

# Learning Gentle Grasping from Human-Free Force Control Demonstration

Mingxuan Li<sup>1</sup>, Graduate Student Member, IEEE, Lunwei Zhang<sup>1</sup>, Graduate Student Member, IEEE,  
Tiemin Li<sup>1</sup>, and Yao Jiang<sup>1</sup>, Member, IEEE

**Abstract**—Humans can steadily and gently grasp unfamiliar objects based on tactile perception. Robots still face challenges in achieving similar performance due to the difficulty of learning accurate grasp-force predictions and force control strategies that can be generalized from limited data. In this article, we propose an approach for learning grasping from ideal force control demonstrations, to achieve similar performance of human hands with limited data size. Our approach utilizes objects with known contact characteristics to automatically generate reference force curves without human demonstrations. In addition, we design the dual convolutional neural networks (Dual-CNN) architecture which incorporating a physics-based mechanics module for learning target grasping force predictions from demonstrations. The described method can be effectively applied in vision-based tactile sensors and enables gentle and stable grasping of objects from the ground. The described prediction model and grasping strategy were validated in offline evaluations and online experiments, and the accuracy and generalizability were demonstrated.

## I. INTRODUCTION

Humans are skilled in grasping objects. Through tactile perception, humans can grasp objects of different shapes and textures stably and safely [1]. Even if the material and weight of the object are unknown, we can quickly adjust the applied grasping force by estimating contact stability [2], [3]. Usually, the grasping force used unconsciously by humans is not much larger than the minimum grasping force (by no more than 60%, according to [4]). Gentle grasping ensures a moderately sized grasping force that can lift an object stably without destroying the object and limiting dexterity [5].

Implementing gentle grasping in robotic systems is challenging [6]. In recent years, the rapid development of tactile sensing technology has provided a promising path for robotic grasping [7], [8]. Tactile sensors can measure objects' surface characteristics and contact states for inferring grasp stability [9]. Among the existing tactile sensors, vision-based tactile sensors are valued for their ability to offer rich tactile information [10]. Techniques based on photometric stereo [11] or marker displacement methods [12] can effectively capture geometric information of the contact surface, which can be used to compute various tactile features. These sensors offer two promising ways to achieve gentle grasping: slip-based and force-based approaches.

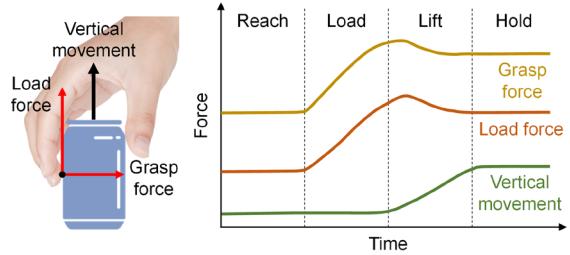


Fig. 1. A sequence of action phases during the process of grasping objects [1]. During the load phase, the object's vertical movement is very small, but the grasp force and load force have increased to near their peak values.

Regarding slip-based methods, Dong *et al.* predicted slip using the deviation of rigid-body motion fitting [13]. T-MO used a support vector machine (SVM) to respond to stabilize the grasping [14]. The GelStereo sensor employed a bimodal ConvLSTM network to learn the slip-feedback adaptive control strategy [15]. Zhao *et al.* predicted the grasp reliability using normalized differential convolution (NDC) [16]. Lu *et al.* proposed a self-attention slip detection method based on spatiotemporal fusion [17]. Since the transition from stable contact to macro slip occurs gradually when an object is held, the area ratio of the sticking region to slipping region, known as the stick ratio, is often used to estimate the safety margin of grasping [18]-[20]. However, Sui *et al.* demonstrated that using only the stick ratio to assess incipient slip is insufficient [21]. More detailed information regarding friction and force needs to be supplemented to enhance the quantification.

Force-based methods focus on direct indicators of slip or contact modeling [22]. For vision-based tactile sensing approach, force-related indicators are usually achieved through end-to-end methods. Griffa *et al.* used a UNet model to learn the 3-D distributed force during contact and developed a physics-based slip detection and force control method [23]. Boonstra *et al.* employed ShuffleNetV2 to learn the safety margin from tactile images of the ChromaTouch sensor, enabling gentle grasp force control [24]. Since force distribution is often obtained through data-driven methods, the sensitivity of the training data always makes consistent data collection challenging. Physics-based modeling is required to decouple the sensor's contact mechanics from the task-related tactile feedback model.

Therefore, the ability to achieve gentle grasping still needs improvement. Unlike maintaining grasp force or actively controlling in-hand object pivoting, the grasp-lift task from the ground involves three stages, as shown in Fig. 1. Starting from contact with the object, the grasp force needed to hold the object and the load force required to lift the object gradually

\*This work was supported by the National Natural Science Foundation of China under Grant 52375017. <sup>1</sup>Mingxuan Li, Lunwei Zhang, Tiemin Li, and Yao Jiang (corresponding author) are with the Institute of Manufacturing Engineering, Department of Mechanical Engineering, Tsinghua University, Beijing 100084, China (e-mail: [mingxuan-li@foxmail.com](mailto:mingxuan-li@foxmail.com); [zlw21@mails.tsinghua.edu.cn](mailto:zlw21@mails.tsinghua.edu.cn); [litm@mail.tsinghua.edu.cn](mailto:litm@mail.tsinghua.edu.cn); [jiangyao@mail.tsinghua.edu.cn](mailto:jiangyao@mail.tsinghua.edu.cn)).

increase. During this process, the object is slightly lifted, but due to the small vertical movement (usually only a few millimeters), the geometric changes on the contact surface are not significant. Mechanical features (friction and distributed force) are required in such situations. To enable a quick grasp, it is necessary to predict the target force (within 60% of the safety margin) to fully lift the object as early as possible. Also, the grasp force cannot be too small to ensure stability.

The combined demands for speed and stability necessitate providing the robot with expert annotations for the entire process (i.e., a suitable grasp force control curve). The most common method for annotating data relies on manual human intuition [16]. However, designing the curve and collecting the data is often labor-intensive. Unlike detecting whether an object is slipping, expert data for gentle grasping tasks need to include the target force at each moment, which cannot be achieved manually. An automated method for generating reference force control curves is needed to achieve timely and effective grasp force control.

This article proposes a strategy for learning gentle grasping through a human-free force control demonstration. This is a scheme for generating lightweight models. It utilizes pre-measured frictional properties of objects to construct an ideal force control demonstration, which is used for the robot to perform behavior cloning. This strategy can automatically generate expert trajectories without human involvement. Besides, to reduce training data while maintaining prediction accuracy, we incorporate a physics-based mechanical module into the training network. The force distribution obtained through the finite element method is not dependent on specific contact conditions, and is more sensitive to changes in grasping force and safety margins. As a result, gentle and stable grasping is achieved with a limited amount of data, and the model can generalize to unfamiliar objects.

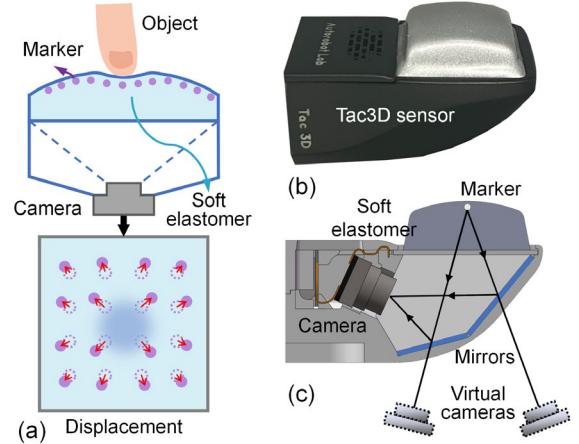
The remainder of this article provides the following:

- The developed sensor and force reconstruction method.
- The control strategy for generating demonstrations and the network to learn from the demonstrations.
- Ablation studies and online experiments that evaluate the effectiveness of gentle grasping control.

## II. MATERIALS AND METHODS

### A. Tactile sensing and force reconstruction

We customized the vision-based tactile sensor, Tac3D [25], as the robot finger. Fig. 2(a) illustrates the principle of vision-based tactile sensors. When an external object contacts the soft elastomer of the sensor, the displacement of markers on the contact surface can be used to discretely sample contact deformations. The structure of Tac3D sensor is shown in Fig. 2(b). An internal camera (1920×1080 pixels) continuously captures the deformation of the elastomer at a frequency of 30 Hz. The surface of the elastomer is engraved with a 20×20 density marker pattern. Tac3D employs a virtual binocular vision system [26], as shown in Fig. 2(c). The key feature of this design is the use of mirrors to split the reflected light from the marker pattern into two paths. Similar designs can also be



**Fig. 2.** (a) Marker displacement method used in vision-based tactile sensors. (b) Tac3D tactile sensor. (c) Virtual binocular vision system (VBVS) used in Tac3D.

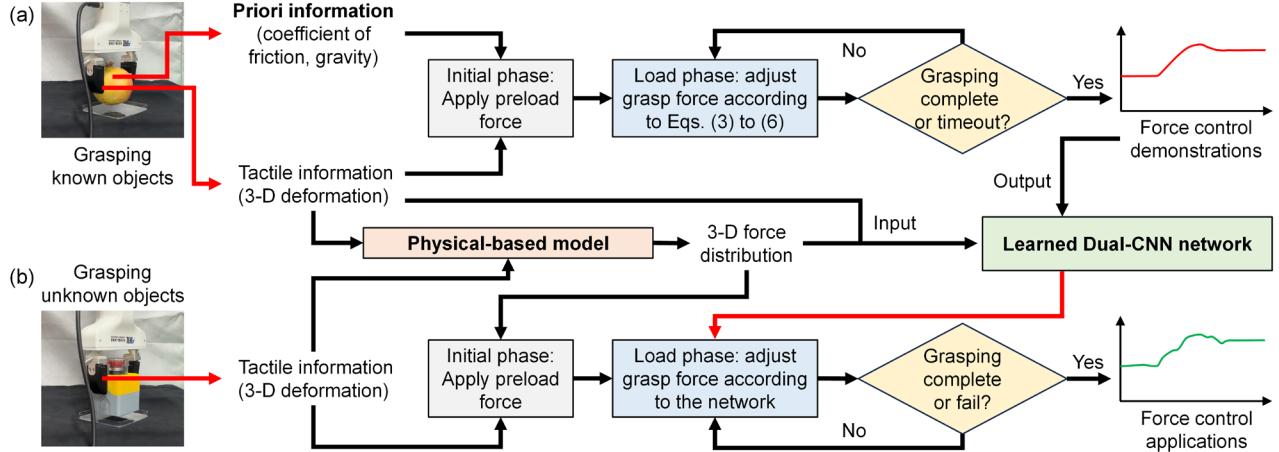
used to extend the sensor's effective range [27]. This solution ensures a balance between measurement accuracy, structural compactness, and synchronized triggering of the sensor.

Extracting distributed force information from deformation data has been shown to improve the assessment of grasping stability, since features with small relation to contact and friction are suppressed. Tac3D constructs a mapping from 3-D deformation to 3-D force distribution based on finite element methods [28], and incorporates various optimization algorithms and calibration techniques [29] to refine the finite element model. The physics-based model is minimally affected by the contact objects and conditions, and does not require training for specific grasping objects.

### B. Generation of Force Control Demonstrations

This article focuses solely on the two-finger parallel gripper as the end-effector. This type of gripper has only one degree of freedom for adjusting the grasp force. Due to its simple controllability, it is commonly used in research on grasp force control. For the grasping process starting from the ground, as shown in Fig. 1, grasp force adjustment mainly occurs during the load phase. A basic strategy is to control the grasp force during the load phase based on the currently measured load force and its increment, making the two forces match (i.e., always ensuring stable contact). During the transition between the load and lift phases, the object is already completely off the ground. At this point, the gripper should have reached the desired appropriate grasp force (i.e., the product of the safety margin and the minimum grasp force). After entering the lift phase, the gripper maintains the appropriate grasp force obtained in the load phase, and the robotic arm quickly lifts the object to complete the grasping.

Considering deformable fingers, the interaction between the object and the fingers during contact includes the normal force  $F_n$ , the tangential force  $F_t$ , and the torque  $M$ . For each position on the contact surface, select an element  $i$  and assume that the micro-element tangential and normal forces are  $f_t^i$  and  $f_n^i$ , respectively. According to Coulomb's friction, the condition for element  $i$  not to undergo local slip is:



**Fig. 3.** Learning gentle grasping from human-free force control demonstration. (a) Generation of force control demonstrations. (b) Force control based on the Dual-CNN network learned from demonstrations.

$$\mu > f_t^i / f_n^i, \quad (1)$$

where  $\mu$  is the coefficient of friction. According to [21], the contact factor can be defined using the ratio of micro-element resultant tangential force and normal force:

$$cf = 1 - \frac{\int f_t^i \cdot ds}{\int \mu f_n^i \cdot ds} = 1 - \frac{F_t}{\mu F_n}. \quad (2)$$

The value of  $cf$  ranges between 0 and 1. Ideally, all elements on the contact surface should be bonded. However, in practice, incipient slip always exists. Therefore, the closer the  $cf$  coefficient is to 0, the closer the contact state is to macro slip.

The contact factor can effectively describe the slip degree and can be applied to grasping control. However, the friction coefficient of the object to be grasped is always unknown. We cannot be satisfied with allowing the robot to grasp only familiar objects. Even if the friction coefficient is measured online through tactile sensing during the grasping, it usually takes tens of seconds [21]. However, the contact factor is inherently suitable for generating demonstrations to guide the robot in learning to grasp. For objects with known friction coefficients and other prior information, grasp force control curves can be generated by maintaining a certain contact factor to avoid macro slip during the load phase. These curves ensure the grasp and load forces maintain an appropriate ratio (i.e., a safety margin) throughout. By using such curves as demonstrations, the robot can learn the force-following strategy to grasp objects with similar friction characteristics.

Based on the above discussion, we propose a strategy for generating force control demonstrations, as shown in Fig. 3(a). This strategy is not directly applied to grasping unknown objects but is used to collect force control data for grasping known objects, which the robot can then learn from. This concept is similar to imitation learning (also known as programming by demonstration) [30], but without human involvement, as the robot generates the expert data.

The first step involves the meticulous measurement of the friction coefficient of the known objects in the training set.

The method is as follows: the object is grasped with a constant force and lifted from the ground, then suspended in the air for 5 seconds. If no noticeable slip is observed during this process, the grasp force is slightly reduced. This process is repeated until the minimum grasping force is found. The ratio of the tangential force to the normal force at this point is precisely defined as the measured friction coefficient. Next, grasp force control is based on the following procedure: the robot contacts the object with a preload force  $F_g^0 = 0.4N$ . Then, during the load phase, the object is lifted slowly at a speed of 1 mm/s for 3 or 4 seconds. Finally, in the lift phase, the object is rapidly lifted at a constant grasp force (the final grasp force from the load phase) at a speed of 10 mm/s. At frame  $k$ , the tactile sensor measures the concentrated force information  $F_t^k$  and  $F_n^k$ . Then, at frame  $k + 1$ , the robot needs to attempt to control the grasp force  $F_g^{k+1}$  (which is different from  $F_n^{k+1}$ ) so that:

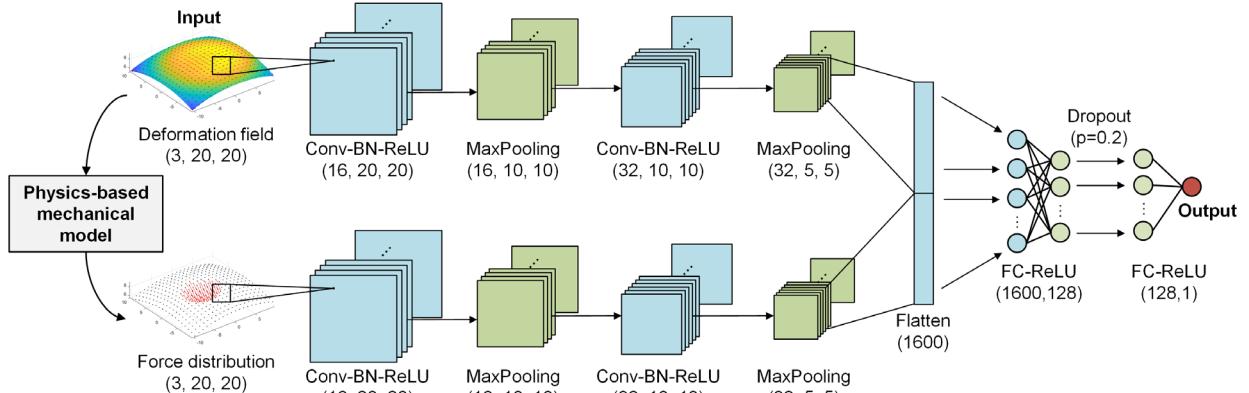
$$F_n^{k+1} = \beta \mu^{-1} \cdot \max(F_t^k, F_m^{k-1}), \quad (3)$$

where  $F_m^{k-1}$  represents the maximum value in  $F_m^i$  and before frame  $k + 1$ .  $F_t^k$  and  $F_n^k$ 's can be computed by resolving the normal direction of each micro-element.

In Eq. 3,  $\beta$  represents the safety margin. Initially, the grasping force is relatively small, and the tangential force increases rapidly, requiring a larger safety margin; as the grasping force gradually approaches the final target value and the tangential force increase rate decreases, a smaller safety margin is needed to prevent overshooting. Therefore, the safety margin is set as a time-dependent function:

$$\beta = \begin{cases} \beta_{\max}, & \text{if } t \geq t_m \\ \beta_{\min} + \frac{\beta_{\max} - \beta_{\min}}{1 + \exp(-k \cdot (t - t_{\text{bias}}))}, & \text{if } t < t_m \end{cases}. \quad (4)$$

Eq. (4) uses a sigmoid function to non-linearly map the safety margin between  $\beta_{\min}$  and  $\beta_{\max}$ . The parameters  $\alpha$  and  $t_{\text{bias}}$  control the shape of the sigmoid curve, thereby adjusting the strategy for selecting the safety margin. Here,  $t$  and  $t_m$  represent the camera frame count. The parameters were set as:



**Fig. 4.** Diagram of the Dual-CNN network for target force prediction based on the mechanical model. Conv-BN-ReLU: convolution layer, batch normalization layer, and ReLU activation layer. MaxPooling: max-pooling layer. Flatten: flatten layer. FC-ReLU: fully connected layer and ReLU activation layer.

$$\begin{aligned} \beta_{\max} &= 2, \quad \beta_{\min} = 1.2, \quad k = 0.1, \\ t_{\text{bias}} &= 20, \quad t_m = 180. \end{aligned} \quad (5)$$

One remaining issue is the difference between  $F_g^{k+1}$  and  $F_n^{k+1}$ , especially when the contact surface is not flat. An effective solution is to dynamically estimate the linear relationship based on historical information [21]:

$$F_g^{k+1} = \begin{cases} F_g^k, & \text{if } F_n^{k+1} = F_n^k \\ F_g^k + (F_g^k - F_g^{k-1}) \cdot \frac{F_n^{k+1} - F_n^k}{F_n^k - F_n^{k-1}}, & \text{if } F_n^{k+1} \neq F_n^k \end{cases}. \quad (6)$$

For the same object, force control curves implemented according to the above strategy may vary little. To further enhance the diversity of the training data, we randomly vary the initial preload force within the range of 0.3 to 0.5 and adjust the grasping positions on the same object. Additionally, to eliminate the influence of the lifting speed on force control, we uniformly sample from the original data sequence to adjust the time intervals between adjacent data points. Ultimately, the implemented demonstrations cover as many feasible paths as possible to ensure sufficient diversity in the training data.

### C. Target Force Prediction

One advantage of prior-based force control demonstration is that it reduces the need to focus on temporal information. Since the generated reference force control follows a fixed strategy, the collected data primarily represents the contact states along the target path, avoiding excessive trivial values and states that would not be reached in a successful grasp. In other words, subtle patterns and dependencies related to temporal sequences are already included in the demonstration. Moreover, the physics-based reconstruction of distributed forces inherently includes direct indicators of slip and contact stability (similar to the contact factor), which can better enhance temporal changes' effects than raw deformation data. Thus, the training can focus on extracting spatial features.

The goal of training is to predict the target grasping force to be controlled at the next moment based on the data collected

by the tactile sensor at the current moment. The overall architecture of the proposed target force prediction network is shown in Fig. 4. The input information to the network is the contact deformation measured by Tac3D (stored in the form of a 3-D coordinate field). We first obtain the displacement field by subtracting the initial coordinate field from the current coordinate field and then map it to a 3-D distributed force field based on the physics-based mechanical method described in Section II-A. The region where the normal force exceeds the threshold value (set to 0.01 mm) is considered as the contact region. We suppress the values of the distributed force outside the contact region using graphics operations to reduce measurement noise at non-contact region. At the  $k$ -th moment, the coordinate and force fields are flattened to form samples:

$$\mathbf{p}^k = \text{reshape}(\mathbf{x}^k) \in \mathbb{R}^{3 \times 20 \times 20}, \quad (7)$$

$$\mathbf{f}^k = \text{reshape}(h^*(\mathbf{x}^k - \mathbf{x}^0)) \in \mathbb{R}^{3 \times 20 \times 20}, \quad (8)$$

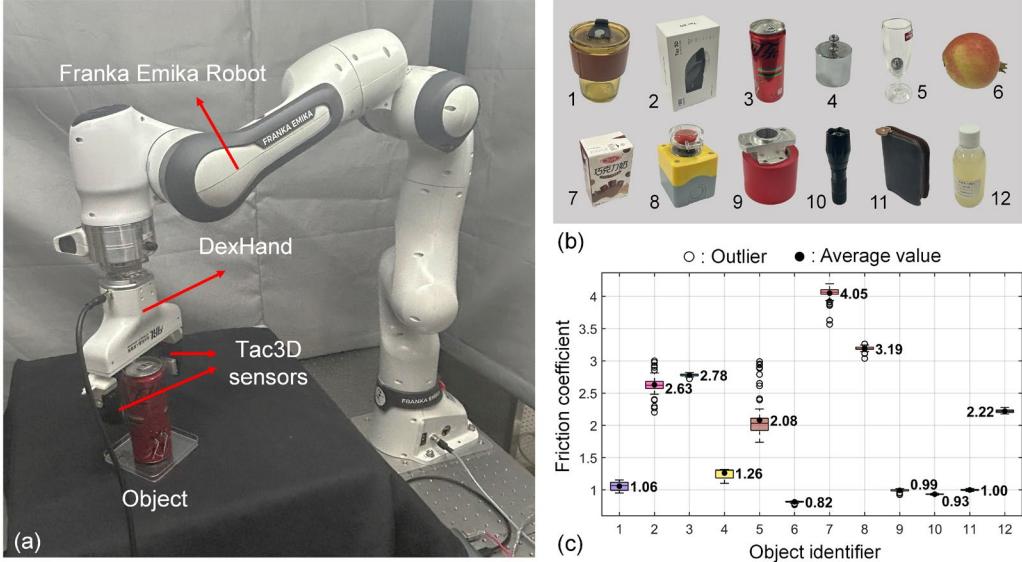
where the reshape(:) operation encodes the tactile features into a two-channel image of  $20 \times 20$  size.

The coordinate and force samples are coordinated in the same format and fed into a Dual-CNN to extract features. The main reason for using CNNs here is to ensure coordinate translation invariance. The CNN module comprises a stack of 2 layers of Conv-BatchNorm-ReLU (CBR) and two pooling layers. Each pooling operation reduces the feature map size by half. The CBR stack includes 2D convolution operation, batch normalization, and ReLU activation. The convolution layers use  $3 \times 3$  kernels with zero padding (padding=1). Let  $X$  be the input, then the output after one CBR operation is:

$$F = \max \left( \gamma \frac{(W * X + b) - \mu}{\sqrt{\sigma^2 + \varepsilon}} + \beta, 0 \right), \quad (9)$$

where  $W$  is the convolutional kernel weight,  $*(:)$  is the convolutional operation,  $b$  is the convolutional bias.  $\mu, \sigma^2, \gamma$ , and  $\beta$  are the parameters in the batch normalization.

After extracting the features, the output features of the two CNN sub-networks are flattened and spliced:



**Fig. 5.** (a) Experimental platform. (b) Objects used for data collection. Objects numbered 1 through 6 were used to train the model, and objects numbered 8 through 12 were used for online grasping experiments. (c) Mean and standard deviation of the friction coefficient estimated for each object. The horizontal coordinate is the serial number of each object corresponding to Fig. 5(b).

$$F = \text{concat} \left( \text{flatten}(\mathbf{y}_p^k), \text{flatten}(\mathbf{y}_f^k) \right), \quad (10)$$

where  $\text{flatten}(\cdot)$  denotes the operation of converting multi-dimensional input into a 1-D vector, and  $\text{concat}(\cdot)$  refers to concatenating operation.  $\mathbf{y}_p^k$  and  $\mathbf{y}_f^k$  are the feature maps obtained from the two CNN sub-networks, respectively. The concatenated features are mapped to the target force prediction using a multi-layer perceptron (MLP). A dropout operation with a rate of 0.2 is applied to prevent overfitting.

Finally, the online grasping force control is implemented based on the control method shown in Fig. 3(b). First, a preload force of 0.4 N is applied to the object. During the load phase, the robotic arm slowly lifts the object vertically at a speed of 1 mm/s. Meanwhile, the data collected by the Tac3D sensor is fed into the Dual-CNN model to generate the current target grasping force. The gripper adjusts and maintains the grasp force until the sensor gets new data. The control lag caused by algorithm processing time and communication delay does not exceed 50 ms. When the grasping time reaches 4 seconds, or if the measured tangential force does not change significantly within 0.2 seconds, The final target grasp force is considered to have been reached. The lift phase is then completed quickly at a speed of 10 mm/s.

### III. EXPERIMENTS AND RESULTS

#### A. Experimental Setup

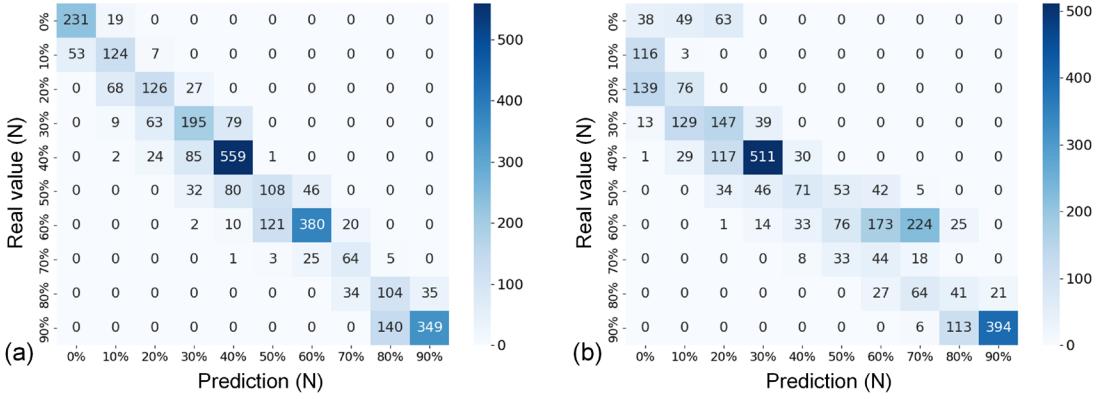
This article conducts two sets of experiments: one involving training the proposed model and ablation studies using offline data, and the other evaluating online grasping performance based on the trained model. The experimental platform is shown in Fig. 5(a). We used a robotic platform from Franka Emika and a custom-designed DexHand gripper. DexHand features good force-following performance for

TABLE I  
PERFORMANCE COMPARISON OF DIFFERENT METHODS

Models	MSE test loss / N <sup>2</sup>	Accuracy score
CNN (trained with deformation only)	0.0236	0.2573
Dual-CNN (trained with deformation and force)	<b>0.0037</b>	<b>0.6933</b>

precise and timely force control. Considering computational efficiency, only the readings from one Tac3D sensor were used. In the offline evaluation, all models were implemented using PyTorch, and the training was performed on a laptop with a 2.30 GHz Intel i7-12700H processor and a GeForce 3060 GPU (6 GB). For the online experiments, the motion control of robot arm, grasp force control, tactile feedback and calculations, and the force prediction pipeline, were run on an Ubuntu 18.04 PC equipped with a 3.6 GHz AMD Ryzen 5 processor and a GeForce 1060 GPU (3 GB).

Twelve household objects with different shapes, materials, and stiffness were used to collect data, as shown in Fig. 5(b). Fig. 5(c) shows the friction coefficients of these objects, measured using the method described in Section II-C. Objects 1 to 7 were used to generate demonstrates and train the model. 100 grasping experiments were conducted for each object. Each grasping process involved capturing 90-120 frames of tactile information. 81,371 sets of 3-D deformation fields (inputs) and corresponding target forces (labels) were collected. We used 80% of this data for training and 20% for testing. The network is trained using stochastic gradient descent (SGD) with a learning rate of  $10^{-6}$ . Mean squared error (MSE) loss function and Adam optimizer are used. The training involves 50 epochs with a batch size of 16, and network weights are initialized using the Xavier normal distribution. A ReduceLROnPlateau scheduler is utilized to adjust the learning rate to prevent overfitting.



**Fig. 6.** Confusion matrix of the real value versus the predicted value. (a) Results for the CNN network trained with deformation only. (b) Results for the Dual-CNN network trained with deformation and force. The more the data is concentrated on the diagonal, the better the model's predictions are.

Besides, objects numbered 8 to 12 were used to evaluate the performance of online grasp force regulation. Note that these objects' friction coefficients were used only for generating the ground truth target grasp force and were unknown to the robot. We added metal pieces on top of or inside lighter objects (e.g., objects 5 and 9 in Fig. 5(b)) to ensure they could not be easily lifted.

### B. Offline Evaluation

The ablation study consists of two experimental groups: the Dual-CNN structure used in this article and a model with only one activated CNN path. The difference between these models is that the former uses the 3-D coordinate field and the 3-D force distribution as inputs. At the same time, the latter removes the physics-based force reconstruction module (i.e., only the CNN path with the 3-D coordinate field is used). The second column of Table I shows the mean squared error loss (MSE Loss). The results indicate that physics-based force reconstruction module improves the model's accuracy, reducing the total MSE loss by nearly seven times.

We divided the true values of the target grasping force into ten intervals to evaluate the model's performance across the entire range, from 0% to 100% of the maximum prediction result. The confusion matrix is shown in Fig. 6. By comparing the diagonal trends in the figure, it can be seen that the target force estimated by the proposed model is positively correlated with the ground truth, and the model with force reconstruction module performs better. The third column of Table I shows the accuracy of the confusion matrix (i.e., the ratio of the number of elements on the diagonal to the total number). Note that since target force prediction is a regression task, this metric does not quantify the model's accuracy (such as precision, recall, and F1 score) but rather provides an intuitive comparison. The results above clearly demonstrate that the physics-based force reconstruction module effectively improves the accuracy of target grasping force prediction.

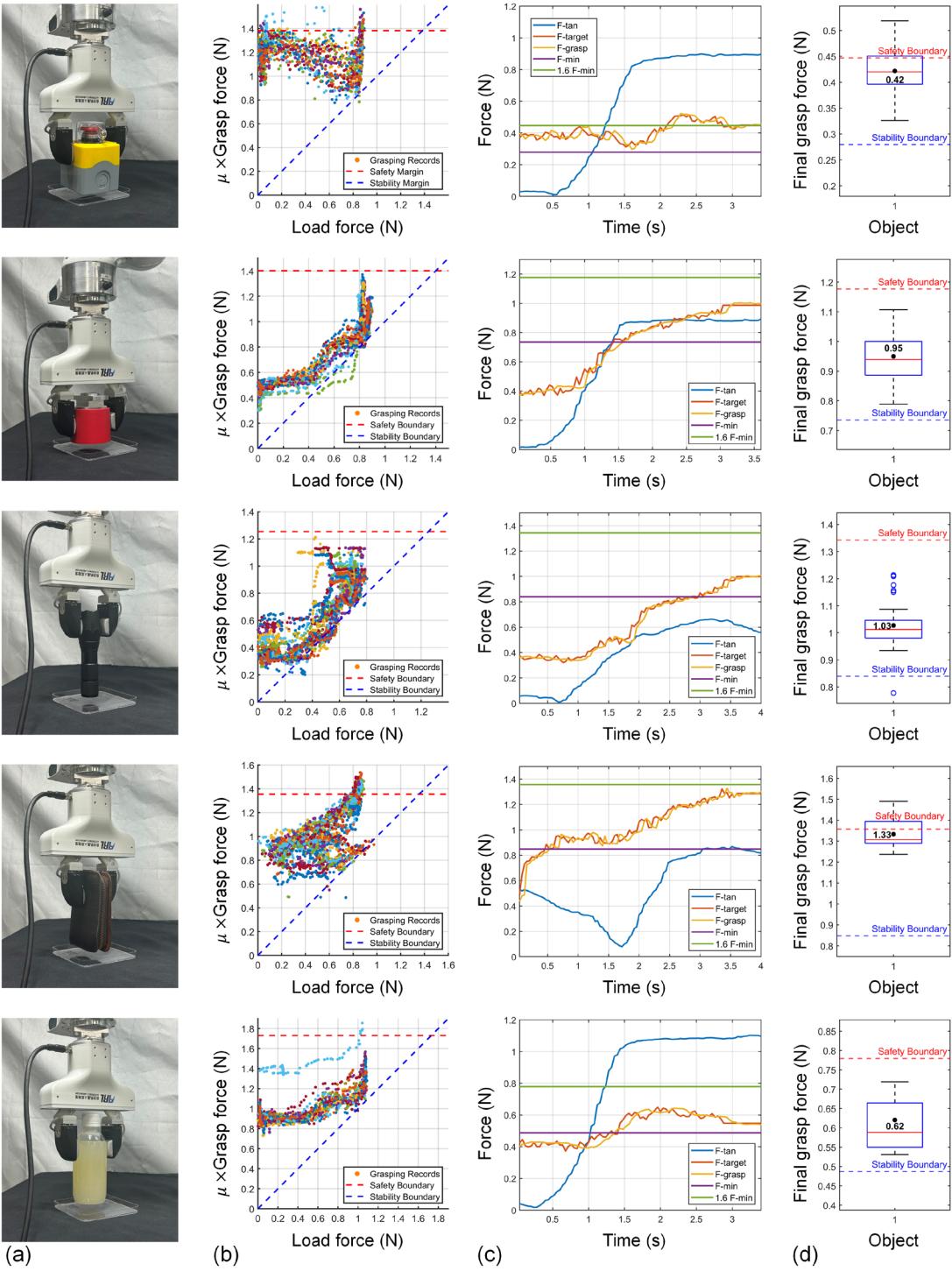
### B. Online Evaluation

We conducted 30 online grasping experiments on objects 8 to 12, respectively [see Fig. 7(a)]. The objects were unknown to the robot, and no safety margins were artificially set.

Therefore, the target force was automatically determined by the proposed model. Fig. 7(b) shows all experiments' load-force/grasp-force states during the load phase. The discrete points of the same color represent the force states during the same grasping trial. We define safety boundaries and stability boundaries to describe the upper limit to prevent excessive grasp forces and a lower limit to prevent falling events. Both boundaries are determined based on the pre-measured friction coefficients but are not known as the prior information.

The proposed method achieved the best performance on objects 10 (flashlight), 11 (wallet), and 12 (plastic bottle), with the force state remaining stable between the safety and stability boundaries throughout the process. For object 9 (tape), the predicted grasp force was relatively small since pivot rotation occurred during grasping (the training set only covered translational slip conditions). Object 10 (plastic block) was relatively light, and the ideal grasping force was close to the initial preload force. Thus, the force state was near the safety boundary in the early Load phase. Ultimately, out of all test cases, only object 9 experienced a single drop case (corresponding to the green discrete points in Fig. 7(b) for that particular grasping attempt). Under the conditions of this experiment, a grasping success rate of 99.33% was achieved.

Our purpose is not limited to successfully grasping objects but to grasping objects as gently as possible while ensuring stability. Fig. 7(c) shows the force curves for grasping five objects (randomly selected for one test). The experimental results show that the described method can quickly load to the final grasp force (between the safety boundary and the stability boundary) for different objects within 2~3 seconds. Compared to existing work [21], the described method eliminates the need to measure the contact characteristics of the object online, and can shorten the time of the loading phase. Despite some measurement noise and delay in the system, the proposed method effectively ensures that the change in grasp force remains synergistic with the change in load force. For object 10, the grasp force adjustment curve showed some overshoot in the middle segment due to the proximity of the ideal grasp force to the initial preload force. Eventually, the grasp force at the end of the load phase can still return to a proper value. Fig. 7(d) shows the final grasp force distribution



**Fig. 7.** Online evaluation of grasp force control. (a) Schematic of the grasping process. (b) Evaluation of load-force/grasp-force status during the whole process. The safety boundary is the product of the minimum grasp force required to pick up the object and a safety margin (set to 1.6 in the experiments), and the stability boundary is defined as the minimum force corresponding to the current load force. (c) Force curve during the whole process. F-tan: load force. F-target: target grasp force (network output). F-grasp: actual grasp force. F-min: minimum grasp force to lift the object. (d) Statistical results of the final grasp force.

for all sub-attempts. Except for the objects 9 and 11, which slightly exceeded the margin of safety, all other tests met the requirements for gentle and stable grasping. Considering that actual objects pass through without being broken by less than twice the minimum grasp force, such results are sufficient for the gentle grasping of most household objects.

#### IV. CONCLUSION

This article proposes a method based on learning force reference trajectories for achieving stable and gentle grasping. We focus on how to generate ideal reference cases for grasp force control, and how to use a lightweight network in

conjunction with physics-based force reconstruction for learning. Not just grasping the object, the described approach achieves grasping the object from the ground as gently as possible. The strategy can provide a more accurate prediction of the target force than methods based on physical modeling. Compared to end-to-end learning methods, the scheme has better adaptability and can generalize to unknown objects. Ablation experiments and online grasping tests demonstrate the effectiveness of the described process.

The limitation of the described method is the dependence on the specific grasping action. The idea of imitation learning, while helping to avoid the complexity of learning on long time-series information, makes the process applicable only to lifting an object vertically from the ground. For objects off the center of gravity that may rotate during the grasping, the network may need to be retrained to include more data related to rotational slip effects. In addition, for objects with particularly small coefficients of friction, the method may not be able to be tuned to the final target grasping force in a very short period. This phenomenon has also been mentioned in other works [25]. For example, when trying to grasp a frozen orange (friction coefficient of about 0.3), the preload force and lifting speed affected the method's reliability. Future work will consider the introduction of network modules capable of processing temporal information and explore techniques for generating differentiated force control demonstrations.

## REFERENCES

- [1] R. S. Johansson, and J. R. Flanagan, “Coding and use of tactile signals from the fingertips in object manipulation tasks,” *Nature Reviews Neuroscience*, vol. 10, no. 5, pp. 345–359, 2009.
- [2] F. Schiltz, B. P. Delhaye, J.-L. Thonnard, and P. Lefèvre, “Grip Force is adjusted at a level that maintains an upper bound on partial slip across friction conditions during object manipulation,” *IEEE Transactions on Haptics*, pp. 1–1, 2021.
- [3] B. P. Delhaye, F. Schiltz, F. Crevecoeur, J.-L. Thonnard, and P. Lefèvre, “Fast grip force adaptation to friction relies on localized fingerpad strains,” *Science Advances*, vol. 10, no. 3, 2024, Art. no. eadh9344.
- [4] A. M. Hadjiosif and M. A. Smith, “Flexible control of safety margins for action based on environmental variability”, *Journal of Neuroscience*, vol. 35, no. 24, pp. 9106–9121, 2015.
- [5] T. Bi, C. Sferrazza and R. D’Andrea, “Zero-shot sim-to-real transfer of tactile control policies for aggressive swing-up manipulation”, *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 5761–5768, 2021.
- [6] Z. Xie, X. Liang, and C. Roberto, “Learning-based robotic grasping: A review,” *Frontiers in Robotics and AI*, vol. 10, 2023, Art. no. 1038658.
- [7] Nathan F. Lepora, “The future lies in a pair of tactile hands” *Science Robotics*, vol. 9, no. 91, 2024, Art. no. eadq1501.
- [8] T. Li *et al.*, “A comprehensive review of robot intelligent grasping based on tactile perception,” *Robotics and Computer-Integrated Manufacturing*, vol. 90, 2024, Art. no. 102792.
- [9] Y. Bekiroglu, J. Laaksonen, J. A. Jorgensen, V. Kyrki, and D. Kragic, “Assessing Grasp Stability Based on Learning and Haptic Data,” *IEEE Transactions on Robotics*, vol. 27, no. 3, pp. 616–629, 2011.
- [10] S. Zhang *et al.*, “Hardware technology of vision-based tactile sensor: A review,” *IEEE Sensors Journal*, vol. 22, no. 22, pp. 21410–21427, Nov. 2022.
- [11] M. K. Johnson and E. H. Adelson, “Retrographic sensing for the measurement of surface texture and shape”, in *2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 1070–1077.
- [12] M. Li, L. Zhang, T. Li, and Y. Jiang, “Marker displacement method used in vision-based tactile sensors—from 2D to 3D: A review,” *IEEE Sensors Journal*, vol. 23, no. 8, pp. 8042–8059, 2023.
- [13] S. Dong, D. Ma, E. Donlon, and A. Rodriguez, “Maintaining grasps within slipping bounds by monitoring incipient slip,” in *2019 IEEE International Conference on Robotics and Automation (ICRA)*, May 2019, pp. 3818–3824.
- [14] J. W. James and N. F. Lepora, “Slip detection for grasp stabilization with a multifingered tactile robot hand,” *IEEE Transactions on Robotics*, vol. 37, no. 2, pp. 506–519, Apr. 2021.
- [15] S. Cui, S. Wang, R. Wang, S. Zhang, and C. Zhang, “Learning-based slip detection for dexterous manipulation using GelStereo sensing.” *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–10, 2023.
- [16] Z. Zhao, W. He, and Z. Lu, “Tactile-based grasping stability prediction based on human grasp demonstration for robot manipulation,” *IEEE Robotics and Automation Letters*, vol. 9, no. 3, pp. 2646–2653, March 2024.
- [17] J. Lu, B. Niu, H. Ma, J. Zhu and J. Ji, “STNet: Spatio-temporal fusion-based self-attention for slip detection in visuo-tactile sensors,” in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 3051–3056.
- [18] W. Chen, H. Khamis, I. Birznieks, N. F. Lepora, and S. J. Redmond, “Tactile sensors for friction estimation and incipient slip detection—Toward dexterous robotic manipulation: A review,” *IEEE Sensors Journal*, vol. 18, no. 22, pp. 9049–9064, Nov. 2018.
- [19] R. Sui, L. Zhang, T. Li, and Y. Jiang, “Incipient slip detection method with vision-based tactile sensor based on distribution force and deformation,” *IEEE Sensors Journal*, vol. 21, no. 22, pp. 25973–25985, Nov. 2021.
- [20] M. Li, Y. H. Zhou, T. Li, and Y. Jiang, “Incipient slip-based rotation measurement via visuotactile sensing during in-hand object pivoting,” in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 17132–17138.
- [21] R. Sui, L. Zhang, Q. Huang, T. Li, and Y. Jiang, “A novel incipient slip degree evaluation method and its application in adaptive control of grasping force,” *IEEE Transactions on Automation Science and Engineering*, pp. 1–10, 2023.
- [22] M. Costanzo, G. De Maria, and C. Natale, “Slipping control algorithms for object manipulation with sensorized parallel grippers,” in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 2018, pp. 3675–3681.
- [23] P. Griffa, C. Sferrazza, and R. D’Andrea, “Leveraging distributed contact force measurements for slip detection: A physics-based approach enabled by a data-driven tactile sensor,” in *2022 IEEE International Conference on Robotics and Automation (ICRA)*, May 2022, pp. 4826–4832.
- [24] D. -J. Boonstra, L. Willemet, J. Luijkx and M. Wiertlewski, “Learning to estimate incipient slip with tactile sensing to gently grasp objects,” in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 16118–16124.
- [25] L. Zhang, Y. Wang, and Y. Jiang, “Tac3D: A novel vision-based tactile sensor for measuring forces distribution and estimating friction coefficient distribution,” Feb. 2022, arXiv:2202.06211.
- [26] Y. Wang, L. Zhang, T. Li, and Y. Jiang, “A model-based analysis-design approach for virtual binocular vision system with application to vision-based tactile sensors,” *IEEE Transactions on Instrumentation and Measurement*, vol. 72, p. 5010916, 2023.
- [27] Y. Ma, J. A. Zhao and E. Adelson, “GelLink: A Compact Multi-phalanx Finger with Vision-based Tactile Sensing and Proprioception,” in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 1107–1113.
- [28] L. Zhang, T. Li, and Y. Jiang, “Improving the force reconstruction performance of vision-based tactile sensors by optimizing the elastic body,” *IEEE Robotics and Automation Letters*, vol. 8, no. 2, pp. 1109–1116, Feb. 2023.
- [29] M. Li, L. Zhang, Y. H. Zhou, T. Li and Y. Jiang, “EasyCalib: Simple and low-cost in-situ calibration for force reconstruction with vision-based tactile sensors,” *IEEE Robotics and Automation Letters*, vol. 9, no. 9, pp. 7803–7810, 2024.
- [30] K. Yamane, Y. Saigusa, S. Sakaino and T. Tsuji, “Soft and rigid object grasping with cross-structure hand using bilateral control-based imitation learning,” *IEEE Robotics and Automation Letters*, vol. 9, no. 2, pp. 1198–1205, Feb. 2024.