Impact of Physical Parameters and Vision Data on Deep Learning-based Grip Force Estimation for Fluidic Origami Soft Grippers

Eojin Rho[1*], Woongbae Kim[2,3,4*], Jungwook Mun[1], Sung Yol Yu[2], Kyu-Jin Cho[2†] and Sungho Jo[1†]

***Abstract*— Knowing the gripping force being applied to an object is important for improving the quality of the grip, as well as preventing surface damage or destruction of fragile objects. In the case of soft grippers, however, an attaching or embedding force/pressure sensors can compromise their softness and adaptability or increase the cost/complexity of the manufacturing process. In this paper, we present a vision-based neural network(*OriGripNet*) that can estimate gripping force by combining RGB image data with key parameters extracted from the physical features of a soft gripper. Real-world force data were collected using a reconfigurable test object with an embedded load cell while simultaneously image data were collected by a single RGB camera mounted on the wrist of a robotic arm. In addition, key geometry information of the pneumatically driven origami gripper extracted from the images, and applied pressure were further used for training of the developed model. The results show that key physical parameters and image information have their own strengths in force estimation, contact estimation, and adaptability to unseen objects, and that they have a synergistic effect on the performance when combined.**

I. INTRODUCTION

Enabling gentle and delicate interaction between robot end-effectors and objects is an essential issue in extending the practical use of robots. By incorporating sensory feedback into the robotic gripping system, the risk of dropping or damaging the objects during pick-and-place operations can be greatly reduced. Visual feedback is generally considered as top priority, and is used to recognize the types, shapes, and positions of objects as well as whether they are being held or not [1], [2]. It is implemented through the procedure of collecting visual data with a camera installed in a workspace or on the robot manipulator and processing the data using computer vision algorithms. On the other hand, force feedback of the gripping system is also important when it is required to prevent surface damage or breakage of vulnerable objects, and to enable a grip that is selectively robust or adaptive to changing conditions. In many cases, rigid force/pressure sensors are attached to the end-tips of the conventional parallel grippers and data is directly measured while positioned between the gripper and the object, offering the advantage of high precision and easy integration of the sensor into the gripper system.

Meanwhile, soft grippers composed of low-stiffness materials [3], [4] are intensively studied to solve challenging tasks that are difficult to handle with conventional systems. These tasks specifically require gentle grip and high adaptability, exemplified by pick-and-place of fragile objects [5], [6], application in food industry [7], [8], picking and inserting coin [9], universal gripping [10], [11], [12], human robot collaboration [13], and underwater operation that handling marine life [6], [14], [15]. Enabling grip force feedback for soft grippers would enhance their usability for the applications and mitigate the disadvantage of difficulty in control due to their inherent nonlinearity [16]. However, direct attachment of the conventional sensors at their soft end-tips eliminates the benefits of conformal and adaptive contact of soft grippers. Soft sensors are also being actively developed as their skin-like perception is suitable for flexible systems, but they are still immature for practical use because the shortcomings associated with robustness, hysteresis, degradation, integration cost, and interconnection to electronics have not been fully addressed [17]. Indirect contact sensing methods such as sensor embedding may not compromise the adaptiveness of the soft grippers [18]. However, these non-contact force sensing methods require a particular internal design and fabrication process of grippers, which limits the design of the gripper and increases the cost. Taken together, the development of a force measurement method that preserves both the conformal contact properties and structural simplicity of soft grippers is expected to increase their usability.

Recently, deep learning-based gripper force estimation methods are being studied to avoid the constraints caused by attaching a force sensor to an end-effector. For parallel grippers, studies that estimate the gripper force using deep learning networks with motor signals [19], vision data(deformable Fin Ray Grippers) [20], [21], or both [22] as inputs have been conducted for several years. In case of soft pneumatic grippers, on the other hand, research of deep learning-based force estimation is in its early stages with only a few studies. To predict the contact force of the soft pneumatic actuator, Thuruthel et al. [23] and Loo et al. [24] used a recurrent neural network(RNN), taking pressure values and embedded strain/flex sensor data as inputs. Ang and Yeow predicted

This work was supported by the National Research Foundation of Korea(NRF) Grant funded by the Korean Government(MSIT)(RS-2023-00208052) and Korea Institute of Science and Technology Institutional Program grant 2E32304.

[1]E. Rho, J. Mun, S. Jo are with Neuro-Machine Augmented Intelligence Lab, School of Computing, KAIST, Daejeon, 08826, Republic of Korea

[2]W. Kim, S. Y. Yu, and K. J. Cho are with Biorobotics Laboratory, Department of Mechanical Engineering, Seoul National University, Seoul, 08826, Republic of Korea

[3]W. Kim is with Artificial Intelligence and Robotics Institute, Korea Institute of Science and Technology(KIST), Seoul, 02792, Republic of Korea

[4]W. Kim is with Korea Institute of Science and Technology Europe(KIST-EUROPE), 66123, Saarbrücken, Germany

**Eojin Rho and Woongbae Kim contributed equally to this work.*

(†Corresponding authors: Sungho Jo and Kyu-Jin Cho.)

(email :shjo@kaist.ac.kr and kjcho@snu.ac.kr)

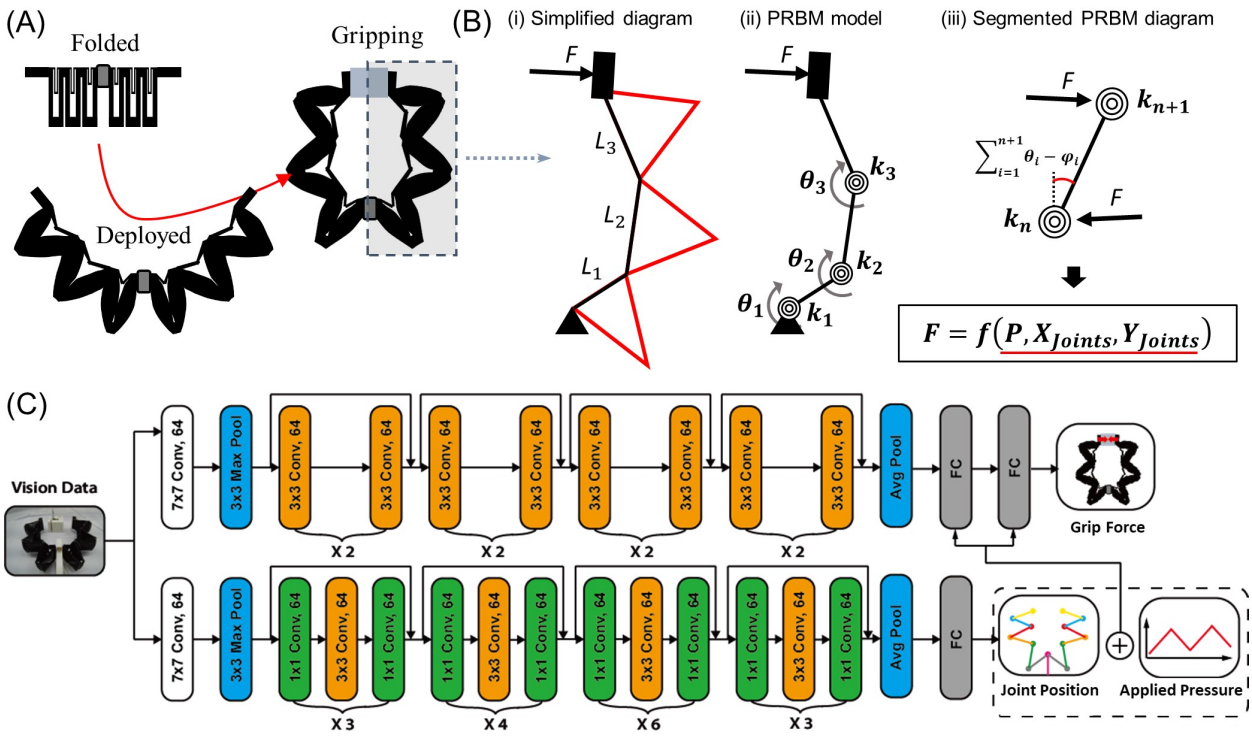

Fig. 1. The dual-origami soft gripper and *OriGripNet* for gripping force estimation. A) The deployment and bending motion of the dual-origami soft gripper. B) Simplified diagram and PRBM model of one finger of a double origami soft gripper when an external force is applied. C) The architecture of *OriGripNet* model.

a contact force of a two-chamber bidirectional pneumatic soft actuator using internal pressures of both chambers as inputs of the LSTM [25]. These works have proven that deep neural networks can continuously predict the gripper forces, when corresponding data sets such as internal pressures or actuator geometry are sufficiently collected. The reason why data-driven deep learning approaches can estimate the soft actuators' force is presumed to be that the geometry of the soft systems is determined according to external force and internal pressure [26], [27], yet existing studies have not directly utilized domain knowledge that may be extracted from mechanical modeling of soft robots. Furthermore, to the best of our knowledge, there are no studies of pneumatic-driven soft grippers' force estimation using visual feedback data and deep neural networks while the visual feedback is commonly used in gripper systems, yet there is a study that estimates the force exerted to the wrist of a robot-arm through the vision data of soft grippers [28].

In this letter, we investigate the impact of parameters extracted from domain knowledge of soft grippers and RGB image data on grip force estimation performance, and present a vision-based deep learning model *OriGripNet*. An intuitive approximation modeling technique, Pseudo-Rigid-Body-Model(PRBM), was inadequate to reflect the complexity of the soft gripper due to the nonlinear stiffness change during pneumatic actuation. Instead, we have extracted parameters related to the grip force based on the PRBM model, applied pressure and joint position information. We collected grip force and image data of dual-origami soft grippers for objects of different sizes and surface geometries on a variety of backgrounds. Then, we evaluated the grip force estimation and contact estimation performances of models that selectively consider parameters and images. The proposed model *OriGripNet*, which takes the image as an input to CNN and the pressure and joint position information as input to the FC layers, demonstrated grip force estimation with a x9 performance improvement over the model using only vision data, x2.63 improvement over models that don't use images directly but only key parameters, and x1.23 improvement over the model using both vision data and pressure data. Additionally, considering joint position information prevented more than half of the false positives for contact estimation. We expect that our results, which show that image, input for actuator, and current gripper geometry information are not only important for grip force and contact estimation but are also synergistic with each other, will provide key insights into the study of vision-based soft gripper grip force estimation.

II. Materials and Methods

A. Dual-origami gripper

The schematic diagram of the dual-origami gripper we developed in a previous study [9] is shown in Fig. 1A. The dual-origami gripper was 3D printed to follow the geometry of the Miura origami polyhedron, with pouch modules of flexible material stacked in a zigzag pattern to form an origami fluid network and an additional origami

strain limiting layer inserted between the pouches. When pneumatic/hydraulic pressure is applied, all crease lines of the origami gripper begin to gradually unfold and the body deploys mainly in the direction of lengthening. Then, once the origami strain-limiting layers are fully unfolded, the origami fluidic network, which is designed to be relatively more stretched, is unfolded alone and the entire body bends. The compact design of the dual-origami soft gripper provides high space utilization when not in use. However, as a trade-off for the space utilization of the folded shape, the origami soft gripper has a complex geometry that makes it difficult to embed appendages such as optical fiber sensors and soft/flexible circuits.

B. Estimation of the tip force of the dual-origami soft gripper via PRBM modeling

Single finger of the soft origami gripper can be represented in a schematic diagram as shown in Fig. 1B(i). When the gripper is deployed by applied fluidic pressure and the external force is applied to the end, most of the deformation occurs in the form of angle changes at the connections between the pouches(we call them 'joints'). Based on the deformation, we applied a pseudo-rigid-body-model in which the rotational deformation occurs only at the joints(Fig. 1A(ii)). For the sake of simplicity, the following assumption(1-3) were made. (1) For the same design, the origami gripper will always unfold to the corresponding geometry for a given pressure value in an environment where no external forces are applied. Namely, the rotation angle of the n-th joint in the absence of an external force(ϕ_n) would be a function of the pressure($\phi = f_\phi(P)$). (2) The object is only gripped at the gripper end, and the reaction force only causes rotation at the joint. (3) The effects of gravity are not considered. Letting the n-th joint be a nonlinear rotating spring with torsional stiffness k_n, the force expression for the segmented PRBM model is given as follow where θ_n is the rotation angle of the n-th joint by external force:

$$\vec{F} = \frac{(k_n\theta_{n+1} - k_n\theta_{n+1})(\vec{i} + \tan\sum_{i=1}^{N}(A_i)\vec{j})}{L_n\{\cos\sum_{i=1}^{N+1}(A_i) + \sin\sum_{i=1}^{N+1}(A_i)\tan\sum_{i=1}^{N}(A_i)\}}$$

where

$$A_i = (\theta_i - \phi_i)$$

The above expression shows that if ϕ and k are known experimentally in advance for the gripper with a given design geometry, the grip force can be estimated by detecting the value of θ.

$$F = f_F(k_{1:N}, \phi_{1:N}, \theta_{1:N})$$

However, we have concluded from simulation and experimental results that k is nonlinear with respect to P and θ due to the complex geometry of the gripper, nonlinear materials, and contact effects between the pouches($k = f_k(P, \theta)$).

C. OriGripNet

As we discussed in the previous section, gripping force estimation through modeling of origami grippers with nonlinear materials and complex geometries is challenging. However, the relationship between the variables shows that the gripping force is determined by the applied pressure *P* and the angular rotation of the joint $\theta_{1:N}$ due to the reaction force during gripping. Since the angle values of the joints can be derived from the position of the joints, the gripping force has the following relationship:

$$F = f_F(P, \hat{J}_{1:N})$$

Based on this domain knowledge, we developed a gripping force estimating deep learning network *OriGripNet* as shown in Fig. 1C. *OriGripNet* is composed of two ResNet-based convolutional neural networks [29]. The first ResNet layer(Force estimation Layer) is based on ResNet18 and extracts vision feature H_0^V, followed by fully connected layers to estimate force. The second ResNet layer(Joint position estimation layer) is based on ResNet50 and aims to accurately predict joint positions $\hat{J}_{0:N}$. The estimated joint value $\hat{J}_{0:N}$ and applied pressure P measured by the pressure sensor are inserted into the FC layers of the Force estimation layer, which is expected to improve performance compared to ResNet with only vision data.

D. Real-world data collection and training

Data collection through FEA simulation and transfer to the real world may be applicable for conventional parallel grippers with relatively simple geometries and deformations. However, for soft origami grippers made of hyperplastic materials, simulation errors are large and FEA convergence is difficult. We set up a data collection platform and collected real world data. As shown in Fig. 2A, the gripper was mounted on the end of the robot arm and an RGB camera(1080P Low Light Wide Angle USB Camera, ArduCam) was installed on the wrist of the robot arm(RB5-850, Rainbow Robotics) to record the entire process of the gripper deployment and bending. In order for the learning model to effectively track each joint, we marked each joint with different colors(Fig. 2A). Each joint location was labeled using the DeepLabCut toolkit [30]. We selected 100 images from the RGB image through k-means clustering [30], and each joint location of the 100 images was manually labeled using the GUI provided by the DeepLabCut toolkit.

Six length modules(1, 10, 15, 20, 30, and 40 mm) and two shape modules(Flat, Round) were 3D printed and assembled with a singe-axis load cell(333FDX, KOYTO) to create a test object. The load cell was placed inside the object so as not to affect the grip. For the diversity of image data to prevent overfitting, three different backgrounds(white, green, and black) were utilized with various objects (Fig. 2B and 2C). The force, image data, and input pressure(was controlled using pressure regulator(ITV1050, SMC) and analog voltage output module(National Instruments)) were simultaneously collected for two different gripper designs(3-joints and 2-joints gripper) using LABVIEW(National Instruments). In

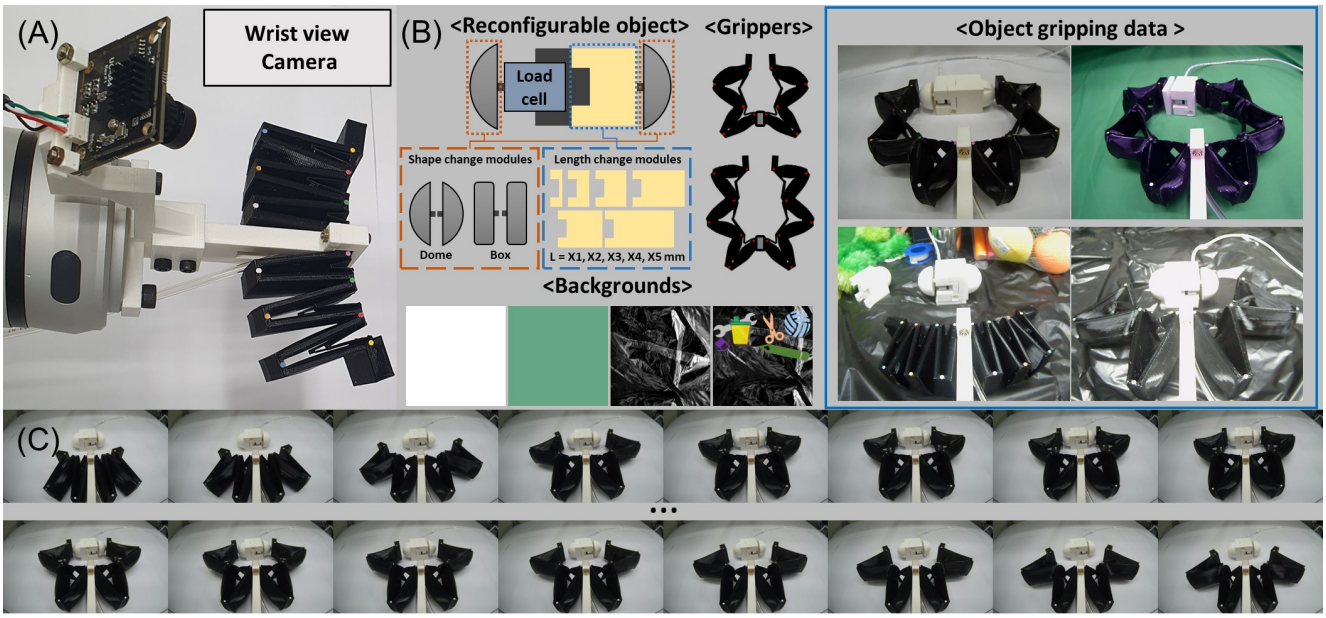

Fig. 2. Collection of visual and force data during object gripping. A) The 6-module dual-origami soft gripper and RGB camera mounted at the end of the robot arm. B) For data diversity, 3D printed reconfigurable objects with variable shape and size, four different backgrounds, and two grippers with different designs were used. (left). Real-world image data. (right). C) Example of continuously collected object grap-and-release image data.

TABLE I
GRIP FORCE ESTIMATION PERFORMANCE FOR EACH GRIPPER

Model	3-joints-gripper(MSE)			2-joints-gripper(MSE)			Comp. Time(S)
	Round	Flat	Round+Flat	Round	Flat	Round+Flat	
Img	0.0084	0.0078	0.0079	0.0079	0.0042	0.0036	0.0061
Img+J	0.0104	0.0106	0.0122	0.0073	0.0013	0.0013	0.0301
Img+P	0.0011	0.0011	0.0009	0.0016	0.0008	0.0011	0.0062
Img+P+J	**0.0008**	**0.0007**	**0.0005**	0.0029	**0.0004**	0.0007	0.0261
P+J	0.0018	0.0007	0.0023	**0.0011**	0.0009	**0.0003**	0.0218

J indicates joint information
P indicates pressure information
Comp. Time indicates Computational Time

addition, gripping and releasing data were collected by varying the pressure profile applied to the gripper. The applied pressure profiles include holding the object for a long time with a constant pressure, slowly changing the pressure continuously or step-wise, and quickly gripping and releasing the object with a rapid applied pressure increase followed by a rapid decrease(Fig. 3). In addition, the data of lifting an object by moving the robot arm in the opposite direction of gravity and then moving it back and forth and side to side were also included.

The network models proposed in this study undergoes two training processes. First, the Joint layer was trained to accurately predict the positions of each joint. The Joint layer was trained for 100,000 iterations with a learning rate of 0.001 and weight decay of 0.01 using the Adam Optimizer. Subsequently, the pretrained Joint layer was used to train the Force estimation model. The Force estimation model was also trained with a learning rate of 0.001 and weight decay of 0.01 using the Adam Optimizer and was trained for 100 epochs. Model training was performed using three of the same GPUs(Titan V, NVIDIA, USA). Both joint estimation and force estimation used Mean Square Error(MSE) loss functions.

III. RESULT

For convenience of description, we refer to *Img* as the input of the image set to the force estimation layer, *P* as the input of the pressure values to the FC layers, and *J* as the input of the joint positions estimated from the joint position estimation layer to the FC layers. For example, *OrigripNet* can be denoted as *Img+P+J*. When only pressure and joint data were directly used and image data was used only for joint position estimation, it is denoted as *P+J*.

A. Grip force estimation performance

Table I shows the performance results of grip force estimation networks that selectively used information from the parameters. The grip force estimation networks were trained with data sets of only Flat objects, only Round objects,

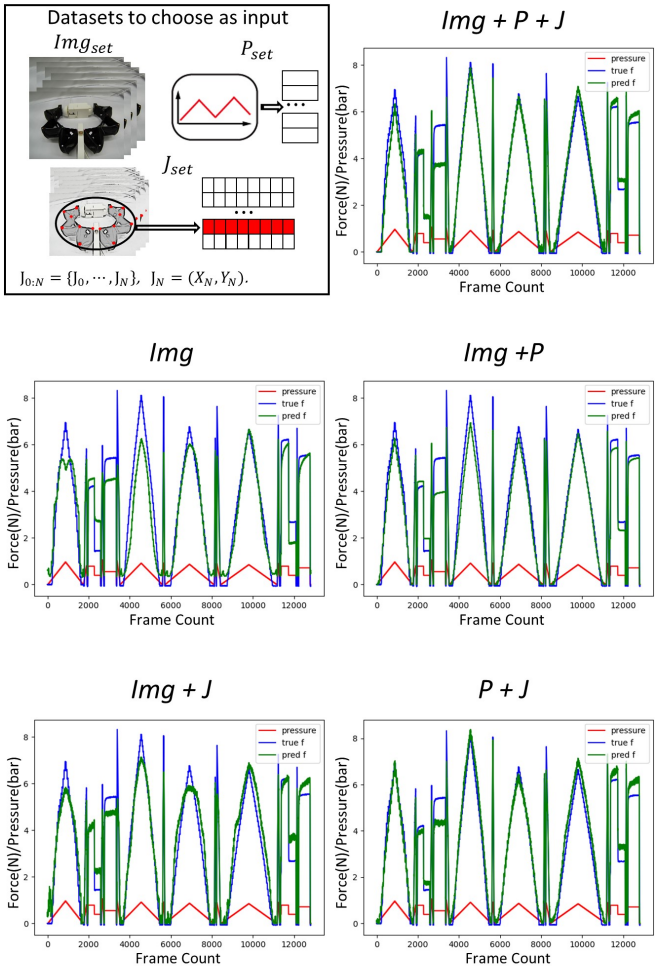

Fig. 3. Visualization of the actual applied force(true f) and the predicted force of each deep learning model(pred f) when gripper grip the test objects with varying force and duration.

or both(Flat+Round), and tested on all types of datasets for the corresponding gripper design(2-joints and 3-joints). Additionally, Fig.3 shows the results of the force estimation networks on the same dataset with various force/pressure profiles. Both the numerical and visualization results show that image information alone is an inaccurate predictor of grip force, but that the addition of pressure information is significantly helpful in predicting grip force. For example, adding pressure information increased performance by more than x8 for a 3-joint gripper(from 0.0079 to 0.0009) and x3 for a 2-joint gripper(from 0.0036 to 0.0011), each trained on Round+Flat dataset. On the other hand, the effect of adding joint information alone(*Img+J*) was inconclusive, as it degraded the performance of the 3-joint gripper and improved the performance of the 2-joint gripper. Interestingly, however, when joint information was added to *Img+P*, i.e. *OriGripNet*, the performance generally increased(five of six cases). The force estimation performance of *OriGripNet* for the 3-joints-gripper and 2-joints-gripper trained on the Round+Flat dataset was 0.0005 and 0.0007, respectively,

TABLE II

GRIP FORCE ESTIMATION PERFORMANCE WITH SELECTIVE FALSE DATA

Model	Rand. Info.	MSE(N^2)
Img	*None*	0.0042
Img+P	*P*	0.0055
Img+P+J	*P+J*	0.0073
Img+P+J	*P*	0.0032
Img+P+J	*J*	0.0018
Img+P+J	*None*	**0.0004**

J indicates joint information
P indicates pressure information
Rand. Info. indicates random information added to the model

which is 1.8x and 1.57x better than the performance of *Img+P*(0.0009 and 0.0011, respectively). It was also noteworthy that networks that do not use image data directly, but only utilize key parameter information(*P+J*), performed as well as or better than *OriGripNet* in few cases. This result supports our claim that grip force can be estimated from only the important parameters($F = f_F(P, \hat{J}_{1:N})$), and implies that other kinds of sensors besides camera providing the joint position information of the gripper(e.g., embedded bending sensors) can be also utilized to estimate the grip force. However, we also suspect that the similarity of the test and training subjects may have contributed to the high performance of *P+J*, as the performances of the network are relatively weak for the contact and release detection and unseen objects which are presented in the following sections(III-B and C).

For network computation time, *OriGripNet* on average took 0.0261 seconds which is about 38.3 calculations per minute. This result shows that when sensing with a real-time camera at 30 fps, *OriGripNet* can also process the data in real-time to estimate the grip force. In comparison, *Img* and *Img+P* were more than x4 faster than *OriGripNet*(0.0061 second an 0.0062 second, respectively). This is because these networks do not utilize the joint position information, while the separate process of extracting joint information requires a relatively long computation time.

To further validate the result that gripper's state information *P* and *J* increase the grip force estimation performance, we trained the networks with false datasets of *P* or *J* that are randomly generated. As shown in Table II, the performance of *Img* becomes progressively worse as the incorrect gripper state information *P* and *J* are added(RMSE performance of 0.0042 worsens to 0.0055 and 0.0073, respectively). Also interestingly, *OriGripNet* improves performance over *Img* if at least one of the two gripper state information is genuine, even if the other is false(0.0018 and 0.0032 for genuine *P* or *J*, respectively), yet their performances are significantly worse compared to the reference *OriGripNet* with RMSE of 0.0004. We believe that this result also supports the idea that both *P* and *J* information are important for estimating grip force, and it can also concluded from both Tables I and II that *P* is particularly important.

TABLE III
CONTACT DETECTION PERFORMANCE

Model	Window	Recall(%)		
		Gripping	Releasing	Avg
Img	200	87.6	77.6	82.6
Img+J	200	89.2	78.6	83.9
Img+P	200	90.3	66.3	78.3
Img+P+J	200	**94.7**	**86.2**	**90.5**
P+J	200	82.2	74.6	78.4

J indicates joint information
P indicates pressure information

TABLE IV
GRIP FORCE ESTIMATION PERFORMANCE FOR UNSEEN OBJECTS

Model	MSE(N^2)	MAE(N)	MAPE(%)
Img	0.0808	0.2315	16.32
Img+J	0.0778	0.2184	11.34
Img+P	0.0101	0.0760	7.08
Img+P+J	**0.0082**	**0.0636**	**4.31**
P+J	0.0220	0.1189	6.21

J indicates joint information
P indicates pressure information

B. Contact estimation performance

Determining whether the gripper has contacted an object is a necessary information for assessing the feasibility of gripping and releasing actions. From the force estimation results, we found that force estimation through learning is subject to false positives of contact, which are not simply represented by performance numbers. To evaluate this, grip force estimation models proposed in this paper were tested to confirm whether they could accurately detect contact with the object. We have manually labeled the gripping and release point, and only data around the actual contact(200 frames around the contact) was used as test data to obtain recall(true positive rate for all actual contact, $TP/(TP + FN)$) near where the actual contact occurs. Each model was trained on data of gripping flat objects against a black background. The result for the recall of each learning model in determining contact is shown in Table III. As a result, *OriGripNet* demonstrated the best recall among the proposed networks, 94.7% for gripping and 86.2% for releasing with an average recall of 90.5%. On the other hand, the *P+J* model averaged 78.4% contact detection performance, which is about 12% lower than *OriGripNet*, showing that image information obviously plays an important role for the contact decision. The average recall of the *Img* model was 82.6%, and there was a slight performance increase when adding joint position information(83.9% for *Img+J*). We believe this is because the detailed gripper geometry information derived from *J* is intuitively important in determining whether contact is made. Also, the *Img+P* model showed more than 12% performance degradation in contact information during releasing compared to the *Img* model. However, when comparing the *Img+J* model to *OriGripNet*, we see that *OriGripNet* has better contact information recall performance during both gripping and releasing, suggesting that pressure information helps improve the contact detection performance when combined with detailed geometry information(*J*) of the gripper.

C. Grip force estimation for unseen objects

To investigate the generalizability of our learning models, we evaluated the performance of our proposed model with unseen objects of different shapes commonly encountered in everyday life. Three test objects - a cylinder, a sphere, and a cup - were 3D printed in half-form to allow for the load cell to be built inside and assembled as shown in Fig 4A. These unseen objects gripping data were also collected for various applied pressure profiles in different background environments as described in session II-D(Fig. 4B). The performance of pretrained models for the unseen objects is shown in Table IV(For MAPE, a small value ϵ was introduced for cases where the reference ground truth is zero) and Fig. 4C.

OriGripNet clearly performed the best for force estimation on unseen objects, with MSE of 0.0082, mean average error(MAE) of 0.0636, and mean absolute percentage error(MAPE) of 4.31%. For the models do not use pressure data *P*(*Img* and *Img+J*), the MSE values were more than x9 higher than OriGripNet, which are unacceptably poor performances for use as shown in Fig. 4. The performance of *Img+P* was 23% lower than *OriGripNet* based on the MSE value, again showing a significant performance improvement when combining *J* data with *P*. We also found that the *Img+P* model tended to estimate false positives for unseen objects even before gripping occurred, which is well shown in the cases where the gripping force increases continuously/step-wisely(Fig. 4C). We suspect this is because the *Img+P* model relies too heavily on the pressure input value, while it does not extract detailed geometry of the gripper from the image alone(especially with background data that is the same color as the gripper, black). In other words, the *Img+P* model seems to be estimating the false positive values just because the input pressure value is increasing. The false positive problem is mitigated in models with both *P* and *J* data, *OrigripNet* and *P+J*, indirectly supporting our speculation. Finally, in contrast to the good performance of the *P+J* model on known objects in Table I, the performance on unseen objects was half of OriGripNet(more than x2 based on MSE value). We believe that this is an intuitive result that demonstrates the effectiveness of *Img* data in "seeing" unseen objects.

IV. CONCLUSIONS

In this letter, we explored the impact of parameters related to the gripping force of a deployable origami soft gripper on the force estimation of a vision-based deep learning model. We built a data collection setup and collected real-world data of gripping 3D printed reconfigurable objects with various backgrounds and applied pressure profile. As a result,

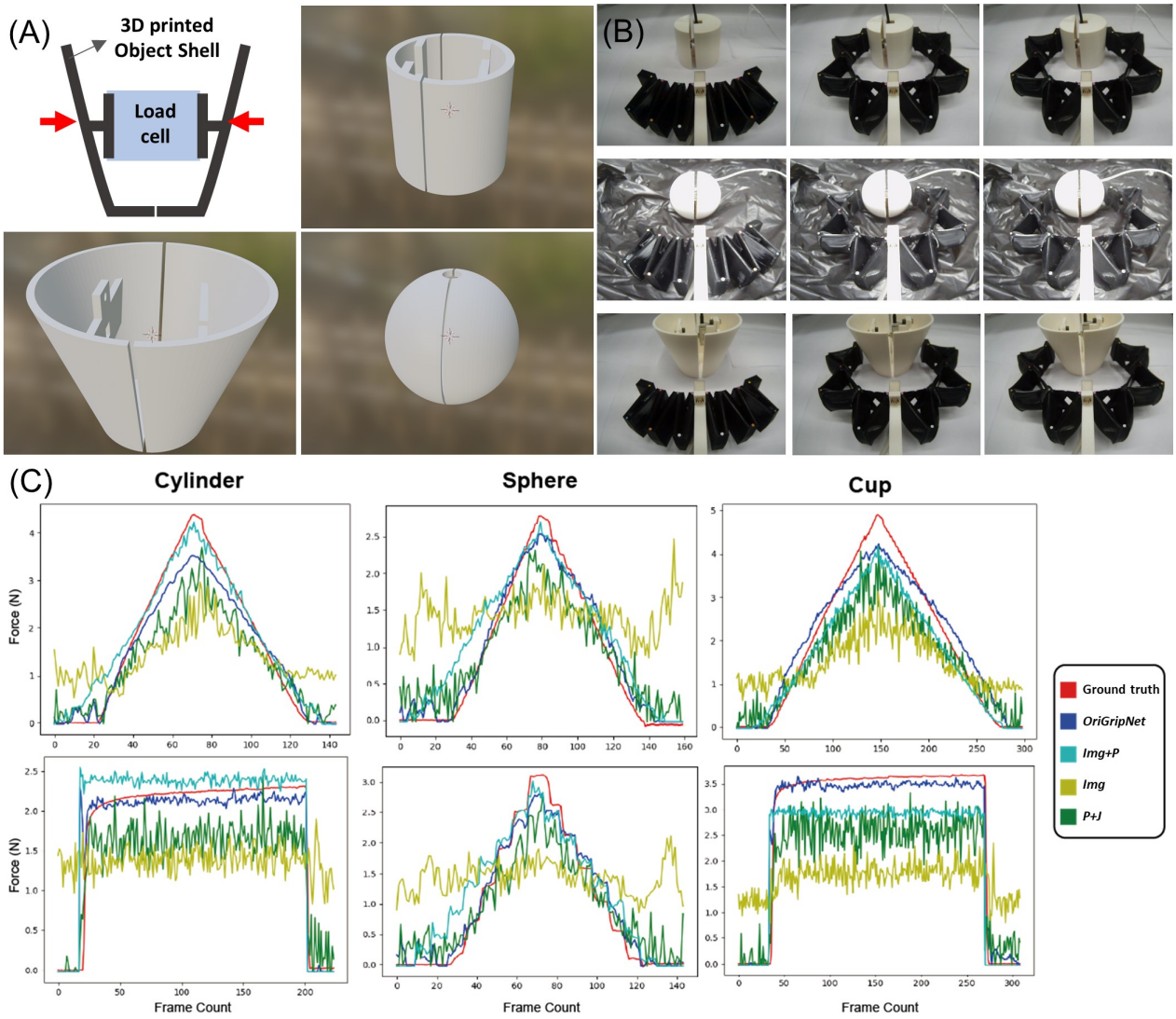

Fig. 4. Test results for unseen objects. A) Design of the unseen objects with a built-in load cell. B) Image data of the soft gripper gripping the unseen objects. C) Grip force estimation results for unseen objects.

we found that the model which utilizes only image and actuation level(applied pressure) data showed 'acceptable' performance, but there was a drawback that it seemed to rely too much on the pressure value rather than the image, resulting in false positives even when the gripper does not make contact with the object. On the other hand, when the model considers the joint positions extracted from the markers in the image, the force estimation performance increased by at least 25%, and the incorrect contact detection was more than halved, from 21.7% to 9.5%. The results showed that the key parameters, P and J, and image data not only have their own distinct roles in force estimation, but also have a synergistic effect in improving performance when considered together. Finally, the computational performance of approximately 38 fps and the force estimation performance of MSE 0.0082 showed the usability of the proposed model, *OriGripNet*. We believe our approach provides fruitful insights into the utilization of domain knowledge for soft machines and the practical application of deep learning models using real-world data.

On the other hand, because our work focused on identifying the impact of key information parameters rather than increasing the absolute performance or practicality, there is much room for improvement. For example, our study only estimated the force of a single axis in the direction of the object being pressed, which is the most dominant. However, for more delicate object handling, a 6-axis force torque sensor can be utilized instead of a 1-axis load cell to obtain the data, and then similar method presented in our work can be applied. Moreover, we conducted training and verification only on objects that are rigid and not easily deformed, and the future work should be conducted on soft objects since soft grippers mainly handle them. At this point, the force estimation model that not only considers the physical property of the gripper but also the physical property of object can be designed. For example, our previous work that

estimates the contact force between tendon-driven wearable robot and deformable objects utilized the estimated stiffness of objects from actuation data [31]. Similarly, a stiffness estimation model that utilizes vision and gripper data could be developed and applied to *OriGripNet* to improve performance, especially for soft objects.

REFERENCES

[1] S. Levine, P. Pastor, A. Krizhevsky, J. Ibarz, and D. Quillen, "Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection," *The International Journal of Robotics Research*, vol. 37, pp. 421–436, 4 2018.

[2] H. Zhang, J. Peeters, E. Demeester, and K. Kellens, "Deep learning reactive robotic grasping with a versatile vacuum gripper," *IEEE Transactions on Robotics*, vol. 39, pp. 1244–1259, 4 2023.

[3] I. Andrade-Silva and J. Marthelot, "Fabric-based star soft robotic gripper," *Advanced Intelligent Systems*, 4 2023.

[4] B. Chen, Z. Shao, Z. Xie, J. Liu, F. Pan, L. He, L. Zhang, Y. Zhang, X. Ling, F. Peng, W. Yun, and L. Wen, "Soft origami gripper with variable effective length," *Advanced Intelligent Systems*, vol. 3, p. 2000251, 10 2021.

[5] A. Gunderman, J. Collins, A. Myers, R. Threlfall, and Y. Chen, "Tendon-driven soft robotic gripper for blackberry harvesting," *IEEE Robotics and Automation Letters*, vol. 7, pp. 2652–2659, 4 2022.

[6] N. R. Sinatra, C. B. Teeple, D. M. Vogt, K. K. Parker, D. F. Gruber, and R. J. Wood, "Ultragentle manipulation of delicate structures using a soft robotic gripper," *Science Robotics*, vol. 4, 8 2019.

[7] Z. Wang, K. Or, and S. Hirai, "A dual-mode soft gripper for food packaging," *Robotics and Autonomous Systems*, vol. 125, 3 2020.

[8] Z. Wang, R. Kanegae, and S. Hirai, "Circular shell gripper for handling food products," *Soft Robotics*, 2020.

[9] W. Kim, J. Eom, and K.-J. Cho, "A dual-origami design that enables the quasisequential deployment and bending motion of soft robots and grippers," *Advanced Intelligent Systems*, vol. 4, p. 2100176, 3 2022.

[10] D. Sui, Y. Zhu, S. Zhao, T. Wang, S. K. Agrawal, H. Zhang, and J. Zhao, "A bioinspired soft swallowing gripper for universal adaptable grasping," *Soft Robotics*, vol. 9, pp. 36–56, 2 2022.

[11] C. Tawk, R. Mutlu, and G. Alici, "A 3d printed modular soft gripper integrated with metamaterials for conformal grasping," *Frontiers in Robotics and AI*, vol. 8, 1 2022.

[12] W. Kim, J. Byun, J. K. Kim, W. Y. Choi, K. Jakobsen, J. Jakobsen, D. Y. Lee, and K. J. Cho, "Bioinspired dual-morphing stretchable origami," *Science Robotics*, vol. 4, p. eaay3493, 2019.

[13] P. Polygerinos, N. Correll, S. A. Morin, B. Mosadegh, C. D. Onal, K. Petersen, M. Cianchetti, M. T. Tolley, and R. F. Shepherd, "Soft robotics: Review of fluid-driven intrinsically soft devices; manufacturing, sensing, control, and applications in human-robot interaction," *Advanced Engineering Materials*, vol. 19, p. 1700016, 12 2017.

[14] D. M. Vogt, K. P. Becker, B. T. Phillips, M. A. Graule, R. D. Rotjan, T. M. Shank, E. E. Cordes, R. J. Wood, and D. F. Gruber, "Shipboard design and fabrication of custom 3d-printed soft robotic manipulators for the investigation of delicate deep-sea organisms," *PLOS ONE*, vol. 13, p. e0200386, 8 2018.

[15] K. C. Galloway, K. P. Becker, B. Phillips, J. Kirby, S. Licht, D. Tchernov, R. J. Wood, and D. F. Gruber, "Soft robotic grippers for biological sampling on deep reefs," *Soft Robotics*, vol. 3, pp. 23–33, 3 2016.

[16] J. Wang and A. Chortos, "Control strategies for soft robot systems," *Advanced Intelligent Systems*, vol. 4, 5 2022.

[17] M. S. Xavier, C. D. Tawk, A. Zolfagharian, J. Pinskier, D. Howard, T. Young, J. Lai, S. M. Harrison, Y. K. Yong, M. Bodaghi, and A. J. Fleming, "Soft pneumatic actuators: A review of design, fabrication, modeling, sensing, control and applications," *IEEE Access*, vol. 10, pp. 59442–59485, 2022.

[18] H. Zhao, K. O'Brien, S. Li, and R. F. Shepherd, "Optoelectronically innervated soft prosthetic hand via stretchable optical waveguides," *Science Robotics*, vol. 1, 12 2016.

[19] X. Liu, F. Zhao, S. S. Ge, Y. Wu, and X. Mei, "End-effector force estimation for flexible-joint robots with global friction approximation using neural networks," *IEEE Transactions on Industrial Informatics*, vol. 15, pp. 1730–1741, 3 2019.

[20] D. D. Barrie, M. Pandya, H. Pandya, M. Hanheide, and K. Elgeneidy, "A deep learning method for vision based force prediction of a soft fin ray gripper using simulation data," *Frontiers in Robotics and AI*, vol. 8, 5 2021.

[21] O. Faris, R. Muthusamy, F. Renda, I. Hussain, D. Gan, L. Seneviratne, and Y. Zweiri, "Proprioception and exteroception of a soft robotic finger using neuromorphic vision-based sensing," *Soft Robotics*, vol. 10, pp. 467–481, 6 2023.

[22] D.-K. Ko, K.-W. Lee, D. H. Lee, and S.-C. Lim, "Vision-based interaction force estimation for robot grip motion without tactile/force sensor," *Expert Systems with Applications*, vol. 211, p. 118441, 1 2023.

[23] T. G. Thuruthel, B. Shih, C. Laschi, and M. T. Tolley, "Soft robot perception using embedded soft sensors and recurrent neural networks," *Science Robotics*, vol. 4, 1 2019.

[24] J. Y. Loo, Z. Y. Ding, V. M. Baskaran, S. G. Nurzaman, and C. P. Tan, "Robust multimodal indirect sensing for soft robots via neural network-aided filter-based estimation," *Soft Robotics*, vol. 9, pp. 591–612, 6 2022.

[25] B. W. K. Ang and C.-H. Yeow, "A learning-based approach to sensorize soft robots," *Soft Robotics*, vol. 9, pp. 1144–1153, 12 2022.

[26] C. D. Santina, R. L. Truby, and D. Rus, "Data–driven disturbance observers for estimating external forces on soft robots," *IEEE Robotics and Automation Letters*, vol. 5, pp. 5717–5724, 10 2020.

[27] Z. Wang and S. Hirai, "Soft gripper dynamics using a line-segment model with an optimization-based parameter identification method," *IEEE Robotics and Automation Letters*, vol. 2, pp. 624–631, 4 2017.

[28] J. A. Collins, P. Grady, and C. C. Kemp, "Force/torque sensing for soft grippers using an external camera," pp. 2620–2626, IEEE, 5 2023.

[29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," pp. 770–778, IEEE, 6 2016.

[30] A. Mathis, P. Mamidanna, K. M. Cury, T. Abe, V. N. Murthy, M. W. Mathis, and M. Bethge, "Deeplabcut: markerless pose estimation of user-defined body parts with deep learning," *Nature Neuroscience*, vol. 21, pp. 1281–1289, 9 2018.

[31] E. Rho, D. Kim, H. Lee, and S. Jo, "Learning fingertip force to grasp deformable objects for soft wearable robotic glove with tsm," *IEEE Robotics and Automation Letters*, vol. 6, pp. 8126–8133, 10 2021.