Article

An End-to-End Grasping Stability Prediction Network for Multiple Sensors

Xin Shu ^{1,2}, Chang Liu ¹ and Tong Li ^{1,*}

¹ National Institute of Standards and Tec State Key Laboratory Transducer, Institute of Electronics, Chinese Academy of Sciences, Beijing 100190, China; shuxin16@mails.ucas.ac.cn (X.S.); liuchang003968@aircas.ac.cn (C.L.)

² University of Chinese Academy of Sciences, Beijing 100049, China

* Correspondence: tli@mail.ie.ac.cn

Received: 18 February 2020; Accepted: 10 March 2020; Published: 14 March 2020

Featured Application: This work can be applied in the grasping operation of the manipulator. It helps to predict the grasping stability.

Abstract: As we all know, the output of the tactile sensing array on the gripper can be used to predict grasping stability. Some methods utilize traditional tactile features to make the decision and some advanced methods use machine learning or deep learning ways to build a prediction model. While these methods are all limited to the specific sensing array and have two common disadvantages. On the one hand, these models cannot perform well on different sensors. On the other hand, they do not have the ability of inferencing on multiple sensors in an end-to-end manner. Thus, we aim to find the internal relationships among different sensors and inference the grasping stability of multiple sensors in an end-to-end way. In this paper, we propose the MM-CNN (mask multi-head convolutional neural network), which can be utilized to predict the grasping stability on the output of multiple sensors with the weight sharing mechanism. We train this model and evaluate it on our own collected datasets. This model achieves 99.49% and 94.25% prediction accuracy on two different sensing arrays, separately. In addition, we show that our proposed structure is also available for other CNN backbones and can be easily integrated.

Keywords: robotic grasping; tactile perception; intelligent manipulation; stability prediction

1. Introduction

With the recent developments in computer vision and range sensing, robots can detect objects reliably. However, grasping remains to be a question even with the correct location and pose of the object. The main reason is that predicting the grasping stability of the object in autonomous robotic manipulation tasks is still an important and difficult research topic. Grasping stability is defined as the capacity of the grasping to resist external forces and disturbances. A stable grasping can be viewed as a static equilibrium state of the object that is maintained when the grippers are closed. In the case of stable grasping, the grasped objects will keep the static equilibrium state when the manipulator moves. If the grippers do not grasp the object stably during the operation of the manipulator, it can lead to a state in the desired action sequence from which the system cannot recover easily.

In comparison, humans can easily rely on remarkable tactile sensing capabilities and perceive the stability in grasping [1]. We can quickly identify textures and fine features using our fingertips, and while holding objects we subconsciously control grip forces and prevent slippage. When it comes to grasping with tools, we act as if they are extensions of our limbs, which requires not only dexterous

manipulation but a sophisticated and implicit understanding of how the sensations at our fingertips related to the interactions between the tool and the environment.

As for manipulator, it is similar and tactile sensors are applied in grasping. There are kinds of tactile sensors based on electric [2], optic [3], acoustic [4], or other effects [5,6]. Developments in tactile sensors also push the method of grasping stability prediction with these advanced sensors. In some cases, traditional tactile features or descriptors are used as the input. Ref [7] builds tactile features and they are fed to a support vector machine. [8] presents a modified SIFT descriptor, which deems the input as an image. In other cases, raw input is directly fed into the machine learning models or deep learning models. [9] uses support vector machines and random forests to predict the sensor. Similarly, [10] uses a multi-layer perceptron to classify grasping stability. As deep learning models have achieved large success in classification tasks in other domains such as vision and audio, [11] explores the possibilities of convolutional neural networks for the task of stability prediction. In addition, [12,13] use the graph convolutional network to predict the grasp stability with tactile sensors, which also receives a good performance.

Most of the current methods to study this problem focus on how to identify the stability of the entire grasping process, while our goal is to predict the grasping process stable after the gripper has caught the object and before lifting it. Some works pay attention to this prediction such as [14] and [15]. While methods mentioned before are all designed to the specific sensing array and models cannot perform well on other sensors. We cannot transfer the existing prediction model and we have to train a new model for the new sensor from scratch. In this paper, we propose a unified tactile representation for grasping stability prediction and provide an end-to-end inference system for multiple sensors.

The main contributions of this work can be divided into two parts. Firstly, a unified tactile representation is proposed for the utilization of transfer learning on different tactile sensors. Secondly, a novel network structure based on the tactile representation is proposed to provide an end-to-end approach to inference the grasping stability prediction for multiple sensors and we call it MM-CNN.

The rest of the paper is organized as follows. In Section 2, we introduce the details of our dataset building. We propose a unified tactile representation and an end-to-end convolutional neural network for multiple tactile inputs with weight sharing mechanism in Section 3. Section 4 reports the experimental results on model justification on several datasets. Finally, we conclude the aforementioned works in Section 5.

2. Dataset Building

The manipulator is usually equipped with grippers to accomplish the operation of grasping and the tactile array is assembled on the grippers of the manipulator. To grasp an object, grippers of the manipulator open to the maximum and then they arrive at a suitable location and pose for grasping. After that, grippers close slowly and stop at the time when the object is held. Thus, how to determine if an object is held or not plays an important role in the grasping task. The human body judges it by the perception of tactile physiological signals rather than vision. In the situation of manipulator grasping, similarly, force distribution on sensing arrays shows more details and preforms better than vision. In order to research and model the relationship between the output of tactile sensing array and grasping stability, we build the dataset by collecting the output and corresponding grasping stability of two different tactile sensing arrays.

2.1. Kitronyx Tactile Sensing Array

The Ms9723 tactile sensing array is a force distribution measurement product of Kitronyx company, which supports multi-touch technologies. It consists of 160 units and these 160 units are distributed in 16 rows and 10 columns. The output of each unit is a natural number normalized with 255. We find that each unit of this array is insensitive to the surface force and has a relatively higher measuring lower boundary. As a result, the output array is easy to be filled with zero and the numerical value is relatively small when grasping objects.

2.2. Self-Made Tactile Sensing Array

Our tactile sensing array also supports multi-touch and can measure the three-dimensional force on it. The sensing material and the electrodes form a sandwich structure. Meanwhile, the electrodes are connected by electrical routings from different rows and columns. Overall, the sensor has 64 units and the area of each unit is 7×7 mm to guarantee a spatial resolution. It works with multiplexers, microcontroller unit (MCU), analog-to-digital converter (ADC), reference resistances, and operational amplifiers to make up the measurement circuit. The measurement principle of the scanning circuit is that the row multiplexer is controlled by MCU to select one of the rows, and then the output of each column of operational amplifiers is connected to the analog-to-digital converter through the column multiplexer and the output voltage is converted to digital signals. As the spatial distribution of this array is more complex than the Kitronyx array, it will be discussed in detail later. Compared with the Kitronyx tactile sensing array, our array has the ability to perceive smaller forces. The output of our array starts to change when the force is about 0.05 N, while the output of the Kitronyx array starts to change when the force achieves around 1 N. In addition, our array consists of fewer units. The output of the sensing unit is also a natural number and there is no normalization applied to the output. The maximum of the output can reach the numerical value of 2000 and there are few zeros in the output because of its smaller measuring lower boundary.

2.3. Data Collection and Statistics

Although these two arrays vary a lot in some aspect as mentioned above, there are still some common characteristics between them. The outputs of tactile sensing arrays are both positively relative to the magnitude of the force on arrays. This common characteristic makes it possible to provide a unified grasping stability prediction model for these two different sensing arrays. Another common characteristic is that when there is no force on the array, units in these two arrays all output zero. The output of tactile sensors also has three dimensions which represent the height, the width, and the channel. While the channel dimension has the length of one, instead of three. Therefore, outputs of the whole array are not only individual numerical values but also can be regarded as an image of a single channel. In addition, objects of different kinds of shapes are grasped during the data collection as we aim to propose a robust prediction model. Therefore, we show our grasping targets during dataset building in Figure 1. We close grippers from maximum step by step. After this process, if the grasping object keeps a static equilibrium state during grasping, the label of corresponding output is assigned to 1, which infers to the stable grasping. Otherwise, the label is assigned to 0, which infers to the unstable grasping. Additionally, we try with different grasping poses to enlarge the dataset. In general, we collect 1562 samples for Kitronyx tactile sensing array and 3478 samples for the self-made array. We guarantee that these samples are roughly balanced among different objects and there are no different manipulator movement parameters used during collection. The samples collected from Kitronyx tactile sensing array form the GKD (grasping Kitronyx dataset) while the 3478 samples from the self-made array form the GSD (grasping self-made dataset). We randomly select around a quarter of samples from GKD and GSD separately to make up the test set. The test part of each dataset is divided into the train set. We make sure that both the train part and the test part of datasets are still balanced among objects after the split of dataset. We merge GKD and GSD into a new dataset called GMD (grasping merged dataset). The approach of fusion is quite simple. We merge the train set of GKD and GSD to the new train set. Similarly, we merge the test set of GKD and GSD. In this way, GMD has 3779 samples in the train set and 1261 samples in the test set. Samples in the GMD have two labels, one infers to the annotated grasping stability and the other infers to the resource dataset (GKD or GSD) of the sample. More details of datasets can be seen in Table 1.

Figure 1. Graspable objects in dataset.

Table 1. Statistics of datasets.

Dataset	Dataset Samples	Train Set Samples	Test Set Samples	Sample Size
GKD	1562	1171	391	160 values and one label
GSD	3478	2608	870	64 values and one label
GMD	5040	3779	1261	(160/64) values and two labels

3. MM-CNN

We formulate the grasping stability prediction as a two-way classification problem. Back to the origin, our initial aim is to find the relationship between the different tactile sensors and propose a convenient model to predict grasping stability for multiple tactile sensing arrays in an end-to-end way. There are two main difficulties to be solved, in general. On the one hand, although the outputs of different arrays are all positively relative to the magnitude of the force, they vary a lot from each other under the same external force. On the other hand, the model needs to have the ability to treat different kinds of input sizes. These two difficulties limit a lot in some way. As the input sizes are relatively small, we first present a shallow network in Figure 2 as the baseline of our models and the input data is normalized before sent to the network. Normalization helps with the case of different output numerical values for different tactile sensors. Next, we extend the framework to a novel end-to-end system with weight sharing for grasping stability prediction.

Figure 2. Structure of baseline model.

3.1. Spatial Information

In the case of Kitronyx tactile sensing array, the distribution of 160 units is shown in Figure 3a. The number on the unit in figure infers the order of array outputs. Thus, it is easy for us to reshape the output 160 numerical values to the 16*10 matrix to recover the spatial information from the output. As for our tactile sensing array, the distribution is more complex. The spatial distribution of our array is illustrated in Figure 3b. The block with full lines in the figure infers to the sensing unit while the block with dotted line infers to a blank space. We put these 64 output numerical values back to their positions in spatial and pad zeros to other positions. Then, it comes out a matrix with the size of 12×12 . As a comparison, we directly reshape the 64 numerical values into a matrix with the size of 8×8 , which loses the spatial information. Experiments show that recovering the spatial information of the sensing array helps in grasping stability prediction.

Figure 3. Spatial distribution of Kitronyx tactile sensing array (a) and spatial distribution of self-made tactile sensing array (b).

3.2. Fully-Convolutional Backbone with Global Pooling

A fully-convolutional backbone with global pooling is designed to process the input of different sizes. As shown in Figure 2, the baseline model is composed of the backbone and the fully-connected classifier. The backbone includes two convolutional layers and two max-pooling layers. The convolutional layers have the convolution kernel of 3×3 . As the stride of the convolutional layer is one, it does not reduce the resolution of spatial. Additionally, the pooling layers with the kernel size of 2×2 have the stride of two, which is deemed as down-sampling on the feature map. These two convolutional layers and two max-pooling layers are collectively referred to as the network backbone. They extract features from the input. Two fully-connected layers are followed to act as the role of the classifier. This structure is defined as the baseline model while it cannot deal with the input of different sizes as the input of a fully-connected layer requires to be fixed. Therefore, we propose the fully-convolutional backbone with global pooling to extract the fixed-size feature. There is a global pooling layer after the second max-pooling layer and it can generate a feature embedding of 128 in length for inputs of different sizes. We define this feature after the global pooling layer as the tactile representation, which performs well when transferring to the prediction models of different sensors. This is a desirable characteristic as it makes the training of the weights of the model much easier than the former work as they do not need to be trained from scratch anymore.

3.3. Multiple Fully-Connected Heads

We propose the structure of multiple fully-connected heads to model different grasping stability predictions for each sensor while these heads share the weights of tactile representations. Generally speaking, different sensors rely on different working effects. Thus, outputs of sensors vary a lot although outputs are all positively relative to the magnitude of the force. It is difficult to model

them in a single classifier, so we decide to remain the parameters in the fully-connected heads to be dependent on the specific sensor, which is effective to maintain the classification accuracy. In this way, the information of the force threshold of grasping is involved in the fully-connected layers and different sensors have their own fully-connected heads. However, this design has the drawback that the number of parameters in the models grows linearly with the number of sensors if we do not utilize weight sharing. To optimize and reduce the number of learnable parameters, these heads for classification share the backbone weights with each other. It is proved to be available and effective in experiments. This weight sharing strategy greatly reduces the number of learnable parameters while being able to maintain the classification accuracies on the dataset of different sensors. Each head outputs a vector of two in length and we concatenate them together to a vector of $2N$ in length as the final output of the grasping stability prediction branch (N is the number of different sensors). Meanwhile, it leaves us a problem of how to determine which fully-connected head to be used during inference and we add an additional mask branch, therefore.

3.4. Mask Branch

We follow this path and design a mask branch to decide which fully-connected head to be used for grasping stability prediction in inference. As shown in Figure 4, the structure of the mask branch is similar to the single head grasping stability prediction branch. The output of the classifier is a vector of N -length, which represents the probability distribution between N sensors. Then, a binary mask of $2N$ -length is generated based on the output according to Equation (1).

$$M_i = \begin{cases} 1, & \text{if } O_{\lfloor \frac{i}{2} \rfloor} \text{ is the maximum in } O_k (k = 1, 2, \dots, 2N) \\ 0, & \text{Otherwise} \end{cases} \quad (1)$$

In the equation, the output of the classifier is described as O while M is the binary mask. For example, we consider the case that $N = 2$. If $O_1 \geq O_2$, then M is equal to $[1, 1, 0, 0]$. Additionally, M is equal to $[0, 0, 1, 1]$ in the case that $O_1 < O_2$. Finally, we multiply the grasping stability prediction output with the mask output to determine which fully-connected branch to use during inference. Argmax is applied to the multiplication to predict grasping stability. Finally, our model has the advantage of parallel to the existing grasping stability prediction systems and thus can be easily integrated with the way of replacing the backbone convolutional neural network.

Figure 4. Structure of mask multi-head convolutional neural network (MM-CNN).

3.5. Implementation Details

As there are two branches in the structure, the training strategy is important for the whole network. During training, we train the grasping stability prediction branch and the mask branch separately. During the training of the grasping stability prediction branch, we first pretrain the tactile representation and one fully-connected classifier with its corresponding dataset (GKD or GSD). In the case that $N = 2$, we utilize the pretrained tactile representation and finetune the other fully-connected classifier with the other dataset in the second step. As for the mask prediction branch, we train it with the data in GMD. As the batch fed to the network should be unified in size, batches alternate between samples from GKD and samples from GSD. Fortunately, we find that the mask prediction branch converges well during training with this strategy. When it comes to inference, the output of the grasping stability prediction branch and the output of the mask prediction branch are multiplied together to determine the final prediction in an end-to-end manner.

What is more, there are some details during training and inference. The ReLU (rectified linear unit) is used as the activation function to increase the feature expression ability of the model. We select the cross-entropy with sigmoid as the classification loss function to optimize the multiple fully-connected heads and the mask prediction branch. Each head in grasping stability prediction branch is a two-way classifier while the mask prediction is processed as a problem of N -way classification. We use a steady learning rate of 0.0005 and select the batch size of 32. We select Adam as the optimizer and it is implemented with the Tensorflow. What is more, there is a dropout of 0.8 after the first fully-connected layer in each branch.

4. Experiments and Results

We make some ablation analysis to prove the advantages of each proposed module in MM-CNN. To make the comparison fair, we use the average and standard deviation of 10 checkpoints after convergence to measure the performance of the model.

For our self-made sensing array, there are two optical kinds of input. The first one is the 8×8 matrix which loses the spatial information and the other is the 12×12 matrix which recovers the spatial distribution of sensing units. We compare two inputs on the baseline structure and the structure with the global pooling layer. As shown in Table 2, we find that the input with spatial information performs better. In addition, although the structure with the global pooling layer cannot perform as well as the baseline model on the GSD, we decide to use it as it can process the input of different sizes. More details about the influence of the global pooling layer will be discussed in the following paragraphs. In general, we use the input with spatial information as the input of self-made sensing array in the following experiments.

Table 2. Comparisons of structures trained on the data w/o spatial information.

Model	Input Size	Avg $\pm$ Std Accuracy on GSD	Highest Accuracy on GSD
CNN baseline	8×8	94.59% $\pm$ 0.50%	95.29%
CNN baseline	12×12	95.45% $\pm$ 0.34%	95.98%
CNN baseline + global pooling	8×8	93.67% $\pm$ 0.76%	94.36%
CNN baseline + global pooling	12×12	93.99% $\pm$ 0.54%	94.60%

We train models on train sets to compare the influence of the global pooling layer on different test sets. We resize the input to the unified size of 16×12 by padding zeros in some experiments. In this way, the baseline model which does not include the global pooling layer can be also trained with the combination of GKD and GSD. On the one hand, this operation has little influence on the results of the experiments. On the other hand, we also pad them into the fixed size of 16×12 in cases where the padding operation is not necessary, to make the comparison fair. The results are shown in Table 3.

We find that deep learning models without global pooling which are trained with the GKD and GSD do not perform well when applying the models of one sensor to the inference of the other unseen dataset and the accuracies are lower than 70%. Additionally, the performances on the other sensor increase at least 10% when we add the global pooling layer. For the evaluation on the GSD, it even increases from 69.56% to 85.14%.

Accuracies are better when we let the model train on the GMD and then evaluate it. If we do not apply the operation of padding zero to the fixed size, we just utilize the model of baseline structure with the global pooling layer and train it with the GMD. In this case, the model comes out with the best checkpoint accuracy of (99.49%, 94.02%) on the test set of GDK and GSD. While it still has the space for improvement.

Table 3. Comparisons of structures w/o global pooling.

Model	Input Size	Train Set	Avg $\pm$ Std Accuracy on GKD	Avg $\pm$ Std Accuracy on GSD	Highest Accuracy on GKD and GSD
CNN baseline	16 $\times$ 12	GKD	99.36% $\pm$ 0.13%	69.56% $\pm$ 2.96%	(99.23%, 74.83%)
CNN baseline + global pooling	16 $\times$ 12	GKD	99.46% $\pm$ 0.08%	85.14% $\pm$ 2.78%	(99.23%, 89.89%)
CNN baseline	16 $\times$ 12	GSD	69.13% $\pm$ 3.70%	95.12% $\pm$ 0.34%	(76.21%, 94.71%)
CNN baseline + global pooling	16 $\times$ 12	GSD	79.00% $\pm$ 3.35%	94.36% $\pm$ 0.34%	(84.40%, 93.91%)
CNN baseline	16 $\times$ 12	GMD	99.03% $\pm$ 0.22%	92.85% $\pm$ 0.72%	(99.23%, 93.68%)
CNN baseline + global pooling	16 $\times$ 12	GMD	99.12% $\pm$ 0.21%	93.13% $\pm$ 0.59%	(99.23%, 93.91%)
CNN baseline + global pooling	16 $\times$ 10 and 12 $\times$ 12	GMD	99.28% $\pm$ 0.19%	93.41% $\pm$ 0.56%	(99.49%, 94.02%)

From this part, two fully-connected heads are applied in our experiments corresponding to two datasets of different sensing arrays. As mentioned before, we pretrain the tactile representation and one fully-connected head on one dataset and then freeze them. Then, we finetune the other head with the other dataset to get the whole model. Results in Table 4 show that the order of dataset does not influence a lot and it comes out with the best checkpoint accuracy of (99.49%, 94.25%) on the test set of GDK and GSD. It outperforms the single head structure as the dependence in each head is sufficient to maintain the classification accuracy. Additionally, it also shows that it is available to utilize the tactile representation with the weight sharing mechanism.

Table 4. Performance of multi fully-connected heads.

Model	Input Size	Train Set	(Avg $\pm$ Std) Accuracy on GKD	(Avg $\pm$ Std) Accuracy on GSD	Highest Accuracy on GKD and GSD
CNN baseline + global pooling	16 $\times$ 10 and 12 $\times$ 12	Pretrain on GKD	99.49%	86.09%	-
		Finetune on GSD	99.49%	94.20% $\pm$ 0.06%	(99.49%, 94.25%)
CNN baseline + global pooling	16 $\times$ 10 and 12 $\times$ 12	Pretrain on GSD	80.05%	94.25%	-
		Finetune on GKD	99.49% $\pm$ 0.00%	94.25%	(99.49%, 94.25%)

In addition, we design the mask branch to predict which head to use during inference. The mask prediction branch is trained with the resource label of GMD. As drawn in Figure 5, we analyze the embedding after global pooling in the mask branch. Embeddings of samples from GSD and embeddings of samples from GKD differ a lot and it is available to distinguish them. The mask predictions are all correct in the GKD and 99.89% of samples are classified correctly in the GSD. We combine this branch

with the grasping stability prediction branch to have an overall prediction. Although the mask branch classifies a sample of GKD incorrectly, we find that the two heads in the grasping stability prediction branch predict similar in this sample. Therefore, as shown in Table 5, the overall accuracy does not decrease. In general, the whole network keeps the best checkpoint accuracy of (99.49%, 94.25%) on the test set of GDK and GSD. This result achieves a comparable level on both datasets with those separate models (with the result of (99.49%, 94.02%)) in Table 3, while this structure can lead to an end-to-end inference.

Figure 5. T-distributed stochastic neighbor embedding analysis of embedding after global pooling in mask branch.

Table 5. Overall results of MM-CNN trained on grasping merged dataset (GMD).

Model	Input Size	Train Set	(Avg $\pm$ Std) Accuracy on GKD	(Avg $\pm$ Std) Accuracy on GSD	Highest Accuracy on GKD and GSD
Mask branch	16×10 and 12×12	GMD	100.00% $\pm$ 0.00%	99.82% $\pm$ 0.06%	(100.00%, 99.89%)
MM-CNN	16×10 and 12×12	Pretrain on GKD	99.49%	94.25%	-
MM-CNN	16×10 and 12×12	Pretrain on GSD	99.49%	94.25%	-

5. Conclusions

We use a convolutional neural network as the baseline to predict the grasping stability through the output of tactile sensing array. In order to propose a unified tactile representation and an end-to-end prediction system, we extend the framework of baseline with the global pooling layer, multiple fully-connected heads, and the mask prediction branch. The global pooling layer can deal with variable input sizes and multiple fully-connected heads can deal with different characteristics of sensors. The feature after the global pooling layer performs well when it is transferred to the grasping stability prediction of other sensors and we define it as the tactile representation. In addition, these two optimizations also have the influence of improving classification accuracy. The mask prediction branch determines which fully-connected to be selected during inference and makes the whole system work in an end-to-end manner. Finally, our proposed MM-CNN achieves the high accuracy of (99.49%, 94.25%) on GKD and GSD.

Author Contributions: Conceptualization, X.S., C.L., and T.L.; methodology, X.S.; validation, X.S.; writing—original draft preparation, X.S.; writing—review and editing, X.S., C.L., and T.L.; visualization, X.S.; supervision, C.L. and T.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 61802363 and 61774157, Key Research Program of Frontier Sciences, CAS, grant number QYZDY-SSW-JSC037, and the Natural Science Foundation of Beijing, grant number 4182075.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Johansson, R.; Westling, G. Signals in tactile afferents from the fingers eliciting adaptive motor responses during precision grip. *Exp. Brain Res.* **1987**, *66*, 141–154.
2. Teshigawara, S.; Tsutsumi, T.; Shimizu, S.; Suzuki, Y.; Ming, A.; Ishikawa, M.; Shimojo, M. Highly sensitive sensor for detection of initial slip and its application in a multi-fingered robot hand. In Proceedings of the 2011 IEEE International Conference on Robotics and Automation (ICRA), Shanghai, China, 9–13 May 2011; IEEE: Piscataway, NJ, USA, 2011; pp. 1097–1102.
3. Yuan, W.; Li, R.; Srinivasan, M.A.; Adelson, E.H. Measurement of shear and slip with a GelSight tactile sensor. In Proceedings of the 2015 IEEE International Conference on Robotics and Automation (ICRA), Seattle, WA, USA, 26–30 May 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 304–311.
4. Lin, C.H.; Erickson, T.W.; Fishel, J.A.; Wettels, N.; Loeb, G.E. Signal processing and fabrication of a biomimetic tactile sensor array with thermal, force and microvibration modalities. In Proceedings of the 2009 IEEE International Conference on Robotics and Biomimetics (ROBIO), Guilin, China, 19–23 December 2009; pp. 129–134.
5. Dahiya, R.S.; Valle, M. Tactile sensing technologies. In *Robotic Tactile Sensing*; Springer: Dordrecht, The Netherlands, 2013; pp. 79–136.
6. Alfadhel, A.; Kosel, J. Magnetic Nanocomposite Cilia Tactile Sensor. *Adv. Mater.* **2015**, *27*, 7888–7892.
7. Kaboli, M.; De La Rosa, A.; Walker, R.; Cheng, G. In-hand object recognition via texture properties with robotic hands, artificial skin, and novel tactile descriptors. In Proceedings of the 2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids), Seoul, Korea, 3–5 November 2015; pp. 1155–1160.
8. Luo, S.; Mou, W.; Althoefer, K.; Liu, H. Novel Tactile-SIFT Descriptor for Object Shape Recognition. *IEEE Sens. J.* **2015**, *15*, 5001–5009.
9. Veiga, F.; van Hoof, H.; Peters, J.; Hermans, T. Stabilizing novel objects by learning to predict tactile slip. In Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, 28 September–2 October 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 5065–5072.
10. Su, Z.; Hausman, K.; Chebotar, Y.; Molchanov, A.; Loeb, G.E.; Sukhatme, G.S.; Schaal, S. Force estimation and slip detection/classification for grip control using a biomimetic tactile sensor. In Proceedings of the 2015 IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids), Seoul, Korea, 3–5 November 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 297–303.
11. Gao, Y.; Hendricks, L.A.; Kuchenbecker, K.J.; Darrell, T. Deep learning for tactile understanding from visual and haptic data. In Proceedings of the 2016 IEEE International Conference on Robotics and Automation (ICRA), Stockholm, Sweden, 16–21 May 2016; pp. 536–543.
12. Zapata-Impata, B.S.; Gil, P.; Torres, F. Tactile-Driven Grasp Stability and Slip Prediction. *Robotics* **2019**, *8*, 85.
13. Garcia-Garcia, A.; Zapata-Impata, B.S.; Orts-Escolano, S.; Gil, P.; Garcia-Rodriguez, J. TactileGCN: A Graph Convolutional Network for Predicting Grasp Stability with Tactile Sensors. In Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14–19 July 2019.
14. Qin, J.; Liu, H.; Zhang, G.; Che, J.; Sun, F. Grasp stability prediction using tactile information. In Proceedings of the 2017 2nd International Conference on Advanced Robotics and Mechatronics (ICARM), Hefei, China, 27–31 August 2017; IEEE: Piscataway, NJ, USA, 2018.
15. Li, T.; Zheng, S.; Shu, X.; Wang, C.; Liu, C. Self-Recognition Grasping Operation with a Vision-Based Redundant Manipulator System. *Appl. Sci.* **2019**, *9*, 5172.