification (Task A)

For this task, the goal is to identify 30 common types of fabric on the market with testing accuracy higher than 99%. For example, the LSTM model should be able to classify whether the sample fabric is cotton, wool, linen, or silk.

In addition, the factor in the data collection process affecting the testing accuracy score is studied in this task.

1.2.2 Rapid Defect Localization (Task B)

The aim of this task is to localize the defect in the fabric as soon as possible by combining tactile data and vision data. The tactile sensor only needs to visit the uncertain areas on the fabric determined by the vision data to speed up the progress. The goal for the combined model is to improve the accuracy score by at least 1% and 10 times faster.

At the same time, the combined model should be able to identify the defect type. For example, the model should be able to tell whether the defect of the fabric is caused by a grease spot or burning in that particular location.

1.3 Scope

There are two main kinds of tactile sensors, namely, “Optical Gelsight Sensor” and “Magnetic Tactile Sensor”. The magnetic tactile sensor developed by professor Pan and his PhD student at HKU is the chosen sensor for the fabric type classification task. The major reason is the simplicity of the hardware design. It provides better malleability, durability, and lower production cost.

Moreover, three-dimensional data is captured by the magnetic tactile sensor simultaneously, it provides vertical and horizontal information about the fabric. That is why the magnetic tactile sensor is a better option for fabric type classification tasks consisting of sliding operations.

On the other hand, the Gelsight sensor is first introduced in [7] in 2009. It is one of the optical-based tactile sensors such as [10], [11] and [12]. The sensor is described to be capable to capture high-resolution reconstruction of contact geometry. Therefore, it is a better choice for rapid defect localization consisting of defect shape reconstruction.

1.4 Magnetic Tactile Sensor

The magnetic tactile sensor comprises three layers, a flexible magnet as the upper part, a silicone elastomer as the middle part and a printed circuit board (PCB) with 9 magnetic sensors as the lower layer. With the contribution of Halbach array technology, the magnet field affects the elastomer and PCB only but does not affect the object above.

Figure 1a demonstrates the circumstance when a normal force is applied to the magnetic tactile sensor. As shown in this figure, the flexible magnet deformation causes the deform of the magnetic field in the elastomer. Thus, affecting the reading on the magnetic sensor in the PCB. The normal force is one of the major tactile data for training the LSTM model. The normal force reveals the hardness and roughness of the fabric which is one of the major features to classify the fabric. Thus, normal force data captured by the magnetic tactile sensor is crucial for the fabric classification task.

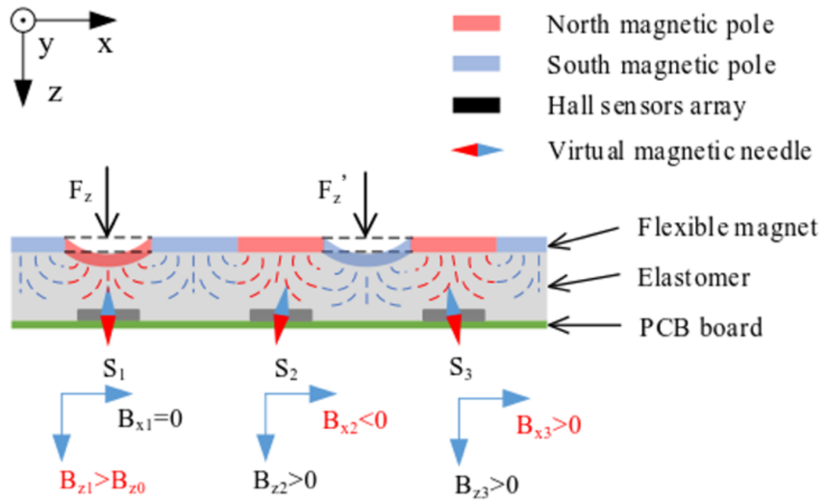


Figure 1a, Illustration of working principle of magnetic tactile sensor under normal force

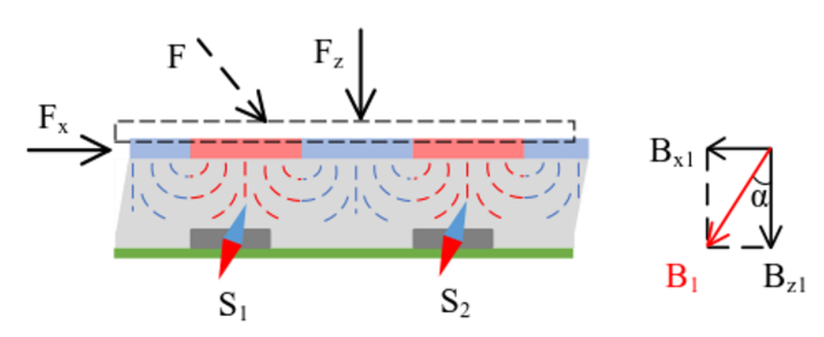


Figure 1b, Illustration of working principle of magnetic tactile sensor under shearing force

Figure 1b illustrates the circumstance when a shearing force is applied to the magnetic tactile sensor. By analyzing the alpha angle shown in this figure, the incoming force (F) is derived into normal force (Fz) and shearing force (Fx). The elasticity, static friction and kinetic friction of the fabric are highly reflected in the shearing force data (Fx). Therefore, the shearing force is another major criterion to identify the type of fabric. Thus, the magnetic tactile sensor is applied in the fabric classification task, as shown in figure 1c.

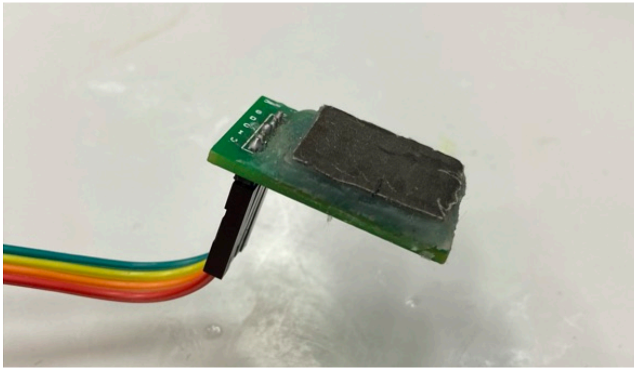


Figure 1c, The magnetic tactile sensor applied in Task A

1.5 Long Short Term Memory

LSTM is a special version of a recurrent neural network (RNN). By adding "forgot gate layer" and "input gate layer" to update the cell state, it is able to remember the important information only.

Figure 2 shows the architecture of an unrolled LSTM model with the detailed structure of an LSTM cell. The sigmoid layer on the left is the forget gate layer, it contributes to forgetting inconsequential information. The sigmoid and tanh function in the middle is the "input gate layer". This layer determines whether the input is significant enough to save in the cell state.

The LSTM model is suitable for the fabric type classification task because of its unique features. The continuous input of this task is the tactile data, the continuous output of this project is the probability of each type of fabric. The LSTM cell saves the significant feature of the fabric and forgets the irrelevant data to reinforce the prediction performance.

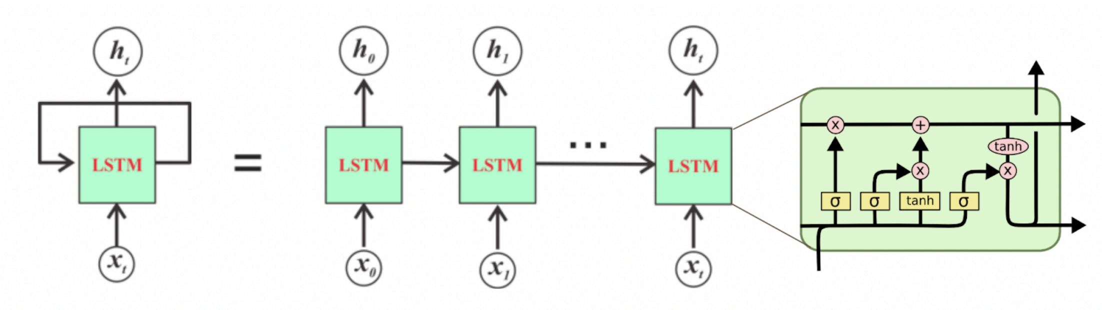


Figure 2, The architecture of an unrolled LSTM model with the detail of LSTM cell [2]

1.6 The Optical Gelsight Sensor

Besides of magnetic tactile sensor, the optical Gelsight sensor is one of the preeminent tactile sensors nowadays, which provides high-resolution tactile data. The working principle and current applications are introduced in this part.

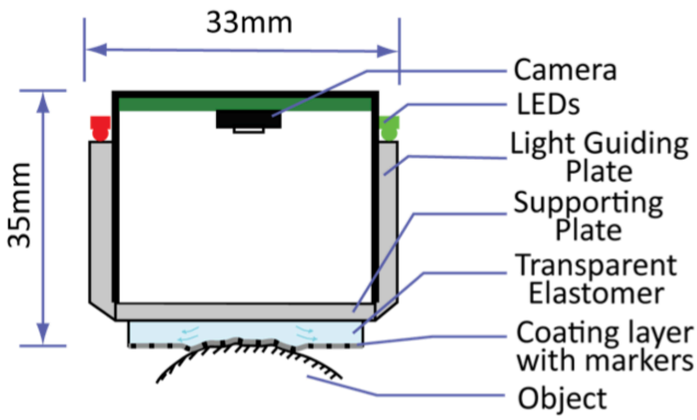


Figure 3a, The schematic of the Finger Tip Gelsight sensor [1]

Figure 3a illustrates the schematic of a Finger Tip Gelsight sensor, which mainly consists of an embedded camera, reflective membrane and illumination system [8]. The elastomer is covered with a coating layer with markers. Such that when the object has contact with the coating layer, the markers are deformed base on the surface texture of the object. Then, the embedded camera captures the deformations of the markers. Therefore, a high-resolution image is achieved, and it can be accurate within microns [9].

Gelsight sensors have been adopted for various applications in the past decade. For instance, Jia, et al used a Gelsight sensor to detect the hard lumps in soft tissues [13] and found that the accuracy is higher than by humans. Li illustrated a Gelsight sensor which is small enough in size to mount on a robot fingertip [14]. The three-dimensional shape of the object is reconstructed by Nicola J. Ferrier et al with a reconstruction algorithm [10]. The GelForce sensor is introduced by K. Sato et al [15]. It is capable to capture the force vector field of the deformation. Furthermore, it can also be used to perform slip detection [16] or hardness estimation [17].

The Gelsight sensor is suitable for the rapid defect localization task of this project, attributed to its high-resolution feature and shape reconstruction ability. The rapid defect localization task requires the tactile data to show the hardness, geometry, texture, and contour of the defect to classify the defect. Therefore, the optical Gelsight sensor shown in figure 3b is applied in this task.

Figure 3b, The Gelsight sensor applied in task B

1.7 Convolution Neural Network

Convolution Neural Network (CNN) is shown to outperform all other algorithms in dealing with high-dimensional and complex 2D images. For instance, LeNet-5 is one of the most influential CNN models developed in 1998 [19]. It consists of 7 layers without considering the input layer as shown in figure 4. The input is a 32*32 pixel image. It is normalized with mean 0 and variance 1 since the normalized image accelerates the learning process [20].

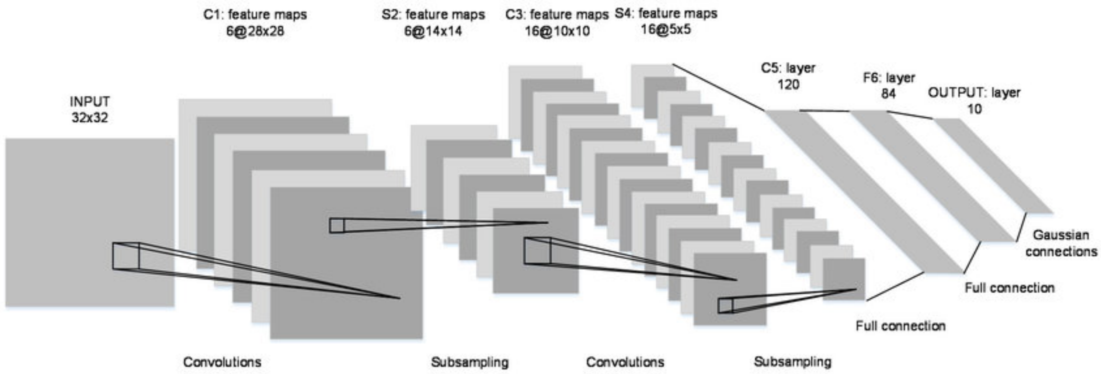


Figure 4, The architecture of LeNet-5 CNN [19]

The LeNet-5 example shown in figure 4 is designed to analyze the MNIST dataset. Therefore, the input size is 32*32 and the image is in greyscale. It has two convolution layers with stride 1 and a kernel size 5. Followed by an average pooling layer, two fully connected layers and an output layer with 10 classes.

Due to its high performance in analyzing images, it is adopted in the rapid defect localization task for analyzing vision data and tactile data from the Gelsight sensor. However, some of the specifications are altered to satisfy the needs of this task. For example, the Gelsight sensor captures high-resolution images, therefore, the input size of the LeNet-5 is increased to 64*64. The colour of the tactile image is crucial for analysis, thus, the input channel is 3. Furthermore, the output dimension is set to 7 for this task, and a SoftMax layer is added to convert the output to probability.

Some studies show that the convolution LSTM model is used to analyze vision-tactile combined data, such as [16] and [17]. The main idea is to use the output of the fully connected layer in CNN as the input of the LSTM model. LSTM has eminent performance in continuous input, however, continuous input is only applicable for the fabric classification tasks. And the magnetic tactile sensor does not capture images. Therefore, the convolution LSTM model is not suitable for this task. For the rapid defect localization task, the input is discrete images, therefore, CNN is more suitable for this task.

2 Methodology

After introducing the concept of the magnetic tactile sensor, the LSTM model, the Gelsight sensor, the CNN model and how they are related to this project. The following part mainly focuses on how to achieve the objectives. The overall architectural design and methodology for each task are as follows.

2.1 Architecture Design

Table 1 illustrates the general architectural design of the whole project from tactile data to GPU allocation. The architectural design includes four main segments from the external to internal, namely, data, data preprocessing, Machine learning model optimization and GPU.

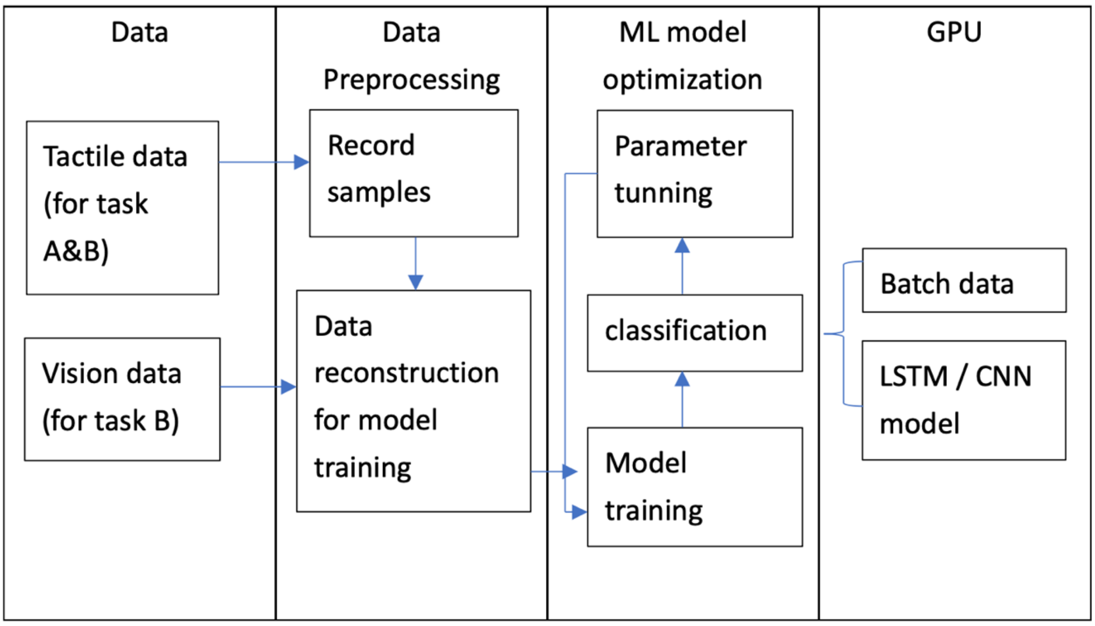


Table 1, The Architectural design from data to GPU

2.2 Hardware and Software Setup

Regarding the hardware part for the fabric type classification task, the magnetic tactile sensor is installed on the tip of the robot arm and connected with a 3D-printed joint. As illustrated in figure 5a, the fabric is fixed on a metal plate by tapes on four edges. Only four edges are fixed on the plate instead of the whole fabric, it ensures that the elasticity along each direction can be captured.

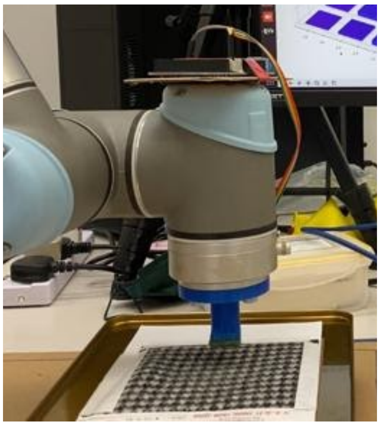


Figure 5a, Hardware setup of Task A

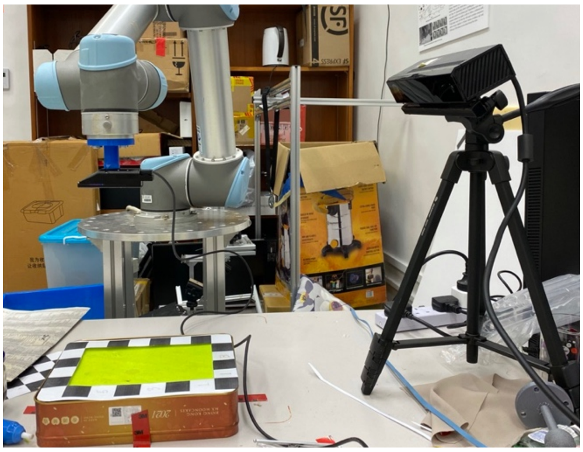


Figure 5b, Hardware setup of Task B

Figure 5c, Fabric before tailoring

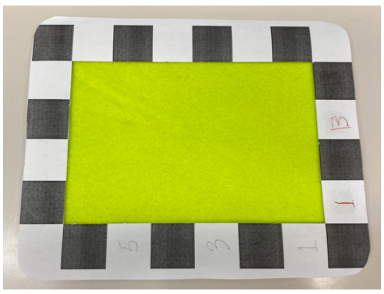


Figure 5d, Fabric divided into 24 grids

Regarding the hardware part for the rapid defect localization task, first, the fabric is tailored into 23*17 cm size. The tailoring process is shown in figure 5c. After that, a paper sign showing 24 grids is placed on top of the fabric, such that the grids are conspicuous during the scanning process. The paper sign is shown in figure 5d, the grid is 3*3 cm in size.

Second, a Microsoft Kinect Sensor V2 is placed behind the sample fabric to capture RBG photos as vision data. Figure 5b illustrates the position of the Kinect camera and the robot arm. The Gelsight sensor is installed on the tip of the robot arm and connected with a 3D-printed joint.

Regarding the software part, the Ubuntu operating system is adopted to provide a robot friendly environment. ROS is installed to manipulate the robot arm.

After understanding the hardware and software setup, the architecture and detailed process for both tasks are depicted step by step in the following.

2.3 Fabric Type Classification

1. Interview two experienced tailors to gain a better understanding of the current difficulties in the quality checking process of the textile industry.
2. The robot arm collects tactile data from the fabric in 3 different trajectories. 6 samples are recorded for each trajectory.
3. The data is reconstructed into 27 dimensions to fit the LSTM model.
4. The LSTM model training and parameter tuning. Keep repeating this step until the accuracy exceeds 99%.

5. Replace the robot arm by hand to repeat the process, to study the factor affecting the accuracy score.

The architecture of the fabric type classification task is relatively simple, it is shown in figure 6a.

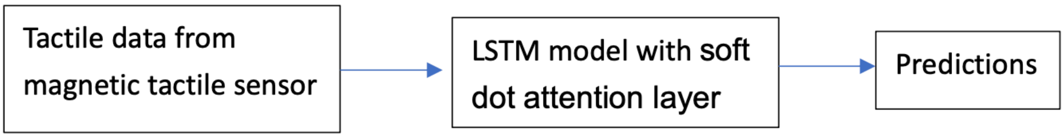


Figure 6a, the architecture of task A

2.4 Rapid Defect Localization

1. Define an area (12cm * 18cm) for sampling and divide the area into grids (4*6 grids), 3*3 cm in size for each grid.
2. Form common defects, such as water stains and grease spots, on the fabric and assign them with a unique label.
3. Capture the area with a Kinect camera and convert the image to grayscale. Provide each grid with a proper label.
4. Train a Vision CNN model to predict the defect for each grid and save the probability output from the SoftMax layer.
5. Calculate the mean probability of correct and incorrect predictions respectively. Define the vision threshold value to be the average of these two values. Such that the predicted probability below this value is treated as uncertain predictions by the vision CNN model.
6. Train a Tactile CNN model, estimate the prediction accuracy for each label and compare the performance to the Vision CNN model.

7. Visit the grid with a predicted probability below the vision threshold by the robot arm. For example, if the probability of the grid is below 0.96, it is difficult to identify the flaws solely by vision data. Therefore, tactile data is required.
8. Scan the uncertain grid by the Gelsight sensor. Predict the defect type using the tactile CNN model. Replace the vision CNN prediction with the tactile CNN prediction only if the tactile CNN model performs better than the vision CNN model in this label (measured in step 6).
9. Calculate the final accuracy with the combined prediction of both models.

The architecture of rapid defect localization task is shown in figure 6b.

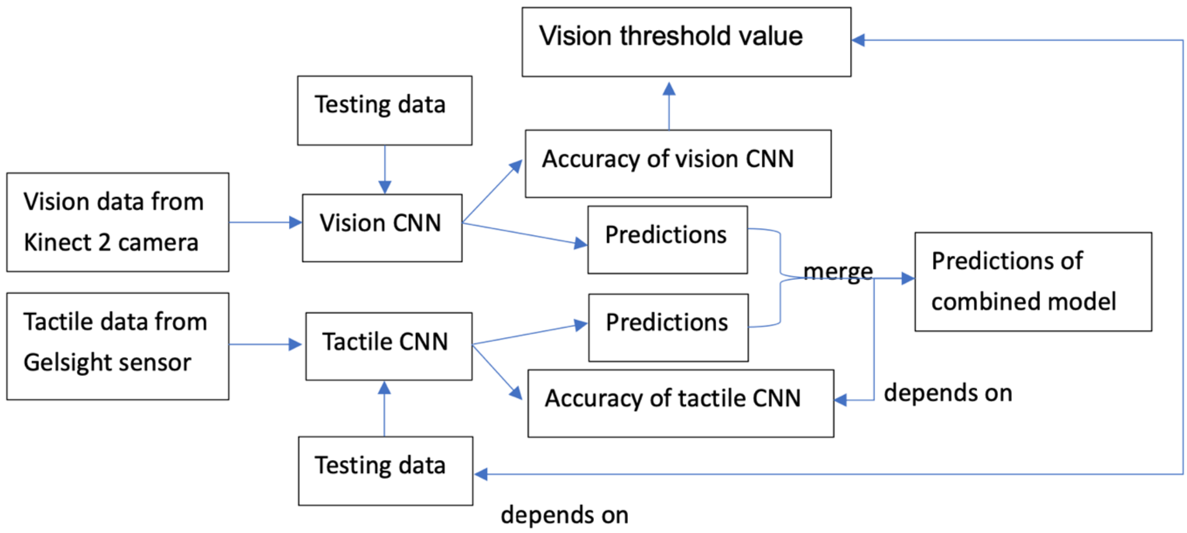


Figure 6b, the architecture of task B

3 The Progress and Result of Task A

After understanding the overall view of both tasks, this section focuses on the work completed. The assessment of the performance of the fabric type classification task is as follows.

3.1 Assessment of Tailor Interview

Mr Kwan is a tailor who worked in a weaving mill for over 30 years. Mr Kwan introduced the properties of the fabric in figure 7a. He demonstrated that the elasticity of fabric in each direction is different. For example, for most cotton, there is no elasticity in the vertical direction. The elasticity along the horizontal direction is weak, however, the elasticity in the oblique direction is strong.

The elasticity in each direction is a major feature of the fabric. Thus, this feature is captured in the fabric type classification task by different trajectories.

Figure 7a, Tailor interview (Mr. Kwan)

Figure 7b, Tailor interview (Mr. Wu)

Mr Wu is working as a tailor for more than 50 years. He is familiar with the quality of the fabric imported from all around the world.

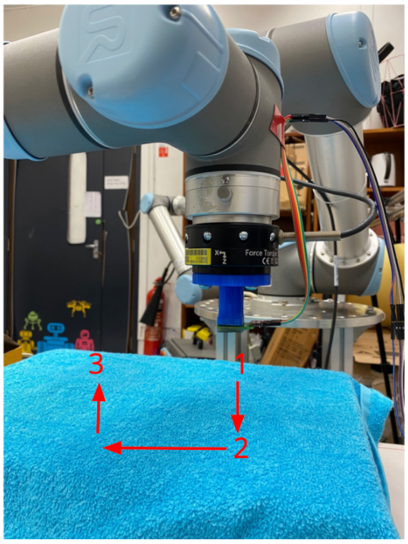
Figure 7b shows the interview process of Mr Wu. He mentioned that water stains and grease spots are the most common flaws on the fabric. It is made by the weaving machine and the improper storage process. All the fabric with water stains and grease spots cannot be used to tailoring clothes, therefore he returns all the defected fabric to the supplier.

Water stains, grease spots and other defects are formed on the fabric manually in the rapid defect localization task in order to simulate the reality.

3.2 Assessment of Hardware Setup and Data Collection

Figure 8a illustrates the moving path of the tactile sensor. The tactile sensor starts recording in position 1 and moves downward to position 2. It presses the fabric and stops for 0.8 seconds in position 2. After that, it slides leftward and upward to position 3. It stops the recording in position 3 and moves to position 1 of the next trajectory. The tactile sensor slides along the fabric at 10mm/sec.

Figure 8b shows the position of three trajectories. Each trajectory separates 120 degrees from the other. Three trajectories are used in this project instead



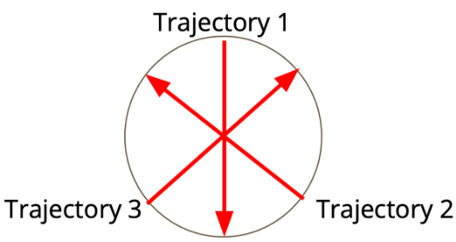


Figure 8b, The three trajectories

Figure 8a, The moving path of the tactile sensor of one trajectory in [3]. This is because three trajectories can capture the elasticity along each direction of the fabric. As mentioned by Mr Kwan, the elasticity along each direction of the fabric are different. The diameter of the circle in figure 4b is 50mm, thus the tactile sensor takes 5 seconds to finish the sliding operation on the fabric.

30 types of fabric are used in the fabric type classification task. They are weaved by fibre, wool, cotton and faux leather. They are weaved by different methods, thus, leading to different textures and patterns. Figure 8c shows the 30 different types of fabric used.

Figure 8c, 30 types of fabric used in task A

Since there are 9 sensors on the PCB, and for each sensor, there are three dimensions. Thus, there are 27 dimensions in total. The sensor is working at 10 Hz, and it takes 10 seconds to scan a sample. Thereby, the data size of each sample is 2700. For each trajectory, 6 samples are taken. For each fabric, there are three trajectories. And there are 30 types of fabric. Thus, there are 540 samples in total. Among them, 450 samples are used to train the model and the training data size is 1,215,000.

3.3 Assessment of The Graph

The result of the data collected is as follows. Few major features of the fabric can be told intuitively from the graph. Figure 9 shows the dynamic of magnetic flux density (uT) from the x, y, and z directions. This graph represents the sample of fabric type 9 collected via robot arm from sensor 0. Trajectory 3 is represented by the red arrow.

The sensor is moving down to the fabric before T0. The sensor is pressing the fabric in T0 - T1. The sensor retains in T1 - T2. After that, the sensor is trying to move but not actually sliding in T2 - T3. The Sensor is sliding in T3 -T4. Finally, the sensor moves upward after T4.

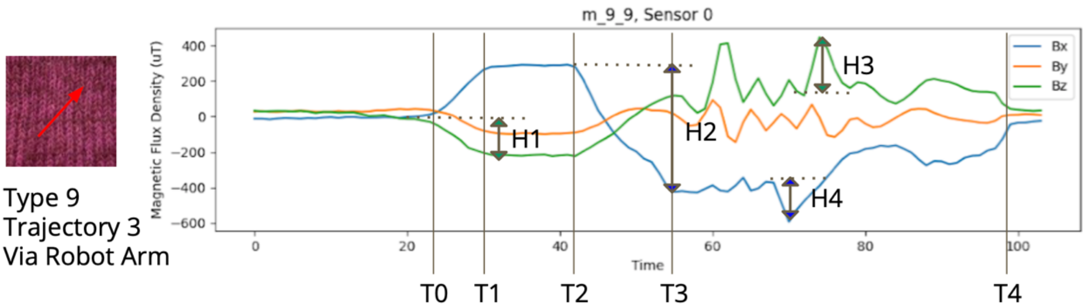


Figure 9, Magnetic Flux Density to Time Graph

The green line shows the magnetic flux density along the vertical direction (Bz) which represent the normal force. Therefore, H1 is directly proportional to the hardness of the fabric. H3 is directly proportional to the roughness of the fabric. The blue line shows the magnetic flux density along the horizontal direction (Bx) which represents the shearing force. Therefore, H2 is directly proportional to the static friction. H4 is directly proportional to the kinetic friction.

There is another indicator of the friction of the fabric. By observation, the vertical magnetic flux density (Bz) of the first row of the sensor along the moving direction is always positive, and the last row is always negative. The difference between these two rows is directly proportional to the friction of the fabric.

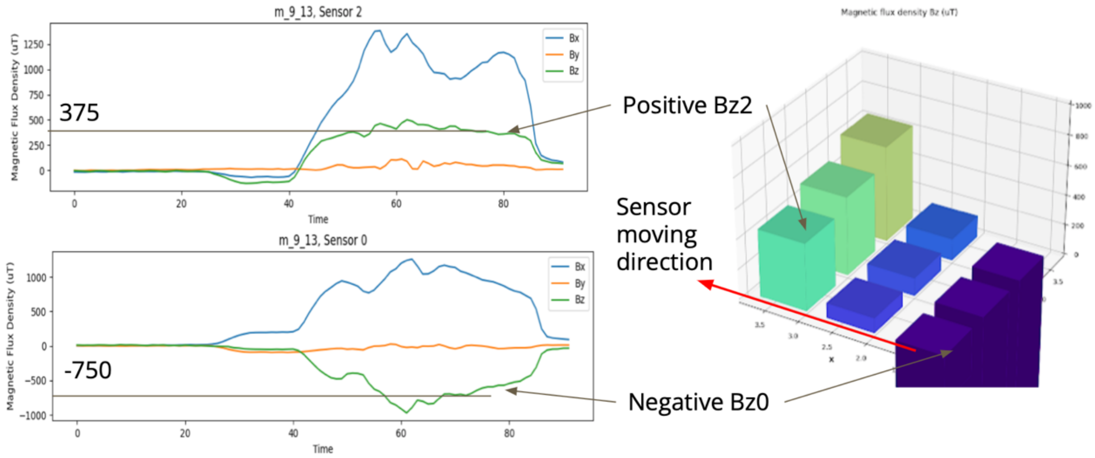


Figure 10, An example of the different between Bz2 and Bz0

Figure 10 shows an example of the different between Bz2 and Bz0. This is a sample of fabric type 9, trajectory 1 collected via robot arm from sensor 2 and 0. The different between Bz2 and Bz0 is 1125, it is directly proportional to the kinetic friction of the fabric.

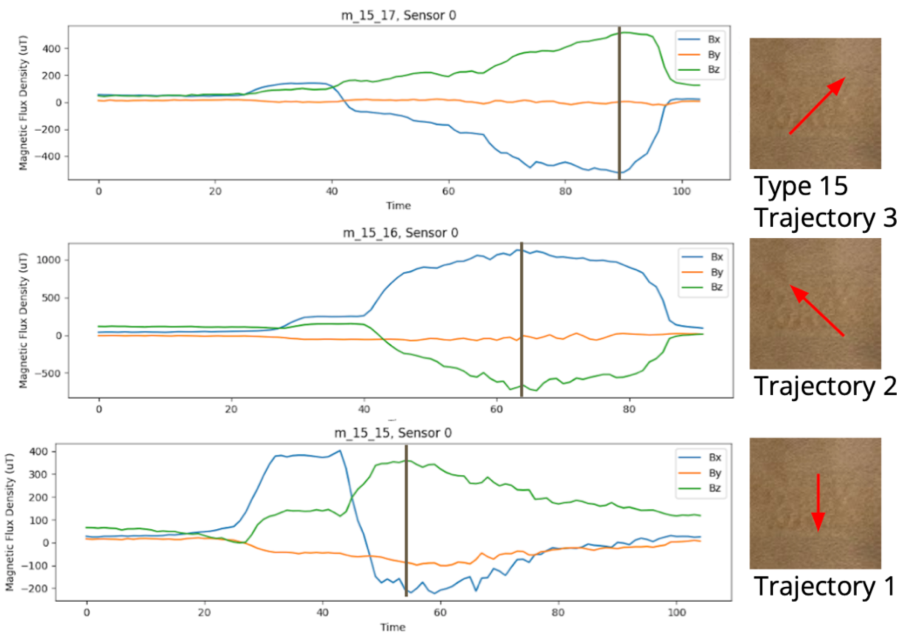


Figure 11, The peak of Bx in three trajectories

Moreover, the peak of the horizontal magnetic flux density (Bx) reveals the elasticity of the fabric in this direction. Figure 11 shows the peak of Bx from three samples from fabric type 15. Three samples are in different trajectories as shown in the red arrow. If the peak appears later, it implies that the elasticity along this direction is the largest and vice versa.

The postponed peak is caused by the accumulated rebounding force when the fabric is stretched. If the fabric can no longer be stretched, it starts to rebound, and the accumulated rebounding force is lost. It causes the peak to appear earlier. Thereby showing the elasticity along this direction is weaker.

3.4 Assessment of the LSTM Model

After the data collection process, 450 samples are used to train the LSTM model introduced in the introduction part. A soft dot attention layer is added to the LSTM model. The soft dot attention layer is proved to improve the testing accuracy by 5% in [3]. Learnable weights are added to the hidden state. Such that the LSTM model only focuses on the useful part of the tactile data and ignores the insignificant information.

3.5 Assessment of The Accuracy Result

<table>
<tr><td>Via Robot Arm</td><td>Via Hand</td><td>Via Robot Arm</td></tr>
<tr><td colspan="3">30 types of fabric</td></tr>
<tr><td colspan="2">3 trajectories</td><td>10 trajectories</td></tr>
<tr><td colspan="2">540 samples (450 for training and 90 for testing)</td><td>180 samples (150 for training and 30 for testing)</td></tr>
<tr><td>Training Accuracy: 100%</td><td>Training Accuracy: 100%</td><td>Training Accuracy: 84%</td></tr>
<tr><td>Testing Accuracy: 100%</td><td>Testing Accuracy: 88.9%</td><td>Testing Accuracy: 43.3%</td></tr>
</table>

Table 2, The accuracy result of Task A

Table 2 shows the accuracy result of 30 types of fabric. The training accuracy of data collected by robot arm and by hand are both 100% and the testing accuracy is 100% while the data is collected via robot arm, 88.9% via hand. Column 3 shows another set of data in contrast. It shows that if the trajectory is increased to 10, the testing trajectory have insufficient training trajectory to support. Two out of ten trajectory is scanned for the second time as the testing set. Therefore, the testing accuracy is very low (43.3%).

4 Findings of Task A

After demonstrating the accuracy result, this part focus on discussing the result and the factors affecting the accuracy score.

4.1 Train Test Data Ratio

For now, the testing accuracy of 30 types of fabric via robot arm is 100%. It is attributed to the good training and testing ratio. For each trajectory, the robot arm will scan it 6 times and one of them is used to test the model. In other words, for each testing sample, there is 5 training sample. As for the 10 trajectories case, for each testing sample, there is only 1 training sample, thus, leading to a low accuracy score.

4.2 Standardized Data Collection Process

The robot arm provides a stable and standardized environment for the data collection process which is essential for a high accuracy score. The magnitude of the pressing force, the direction and the retention time can only be standardized by using a robot arm. By contrast, the magnitude of normal force and shearing force cannot be standardized easily by hand.

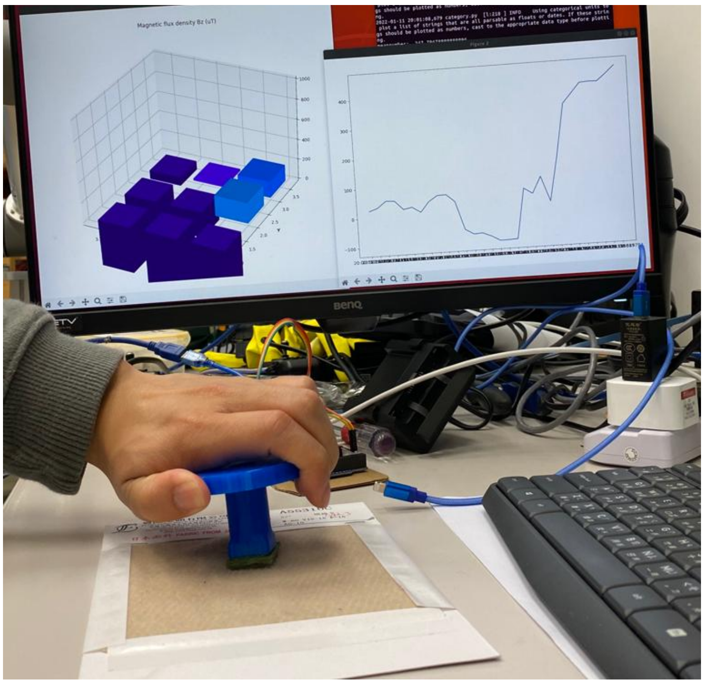


Figure 12, The instability of data collected by hand

Figure 12 shows the data collection process via hand. As shown in the figure, it is difficult to maintain the direction and magnitude of force for over 540 samples. In addition, it is impossible to stop for 0.8 seconds after pressing the fabric by hand. Thereby the testing accuracy score is relatively lower (88.9%).

5 The Progress and Results of Task B

The cloth quality checking process does not only consist of fabric type classification. To better describe the quality of the cloth, the defects on the fabric should be found and located rapidly. Therefore, the assessment of the performance of the rapid defect localization task is as follows.

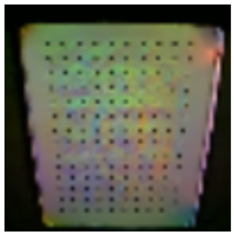
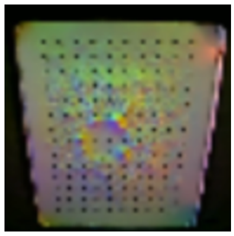
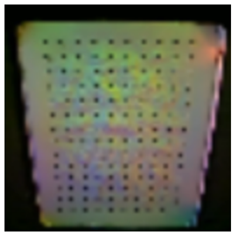
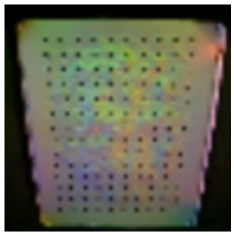
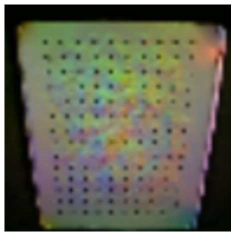
5.1 Assessment of Hardware Setup and Data Collection

Six common defects are adopted and stimulated in this task. Their label, name and stimulation method are shown in table 3. For the grid which has no defect, the label is 0. Water stain and grease spots are suggested by Mr. Wu.

Labels	Defect Name	Stimulation Method
0	No defect	
1	Hole	Rub a hole with a sandpaper
2	Water Stain	4 drops of the water-sauce mixture at a ratio of 10:1
3	Grease spot	1 spray of WD-40 Smart Straw
4	Drop stiches	3 slices with a cutter
5	Worn out	Rub slightly with a sandpaper
6	Misprinting	4 drops of AA - Aron Alpha

Table 3, The label, defect name and stimulation method of task B

The simulation results are shown in table 4. It shows the image of vision data and tactile data for each defect and their occurrence in the training dataset.

Label	Occurrence in training dataset	Defect Name	Vision Image	Tactile Image
0	310	No defect		
1	300	Hole		
2	300	Water Stain		
3	310	Grease spot		
4	300	Drop stiches		

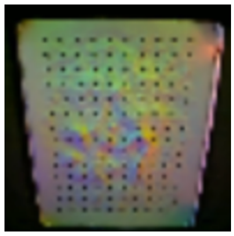
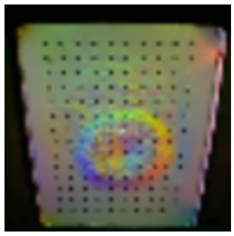
5	300	Worn out		
6	300	Misprinting		

Table 4, The image of vision and tactile data

Figure 13 shows the transformation process of vision data. This is an example of vision data of fabric sample 1. Grid 3 is first turned into greyscale at a size 64*64, to minimize the illumination effect. After that, it goes through a transformation process including rotation, flipping (both horizontal and vertical) and perspective distortion. Such that the distortion and location effect of the Kinect camera is minimized.

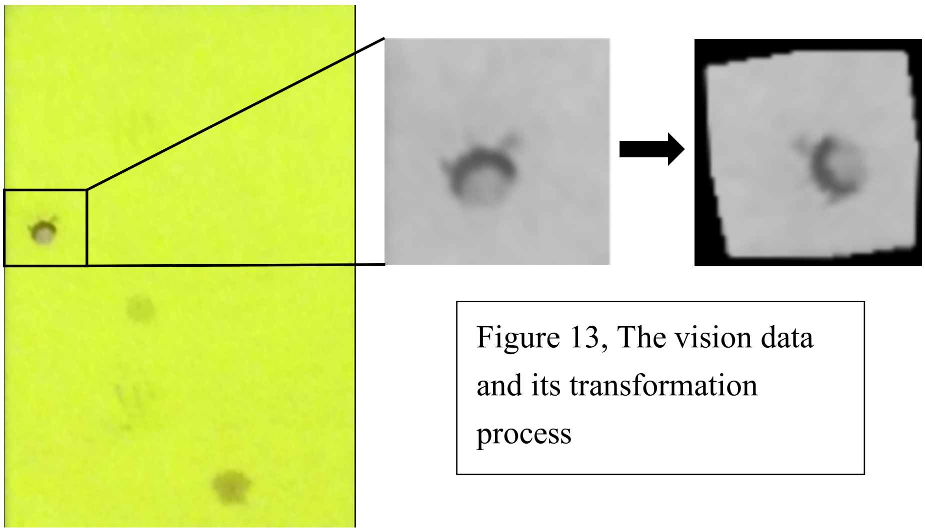


Figure 13, The vision data and its transformation process

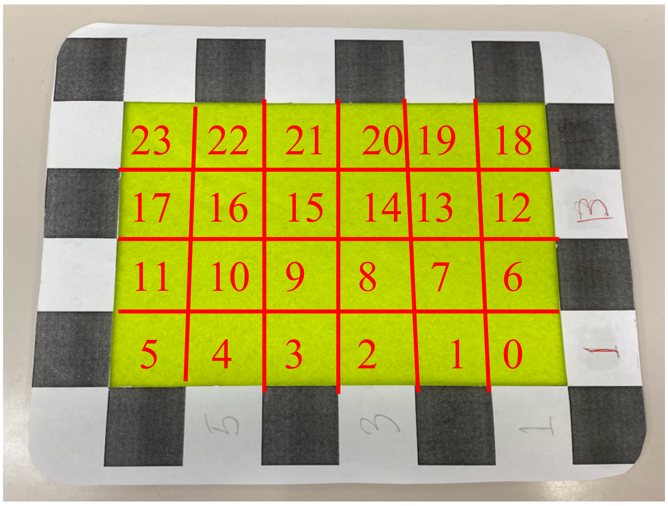


Figure 14, The moving path of the Gelsight sensor

As for the tactile data, the moving path of the Gelsight sensor is shown in figure 14. It first moves to a location above the 0 grid. Then, it moves downward and presses the fabric for a second. During pressing, the tactile image is captured by the Gelsight sensor. After that, it moves to the location above the 1 grid and continues the process until all grids are captured for the training dataset.

There are 30 defective fabric samples in total, 27 of them are used to form the training dataset and 3 of them are testing datasets. For each fabric, 10 vision photo is taken by the Kinect camera. Since there are 24 grids for each sample. Therefore, the total training data size of vision data is 6480 and the total training data size of tactile data is 648. Some of the image with label 0 is dropped in the training dataset of the Vision CNN model to maintain the balance for each label, as shown in table 4.

Only one photo is taken for the testing dataset, thus, the vision testing data size is 72. As for the tactile testing data size, it is expected to be 72 too, however, the actual testing data size depends on the vision threshold value. Therefore, it is reduced significantly.

5.2 Assessment of the Vision CNN Model

After transformation, 6480 photos and their label are trained in a Vision CNN model. The batch size is 30, with 1000 epochs and a 0.0005 learning rate. Since the photo is converted to greyscale, the input channel is 1. The training loss function is shown in figure 15.

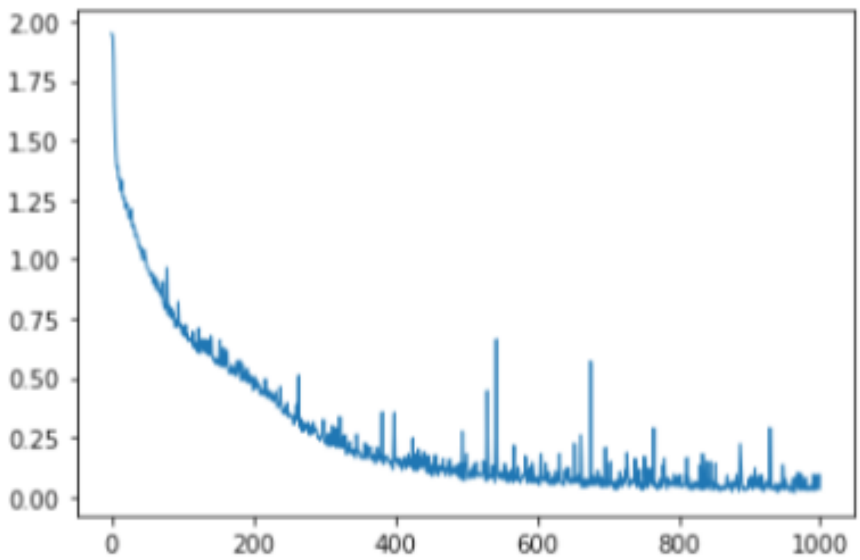


Figure 15, training loss per epoch of Vision CNN model

The testing accuracy on 72 photos is 94.4%. Table 5 shows the accuracy of each prediction. For defects 0,1 and 5, vision CNN model has 100% testing accuracy. But for defects 3 and 6 the testing accuracy is 75% which reveals that, it is difficult to identify grease spots and misprinting only by vision data.

Vision CNN Model		Label							Accuracy (%)
		0	1	2	3	4	5	6	
Prediction	0	48	0	0	0	0	0	0	100
	1	0	4	0	0	0	0	0	100
	2	0	0	4	0	0	0	1	80
	3	0	0	0	3	0	1	0	75
	4	0	0	0	0	4	1	0	80
	5	0	0	0	0	0	2	0	100
	6	0	0	0	1	0	0	3	75

Table 5, The testing accuracy of Vision CNN model

Moreover, by observing the prediction probability output by the SoftMax layer, the average probability of correct predictions is 99.1% and 84.0% for wrong predictions. Therefore, the vision threshold value is defined to be 91.6%. Which tells the robot arm that for the prediction probability below 91.6%, the predictions are uncertain. It requires tactile data to confirm the grid.

For testing fabric 1, grid 7 is uncertain. For testing fabric 2, grids 2 and 16 are uncertain. Grids 7 and 19 are uncertain for testing fabric 3. Therefore, the Gelsight sensor collects tactile data from these five grids as the testing dataset for the tactile CNN model.

5.3 Assessment of the Tactile CNN model

648 photos are used to train the Tactile CNN model. The batch size is 30, with 1000 epochs and a 0.0005 learning rate. Since the photo consists of three primary colours, the input channel is 3. The training loss is shown in Figure 16.

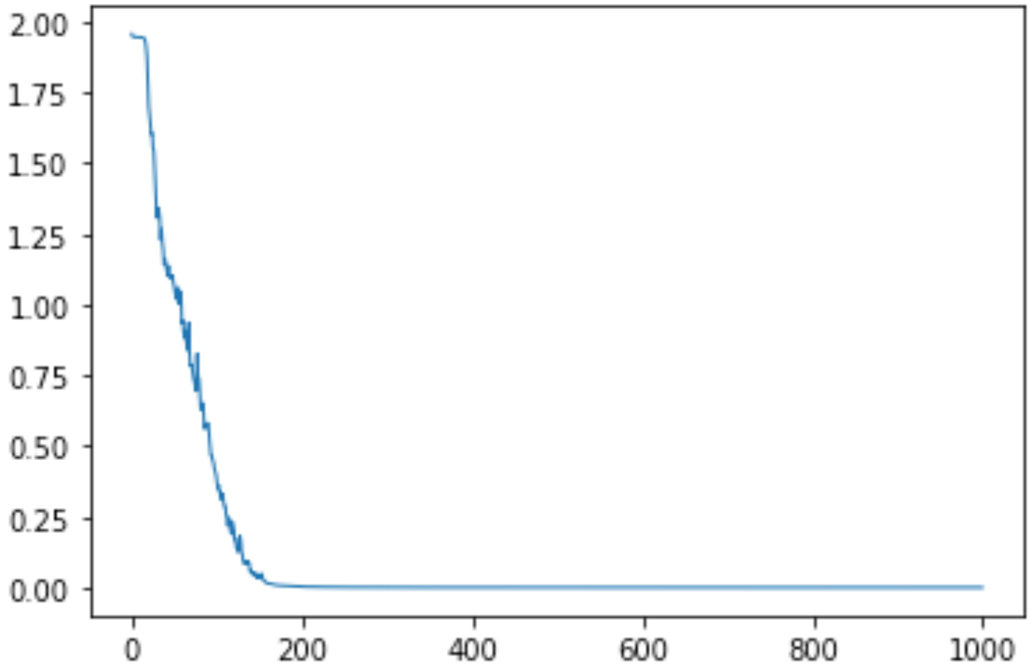


Figure 16, training loss per epoch of Tactile CNN model

Testing accuracy on 72 testing datasets is 52%. Table 6 shows the testing accuracy of each prediction. For defects 0 and 6, the testing accuracy is 100%. For defects 2,3 and 4, the testing accuracy is below 50%. This implies that the tactile CNN model is good at identifying no defects and misprinting. Moreover, by observing the table, all the wrong predictions come from label 0. Eleven of them are predicted to defect 2. It reveals that it is difficult to identify the difference between no defect and water stain.

Tactile CNN Model		Label							Accuracy (%)
		0	1	2	3	4	5	6	
Prediction	0	14	0	0	0	0	0	0	100
	1	1	4	0	0	0	0	0	80
	2	11	0	4	0	0	0	0	26.7
	3	9	0	0	4	0	0	0	30.8
	4	9	0	0	0	4	0	0	30.8
	5	4	0	0	0	0	4	0	50
	6	0	0	0	0	0	0	4	100

Table 6, The testing accuracy of Tactile CNN model

5.4 Assessment of the Vision Tactile Combined Model

By comparing tables 5 and 6. Both Vision and Tactile CNN models have 100 accuracies in predicting defect 0. For defects 1 to 5, the Vision CNN model has better performance. On the other hand, the Tactile CNN model performs better in predicting defect 6. Therefore, the prediction of the vision CNN model is replaced by the tactile CNN model only if the prediction of the tactile CNN model is 6.

The prediction of the 5 uncertain grids mentioned in section 5.2 is 6, 5, 5, 1, 2, respectively. Only the first tactile CNN prediction is replaced in the Vision CNN prediction because it is predicted as defect 6. Finally, the testing accuracy of the vision tactile combined model is increased to 95.8%.

Instead of scanning 72 grids for the Tactile testing dataset, only 5 grids are used to build the testing dataset. Therefore, the time required to collect the tactile testing dataset is 14.4 times faster.

6 Findings of Task B

After demonstrating the accuracy score of the Vision CNN model, Tactile CNN model and Vision Tactile Combined model. This part focuses on discussing the result and the phenomenon.

6.1 The Nature of Defects

Before looking into the design and result of the two models, the natural human reaction to the cloth quality checking process should be discussed first. A simple definition of artificial intelligence is to mimic cognitive functions in the human mind, including how humans learn and their problem-solving skills [18]. Therefore, the tailor’s reaction when dealing with a fabric is as follows.

First, an experienced tailor has seen and touched a lot of defects in fabric before, such that, he was familiar with different types of common defects including their vision features and tactile properties. When there is a new fabric that he should examine, the natural reaction is to look at the fabric under a light source. An experienced tailor should be able to identify most of the defects by vision. This process is known as the “Light Test” mentioned in section 1.1.

Most of the defects should be identified and localized by the tailor with his vision. However, some of them are difficult to identify only by vision, therefore, the tailor moves his palms to the suspicious location and touches it to confirm the defects.

This process is known as the “Uniformity Test” mentioned in section 1.1. After feeling the tactile texture of the defect, the tailor should be able to confirm the defect type with a higher certainty level.

This project assumed that defects in the fabric can be classified into four classes.

1. Defects which is obvious from a visual and tactile perspective.
2. Defects which is obvious from a visual perspective, but implicit from a tactile perspective.
3. Defects which is implicit from a visual perspective, but obvious from a tactile perspective.
4. Defects which is implicit from a visual and tactile perspective.

Even human finds that it is difficult to identify defect class 4. Because it is difficult to identify with human vision and tactile sensation. The defect is believed to be very minor such that maybe it could only be identified under a microscope. Most of the time, both the tailor and customer can accept this kind of defect, therefore, it can be ignored in the quality checking process. Thus, class 4 is not in the scope of this project.

Defect classes 1 and 2, can be identified with high accuracy scores by vision data easily, therefore, tactile data is not required. Thus, the idea of task B is mainly tackling defect class 3, which cannot be identified with vision data certainly. Then, tactile data is used to improve the accuracy score.

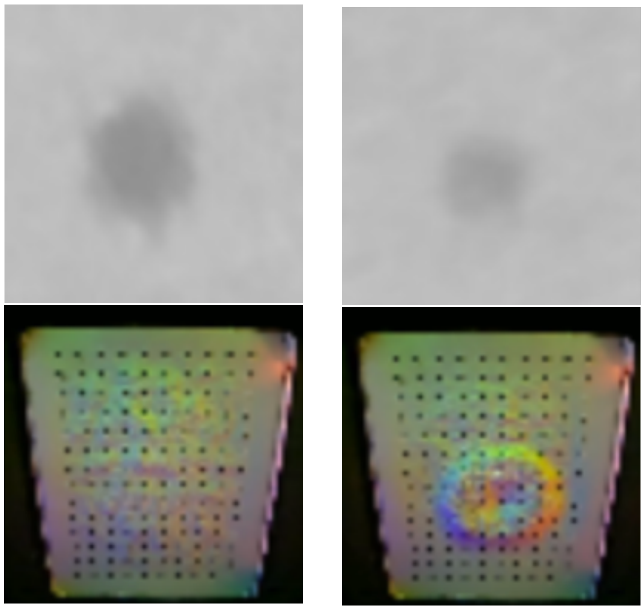


Figure 17, An example of defect class 3

Figure 17 shows an example of defect class 3, the defect on left is a water stain (label 2) and the defect on the right is misprinting (label 6). Besides the size, their visual features are very similar, therefore it is expected to be mixed up with each other. Regarding table 5, 25% of misprinting is predicted as water stain in the Vision CNN model. Therefore, the experimental result is aligned with the expected result.

On the other hand, their difference in tactile data is obvious. Therefore, 100% of water stain is predicted as water stain and 100% of misprinting is predicted as misprinting in the Tactile CNN model in table 6.

Moreover, defect 0 (no defect) belongs to defect class 1. Because both vision and tactile CNN models have 100% accuracy in predicting this defect. Defects

1 to 5 (a hole, water stain, grease spot, drop stitches and worn out) belong to defect class 2 because their prediction accuracy is higher in the Vision CNN model, regarding tables 5 and 6. Misprinting belongs to class 3 because its prediction accuracy is higher in the Tactile CNN model.

6.2 How Tactile Data Helps Improve the Accuracy Score

After understanding the four classes of the defects, the reason for the tactile data improving the accuracy is further discussed in this part. As mentioned in sections 5.2 to 5.4, the accuracy score of the Vision CNN model, Tactile CNN model and the combined CNN model are 94.4%, 52% and 95.8% respectively. The reason for the low accuracy of the Tactile CNN model and how this low accuracy model improves the Vision CNN model by 1.4% are as follows.

First, the low accuracy score of the Tactile CNN model is attributed to the high portion of the class 2 defect chosen in this project. As mentioned in section 6.1, 5 out of 7 (71%) defect types chosen in this project belong to class 2 which is expected to have low accuracy in the Tactile CNN model. In other words, the low accuracy of the Tactile CNN model is not because of the design failure of the CNN model itself, but the design of the defect. If more defects in classes 1 and 3 are included in this task, the accuracy of the Tactile CNN model is expected to increase significantly.

Second, the reason why a low accuracy model improves a high accuracy model is attributed to the contribution of the high accuracy of the class 3 defects in the Tactile CNN model. As shown in Table 6, the Tactile CNN model has 100% accuracy in predicting misprinting. However, the Vision CNN model only has

75% accuracy in predicting misprinting. Thereby, the result of the Vision CNN model should be replaced with the Tactile CNN model if the prediction is 6. Thus, the accuracy is expected to increase. And the experimental result supports this assumption.

Third, the improvement is not significant (1.4%). It is because the accuracy of the Vision CNN model is high, which implies that there is a less uncertain grid. Therefore, only 5 among 72 grids have to be predicted by tactile data. And only 1 of the 5 is predicted as misprinting and replaced. Therefore, the insignificant improvement is as expected.

Moreover, it is normal that most of the defects can be identified using visual identities by humans, and tactile texture only improves the certainty to a small extent. Therefore, using the tactile prediction only improves the accuracy of the Vision CNN model by 1.4%.

Although there is only a 1.4% accuracy improvement, the time used to collect tactile data is reduced significantly. Instead of scanning 72 grids, only 5 grids have to be scanned in this combined model, thus the time is 14.4 times faster. And this is the reason for calling this task “Rapid” Defect Localization. Therefore, both time and accuracy have improved with the vision tactile combined model.

7 Limitation

This section focuses on the difficulties which hinder the process of the fabric type classification task and the rapid defect localization task, and the mitigations applied in the experiment process.

7.1 Elastomer Detachment of The Magnetic Tactile Sensor

Some of the fluffy fabrics such as the fabric shown in figure 18 caused the elastomer to detach from the PCB. It directly stops the progress of data collection. The corresponding solution is using the silica gel to stick it back. This solution is provided by the inventor of the magnetic tactile sensor Mr Yan. It prevents the elasticity of the elastomer to be affected by the glue.

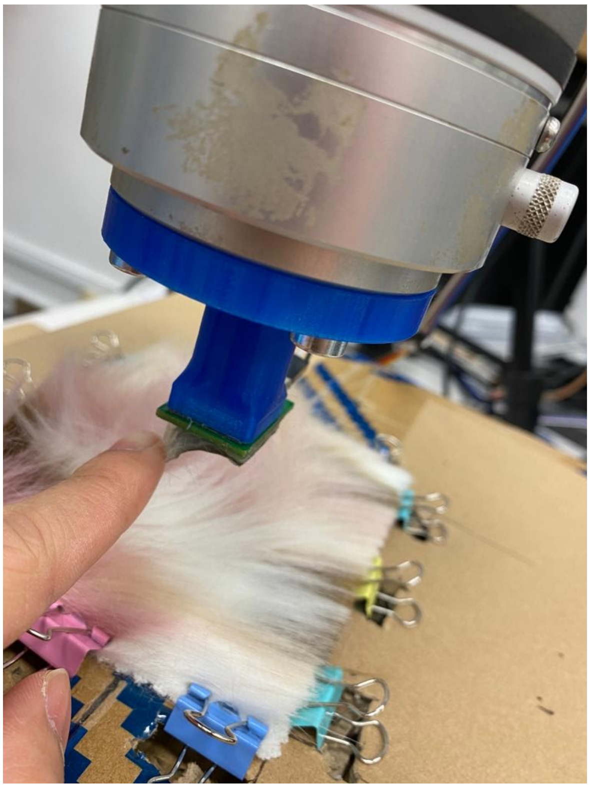


Figure 18, The elastomer detachment

7.2 Fiber Detachment

Some of the fibres are detached after sliding through by the sensor as shown in figure 19. The detached fibres are stuck in between the flexible magnet and the fabric. It decreases the friction of the fabric, thus affecting the result. The mitigation is to remove all detached fibres after each trajectory manually.

However, stopping the robot arm after each trajectory hinder the progress of the data collection progress.

Figure 19, The detached fiber

7.3 Fabric Deformation

Since the sensor stretches the fabric to measure the elasticity of the fabric as shown in figure 20. After stretching the fabric, some of the fabrics are not able to rebound back to the original scale. Thus, affecting the reading of the next trajectory. Therefore, the solution is to smooth out the fabric manually after each trajectory. It hinders the progress of data collection.

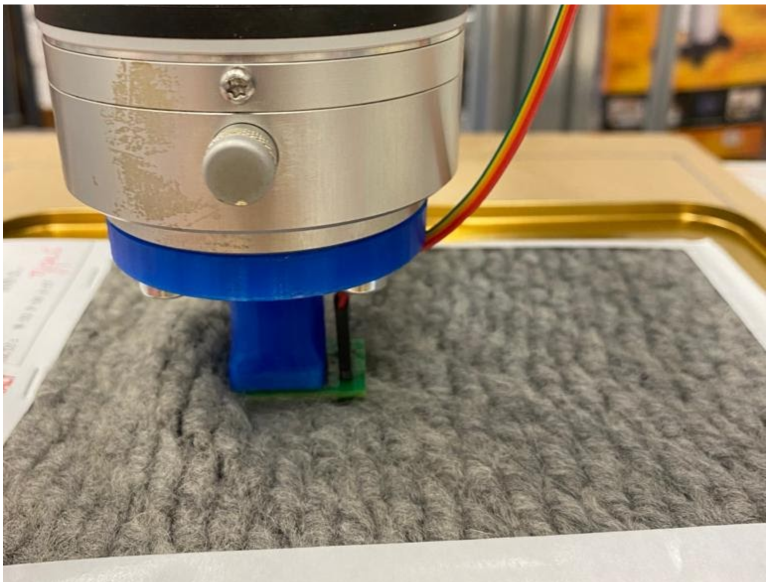


Figure 20, The stretched fabric

7.4 Membrane Contamination of Gelsight Sensor

Defect type 6 (Misprinting) of task B is imitated by adding 4 drops of Aron Alpha glue as mentioned in section 5.1. However, if it is scanned by the Gelsight sensor without fully solidifying, it contaminates the membrane of the Gelsight sensor. The pollution caused by the Aron Alpha glue can be removed with an Aron Alpha Debonder theoretically. However, after removing the glue with the debonder, indelible harm is caused to the membrane. Figure 21 shows the damage caused by the debonder.

The solution to this problem is to use the reserved Gelsight sensor to continue the experiment. However, all fabrics have to be scanned once more to maintain consistency. It hinders the progress of the data collection process of the rapid defect localization task.

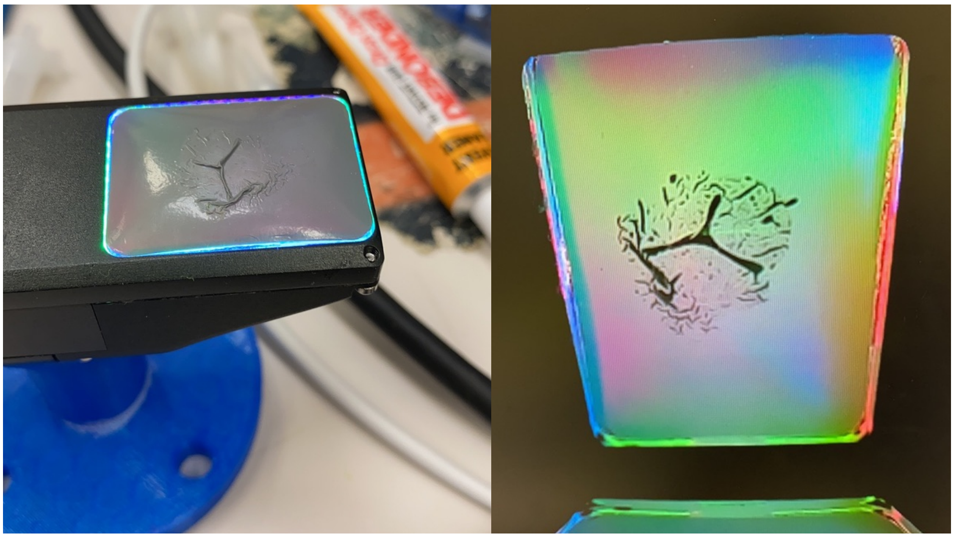


Figure 21, The contaminated membrane

8 Further Enhancement

After understanding the difficulties faced in both tasks, there are various enhancement to be done for further studies. The suggested further enhancement for both tasks are as follows.

8.1 Fabric Type Classification

One of the enhancements is to apply the Gelsight sensor to this task accompanied by the convolution CNN network mentioned in section 1.7. Since the fabric type classification task captures continuous tactile data as an input. Thus, if the Gelsight sensor is used, Image data is captured. Then, it satisfied the input condition of the convolution CNN network, as continuous image data.

This enhancement studies the performance of the convolution CNN network in the fabric type classification task.

8.2 Rapid Defect Localization

First, different types of fabric can be applied to this task. The fabric type can be classified by using vision data and the Vision CNN model or Gelsight tactile data with the Tactile CNN model.

Moreover, The defective texture on a different type of fabric is unique. Thus, it increased the difficulties and complexity of this task. To achieve it, the training data size should be increased significantly to learn the same defect on different fabrics.

Second, cross-grid defects and borderline defects can be further studied. In the reality, the defect may not be exactly lying within a grid, or the defect is too large to be in one single grid. Therefore, these problems should be further studied to enhance the competence of this model. Cross-grid defects and borderline defects can also be used to train and test the model.

Instead of using the grid to define the coordinate of the defect, The Gelsight sensor could scan through the whole fabric by sliding on it. When the defect is scanned, the current position of the robot arm is saved and defined as the coordinate of the defect. After that, a convolution CNN network suggested in [16] can be applied to this continuous image tactile data.

Third, Deep Maximum Covariance Analysis (MWCA) suggested in [4] can also be applied to tactile vision combined data with a different assumption. This task assumed that defects can be classified into four classes in section 6.1. In some instances, the vision data are similar while the tactile textures are different as shown in figure 17. But if it is assumed that, if defects have a similar vision texture, their tactile textures are similar and vice versa, then, DMCA can be applied in this task too. Furthermore, the difference between the performance of the vision tactile combined CNN model and the DMCA model can be further studied.

Fourth, Fast R-CNN in [6] can be applied to the vision data. The region of interest (ROI) layer finds out the coordinates of the region where the CNN should pay more attention. Each ROI is a rectangular window defined by four-tuple (r,c,h,w). (r,c) is the coordinate of the upper left corner and (h,w) is the height and width of the window. It replaces the grid as another coordinate system for the robot arm. Therefore, it can be further applied to the rapid defect localization task.

9 Schedule and Workflow

Table 7 illustrates the schedule and the workflow of the project.

Time	Work Completed
September	• Detailed project plan • FYP Website design
October	• Software & hardware familiarization • Purchase required fabric
November & December	• Work on data collection, data pre-processing and LSTM model training of Task A
January	• First presentation • Intermediate report
February & March	• Work on data collection, Vision Tactile CNN model training of the Rapid defect localization task
April	• Final presentation • Final report
May	• Project exhibition • Project competition

Table 7 The schedule and the workflow of the project.

10 Conclusion

The Vision Tactile combined network is one of the outstanding solutions to the fall behind quality checking technology in the textile industry. The main objective of this project is to check the quality of fabric using magnetic and optical Gelsight tactile sensors. The main objectives are further divided into two sub-objectives, fabric type classification (task A) and rapid defect localization (task B). They are all achieved with a marvellous result.

The work completed for task A includes the interview of two experienced tailors and the data collection of 30 types of fabric by the robot arm and hand. The testing accuracy of 30 types of fabric is 100% by the robot arm. The factors affecting the testing accuracy score include the train to test data ratio and the standardized data collection procedure. Therefore, the testing accuracy of data collected by the robot arm is higher than the hand by 11.1%.

Six types of common defects suggested by the tailors are mimicked in the rapid defect localization task. 6480 samples are used to train the combined CNN model. In addition, all defects are classified into four classes and the combined model mainly tackles the defects with implicit visual texture but obvious tactile texture. Therefore, the Tactile CNN model with relatively low accuracy boosted the accuracy of the Vision CNN model by 1.4%, and the final accuracy of the Vison Tactile CNN model is 95.8%. Thus, the result shows that the Vision-Tactile combined CNN model outperforms the Vision CNN model by 1.4% and the time used is 14.4 times faster.

The difficulties faced in both tasks include the detachment of elastomer, fibre, the deformation of the fabric and the Gelsight membrane contamination. These difficulties hinder the progress significantly and sometimes lead to the recollection of data. However, it can be solved manually with perseverance.

As for the enhancement of both tasks, convolution CNN with continuous tactile image data can be used to study the performance of this network on fabric type and defect classification tasks. Moreover, some other coordinate systems can be used to define the location of the defect. Such as the ROI layer in the fast R-CNN network. Therefore, the defect does not limit by its size and location. In addition, cross-grid defects and borderline defects should also be studied to enhance the capability of this model.

Reference

[1] W. Yuan, R. Li, M. A. Srinivasan and E. H. Adelson, "Measurement of shear and slip with a GelSight tactile sensor," 2015 IEEE International Conference on Robotics and Automation (ICRA), 2015, pp. 304-311, doi: 10.1109/ICRA.2015.7139016.

[2] Understanding LSTM Networks -- colah's blog. (2021). Retrieved 29 November 2021, from https://colah.github.io/posts/2015-08-Understanding-LSTMs/

[3] Yan, Y., Hu, Z., Shen, Y., & Pan, J. (2021). Surface Texture Recognition by Deep Learning-Enhanced Tactile Sensing. Advanced Intelligent Systems, 2100076. doi: 10.1002/aisy.202100076

[4] Luo, S., Yuan, W., Adelson, E., Cohn, A., & Fuentes, R. (2018). ViTac: Feature Sharing between Vision and Tactile Sensing for Cloth Texture Recognition. Retrieved from https://arxiv.org/pdf/1802.07490.pdf

[5] Anagnostopoulos, C., Vergados, D., Kayafas, E., Loumos, V., & Stassinopoulos, G. (2001). A computer vision approach for textile quality control. The Journal of Visualization and Computer Animation, 12(1), 31–44. https://doi.org/10.1002/vis.245

[6] Girshick, R. (2015). Fast R-CNN. 2015 IEEE International Conference on Computer Vision (ICCV). https://doi.org/10.1109/iccv.2015.169

[7] M. K. Johnson and E. H. Adelson, “Retrographic sensing for the measurement of surface texture and shape,” in Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009, pp. 1070–1077.

[8] M. K. Johnson, F. Cole, A. Raj, and E. H. Adelson, “Microgeom- etry capture using an elastomeric sensor,” in ACM Transactions on Graphics (TOG), vol. 30, no. 4. ACM, 2011, p. 46.

[9] M. K. Johnson, F. Cole, A. Raj, and E. H. Adelson, “Microgeom- etry capture using an elastomeric sensor,” in ACM Transactions on Graphics (TOG), vol. 30, no. 4. ACM, 2011, p. 46.

[10] N. J. Ferrier and R. W. Brockett, “Reconstructing the shape of a deformable membrane from image data,” The International Journal of Robotics Research, vol. 19, no. 9, pp. 795–816, 2000.

[11] K. Kamiyama, K. Vlack, T. Mizota, H. Kajimoto, K. Kawakami, and S. Tachi, “Vision-based sensor for real-time measuring of surface traction fields,” IEEE Computer Graphics and Applications, vol. 25, no. 1, pp. 68–75, 2005.

[12] C. Chorley, C. Melhuish, T. Pipe, and J. Rossiter, “Development of a tactile sensor based on biologically inspired edge encoding,” in Advanced Robotics, 2009. ICAR 2009. International Conference on. IEEE, 2009, pp. 1–6.

[13] X. Jia, R. Li, M. A. Srinivasan, and E. H. Adelson, “Lump detection with a gelsight sensor,” in World Haptics Conference (WHC), 2013. IEEE, 2013, pp. 175–179.

[14] R. Li, R. Platt, W. Yuan, A. ten Pas, N. Roscup, M. A. Srinivasan, and E. Adelson, “Localization and manipulation of small parts using gelsight tactile sensing,” in Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on. IEEE, 2014, pp. 3988–3993.

[15] K. Sato, K. Kamiyama, N. Kawakami, and S. Tachi, “Finger-shaped gelforce: sensor for measuring surface traction fields for robotic hand,” IEEE Transactions on Haptics, vol. 3, no. 1, pp. 37–47, 2010.

[16] Zhang, Y., Kan, Z., Tse, Y., Yang, Y., & Wang, M. (2018). FingerVision Tactile Sensor Design and Slip Detection Using Convolutional LSTM Network [PDF]. Retrieved from https://arxiv.org/pdf/1810.02653.pdf

[17] Wenzhen, Y., Chenzhuo, Z., Andrew, O., Mandayam A., S., & Edward H., A. (2017). Shape-independent Hardness Estimation Using Deep Learning and a

GelSight Tactile Sensor [PDF]. Cornell University. Retrieved from https://arxiv.org/pdf/1704.03955.pdf

[18] Russell, S. and Norvig, P., n.d. Artificial intelligence.

[19] Lecun, Y., Bottou, L., Bengio, Y. and Haffner, P., 1998. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), pp.2278-2324.

[20] Cun, Y., Kanter, I. and Solla, S., 1991. Eigenvalues of covariance matrices: Application to neural-network learning. Physical Review Letters, 66(18), pp.2396-2399.

Appendix 1 The Combination of Tactile and Visual Data

[1] Wettels, N., Fishel, J., Su, Z., Lin, C., & Loeb, G. (2009). Multi-modal Synergistic Tactile Sensing [PDF] (9th ed.). Tactile Sensing in Humanoids – Tactile Sensors and Beyond Workshop. Retrieved from https://syntouchinc.com/wp-content/uploads/2016/12/2009_Wettels_Humanoids.pdf

Appendix 2 Deep ConvNets

[1] A. Krizhevsky, I. Sutskever, and G. Hinton. ImageNet clas- sification with deep convolutional neural networks. In NIPS, 2012.

[2] Y. LeCun, B. Boser, J. Denker, D. Henderson, R. Howard, W. Hubbard, and L. Jackel. Backpropagation applied to handwritten zip code recognition. Neural Comp., 1989.