

# Visual Pressure Estimation and Control for Soft Robotic Grippers

Patrick Grady<sup>1</sup>, Jeremy A. Collins<sup>1</sup>, Samarth Brahmbhatt<sup>2</sup>, Christopher D. Twigg<sup>3</sup>,  
Chengcheng Tang<sup>3</sup>, James Hays<sup>1</sup>, Charles C. Kemp<sup>1</sup>

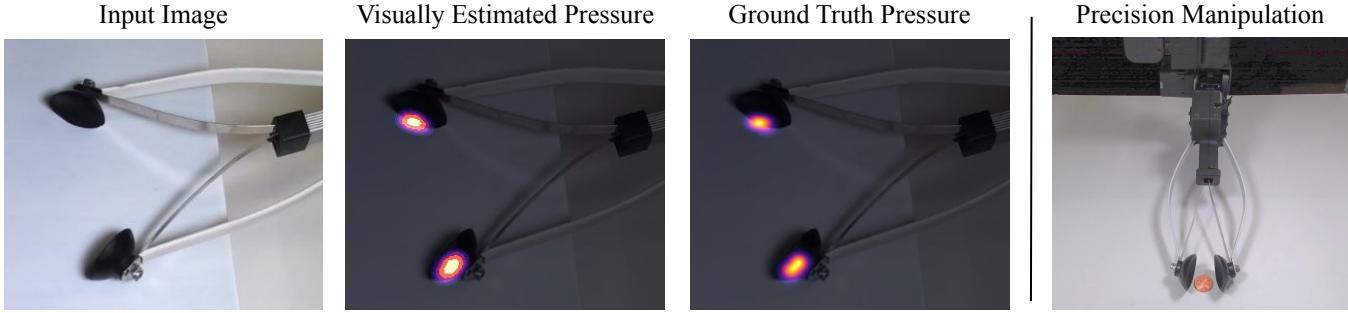


Fig. 1. **Left:** Given a single RGB input image, VPEC-Net estimates the pressure applied by a soft gripper to a flat surface. **Middle Left and Middle Right:** Pressure images are shown overlaid on the input image. VPEC-Net outputs a pressure image with a pressure estimate for each pixel of the input image. The estimated pressure image closely matches the ground truth pressure image obtained with a planar pressure sensing array. **Right:** Using visual servoing with estimated pressure images, a robot grasps small objects, including a penny.

**Abstract**—Soft robotic grippers facilitate contact-rich manipulation, including robust grasping of varied objects. Yet the beneficial compliance of a soft gripper also results in significant deformation that can make precision manipulation challenging. We present visual pressure estimation & control (VPEC), a method that infers pressure applied by a soft gripper using an RGB image from an external camera. We provide results for visual pressure inference when a pneumatic gripper and a tendon-actuated gripper make contact with a flat surface. We also show that VPEC enables precision manipulation via closed-loop control of inferred pressure images. In our evaluation, a mobile manipulator (Stretch RE1 from Hello Robot) uses visual servoing to make contact at a desired pressure; follow a spatial pressure trajectory; and grasp small low-profile objects, including a microSD card, a penny, and a pill. Overall, our results show that visual estimates of applied pressure can enable a soft gripper to perform precision manipulation.

## I. INTRODUCTION

High compliance helps soft robotic grippers conform to the environment and apply low forces during contact, but it also results in significant deformation that makes precise motions more difficult to achieve. For precision manipulation tasks, small position errors due to deformation can lead to failure. For example, tasks using fingertips to press a small button, flip a small switch, or pick up a small object depend on pressure being applied with precision.

<sup>1</sup>Patrick Grady, Jeremy A. Collins, James Hays, and Charles C. Kemp are with the Institute for Robotics and Intelligent Machines at the Georgia Institute of Technology (GT). <sup>2</sup>Samarth Brahmbhatt is with Intel Labs.

<sup>3</sup>Christopher D. Twigg and Chengcheng Tang are with Meta Reality Labs. This work was supported in part by NSF Award # 2024444. Code, data, and models are available at <https://github.com/Healthcare-Robotics/VPEC>. Charles C. Kemp is an associate professor at GT. He also owns equity in and works part-time for Hello Robot Inc., which sells the Stretch RE1. He receives royalties from GT for sales of the Stretch RE1.

One approach to precisely apply pressure is to explicitly model the mechanics of the gripper. Rigid-body models enable precise control of rigid grippers, but result in errors when applied to soft grippers. For example, sliding a soft gripper’s fingertips across a surface can result in large deviations from the gripper’s undeformed geometry. Soft-body models can represent the compliant geometry of soft grippers, but are more complex than rigid-body models and depend on quantities that can be difficult to measure, such as external forces and internal strain. Embedded sensors in soft grippers can make relevant measurements, but increase hardware complexity and often alter gripper mechanics. For both rigid and soft grippers, an explicit model of the gripper needs to be related to the environment to model applied pressure, which often involves additional sensors and calibration.

We present a novel approach that circumvents these modeling and instrumentation complexities by using an external camera to directly estimate the pressure applied by a soft gripper to the world. Our approach relies on two key insights. First, many manipulation tasks only depend on the pressure applied by the gripper to the world, rather than the gripper’s detailed state. For these tasks, directly estimating and controlling pressure applied to the world is sufficient for task success. Second, the pressure applied to the world by a soft gripper can be directly estimated by the gripper’s visible deformation. This takes advantage of high compliance, since larger deformations are more easily observed by an external camera.

Our method, visual pressure estimation & control (VPEC), uses a convolutional neural network, VPEC-Net, to infer a 2D *pressure map* overlaid on the input RGB image from an external camera (see Figure 1). In other words, contact

locations and pressure are estimated *in the image space* with an estimated pressure for each pixel in the image. A control loop achieves pressure objectives in the image space, enabling a robot to precisely control pressure applied to the world and thereby grasp a small object observed by the camera. In addition to gripper deformation, VPEC-Net has the potential to use other information, such as cast shadows and motion blur.

For this paper, we consider contact with a horizontal plane, which is a common surface relevant to manipulation. To construct a training dataset, we hand-operated a tendon-driven soft gripper and a pneumatic soft gripper to make contact with a high-resolution planar pressure sensor. We capture these interactions with four RGB cameras and use camera extrinsics to project the pressure sensor data onto the RGB images, creating a labeled dataset for training VPEC-Net. We collected approximately one hour of data for each gripper, yielding a dataset of 650K frames. Our contributions include the following:

- **VPEC:** An algorithm that infers pressure applied by a soft gripper to a planar surface using a single RGB image.
- **Precision Manipulation with VPEC:** Evaluations in which a mobile manipulator with a soft gripper achieves pressure objectives via closed-loop control and grasps small objects, including a washer and a coin.
- **Release of dataset, trained models, and code:** We will release our core methods online to support replication of our work.

## II. RELATED WORK

Our work builds on prior efforts to visually infer pressure applied by human hands [1]. We use the same neural network architecture, but apply it to inference and control of soft robotic grippers.

We evaluate our approach with the task of precision manipulation on a flat surface. Our method controls the pressure applied by a soft gripper to the surface to grasp small objects. Similarly, humans often slide their fingertips across flat surfaces when picking up small objects [2], [3], which has inspired robotic grasping methods [4], [5], [6]. Work on grasping with soft end effectors has focused on larger objects than we consider [7], [8], [9]. Grasping smaller objects tends to be more sensitive to gripper deformation, such as deformation due to sliding while in contact.

Prior work has used internal sensors to infer contact and pressure based on deformation of the gripper’s surface [10], [11], [12], [13], [14], changes to gripper vibration [15], deflection of the gripper’s compliant joints [16], and changes to the gripper’s motion [17]. We expect that our method can use similar information by observing a soft gripper with an external camera. For this paper, estimation only uses a single image, which can have motion blur but lacks information available in sequences of images.

Research on image-based force estimation has focused on inferring force and torque applied by a rigid tool to a deformable object [18], [19], [20], [21], [22], [23]. Early

work inferred grip force for a microgripper [24]. Cross-modal research has used vision to predict the output of robot-mounted tactile sensors [25], [26], [27]. Our approach relies on soft grippers that visibly deform to infer the output of tactile sensors mounted to the world.

Our method uses visual servoing [28] to achieve pressure objectives in the camera’s image space. Prior work has demonstrated visual control of a robot’s arm relative to flat surfaces based on shadows [29]. Marker-based visual servoing has achieved precise in-hand manipulation with a soft gripper [30]. Visual object tracking has enabled precision insertions with a soft gripper [31]. Our system enables a soft gripper to grasp small, low-profile objects from a flat surface.

Our system can be thought of as using a *virtual* tactile sensor array mounted to the world with which it attempts to achieve pressure objectives. As such, our approach is similar to tactile servoing with real tactile sensor arrays [32], [33], [34], [35], [36]. Notably, our system reports inferred pressure for each pixel of the input RGB image. This enables our system to directly relate pressure and vision. For example, in our grasping evaluation, the robot uses the RGB image to find the centroid of the target object, which determines key pressure objectives.

## III. VISUAL PRESSURE ESTIMATION

In this section, we describe the grippers used and the capture process to create our dataset. We also describe the network architecture and training procedure of VPEC-Net.

### A. Selected Grippers

For a vision-based approach to successfully estimate contact and pressure, there must be visual cues to indicate the presence of these quantities. As a result, we train and test VPEC-Net with soft robotic grippers. These grippers are compliant and deform when in contact with an object. We select two different models of grippers that are examples from common classes of grippers used by researchers.

**Tendon-Actuated Gripper:** The first gripper we consider is the *Stretch Compliant Gripper* that comes with the *Stretch RE1* mobile manipulator by Hello Robot Inc. This gripper has suction-cup-shaped soft rubber fingertips supported by spring steel flexures that bend when the gripper closes. To close the gripper, an actuator uses a tendon to pull the inner flexures. In a user study, a similar commercially available grabber tool was found to be adept at manipulating various household objects [37]. During the collection of the dataset (Sec III-B), we used a hand-operated version of this gripper.

The gripper displays several visual cues that may indicate pressure when grasping. The rubber fingertips visibly deform to match the contours of the grasped object or when in contact with a surface, and the steel flexures bend when in contact with a surface (Figure 2).

**Pneumatic Gripper:** The second gripper we consider is a pneumatic gripper sold by *SoftGripping GmbH* [38]. The gripper is made of a flexible silicone, and contains hollow cavities for pressurized air. When inflated, the sides of the finger expand asymmetrically, resulting in the fingers closing.

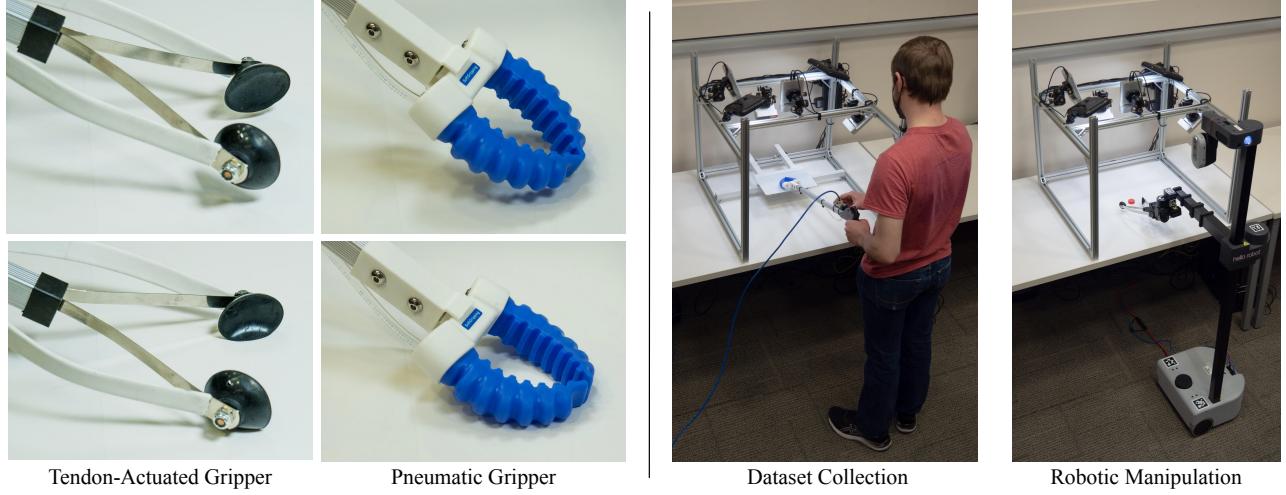


Fig. 2. **Left:** The two grippers used to train and test VPEC-Net. In the top images, the grippers are hovered in mid-air, and in the bottom images, the grippers are pressed against the surface. Notice the deflection in the pneumatic gripper and the deformation in the tips of the tendon-actuated gripper. **Right:** We use hand-operated grippers to make contact with a high-resolution pressure sensing array to collect training data. During robotic manipulation experiments, we command a Hello Robot Stretch RE1 robot to pick up a variety of small objects from the table.

Due to its silicone construction, the gripper is soft, and the entire finger deforms when in contact (Figure 2). Additionally, as the pressure in the finger cavity increases, the exterior of the gripper expands.

#### B. Data Capture Setup

We built a custom data capture rig to collect RGB images with synchronized ground truth pressure. The rig uses aluminum framing to rigidly support a pressure sensor and cameras. The parts of the rig visible to cameras are covered with a white vinyl covering to provide a consistent visual background.

To record pressure data, we use a Sensel Morph [39] sensor. The Morph is a flat pressure sensor with an active area of  $23 \times 13$  cm and features a grid of  $185 \times 105$  individual force-sensitive resistor (FSR) elements. The sensor produces high-resolution pressure data at approximately 100 Hz.

Four Azure Kinect cameras are mounted at different locations around the cage to observe the pressure sensor and gripper from a variety of viewpoints. The RGB feed from the cameras is captured in 1080p at 30 Hz. The capture rig additionally has two lights mounted to provide illumination which can be turned on or off in any combination. Bright lighting reduces the effect of motion blur, but during fast motions, some blurring is still visible.

The cameras and pressure sensor are calibrated before each recording session using a specialized fiducial board. The board uses ChArUco [40] markers on the top for localization in camera space, while pins on the bottom push into the edges of the pressure sensor, allowing consistent positioning.

#### C. Data Capture Protocol

While our work is targeted toward robotic grippers, we operate the grippers by hand for data collection (Figure 2). Collecting data with a human operator allows for efficient

capture of a diverse dataset including a wide range of pressure levels, orientations, grasp styles, and speeds. The grippers are mounted on a handle 60 cm in length to allow a person to operate the gripper easily. In Section V, we show that a network trained on this data can be used to control the position of a robot-actuated gripper.

We designed a capture protocol to systematically collect data from the grippers. We studied actions where the gripper makes contact with the surface of the planar pressure sensor. Our data is divided into three classes of actions: *make contact*, where both fingers of the gripper are lowered onto the surface, *slide*, where the gripper is translated along the surface, and *close gripper*, where the gripper is closed while in contact with the surface. We additionally collect *no contact* data, where the gripper held just above the surface without making contact to provide adversarial training data.

Data collection is further divided by the amount of force applied, the lighting configuration, the speed of the human operator, and the approach angle of the gripper. At least 30 seconds of data is collected for each combination of parameters, where the operator approaches the sensor and performs multiple grasps. Between individual grasps, the operator varied the translation and angle of the gripper. We record 32 actions each with 3 lighting conditions, resulting in 96 individual sequences for each gripper. Approximately 1 hour of data is collected for each gripper.

We randomly remove 20% of the sequences to create a held-out test set.

#### D. Network Architecture

We develop VPEC-Net, a neural network to estimate pressure in *image space*. The network uses a single RGB image as input and produces an estimated pressure for each input pixel. To generate ground truth pressure data for this approach, the data measured by the pressure sensor is warped

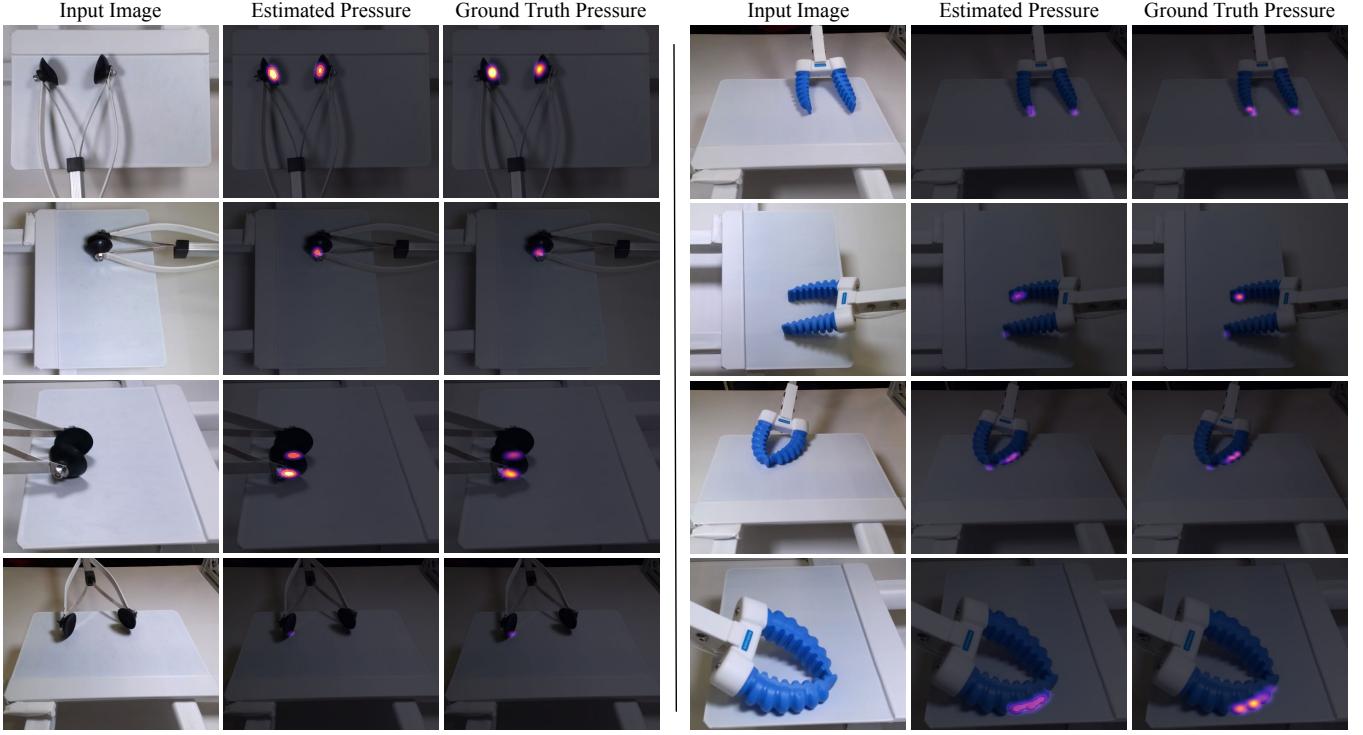


Fig. 3. **Left:** Examples of pressure estimation for the tendon-actuated gripper. **Right:** Examples of pressure estimation for the pneumatic gripper. **Input Image Column:** The input RGB image used by VPEC-Net to infer pressure. **Estimated Pressure Column:** The estimated pressure image overlaid on the input image. **Ground Truth Pressure Column:** The ground truth pressure measurements from a pressure sensing array overlaid on the input image.

Method	Temporal Acc.	Contact IoU	Volumetric IoU	MAE
Tendon-Actuated Gripper	95.9%	73.8%	58.2%	5.3 Pa
Pneumatic Gripper	95.1%	63.3%	52.0%	9.7 Pa

TABLE I  
RESULTS OF VISUAL PRESSURE ESTIMATION

into image space using a homography transform. This allows directly overlaying pressure information onto the image.

VPEC-Net’s architecture takes inspiration from image-to-image translation neural networks used in the semantic segmentation literature. For an input RGB image  $I$ , a pressure image  $\hat{P} = f(I)$  is estimated. The network uses an encoder-decoder architecture with skip connections. An SE-ResNeXt50 network [41], [42], [43], [44] with weights from pretraining on ImageNet [45] is used for the encoder, and an FPN network [46] is used for the decoder.

The task of pressure estimation is framed as a classification problem. The pressure range is divided into 8 discrete bins placed evenly in logarithmic space, including an additional *zero pressure* bin. For each pixel in the output pressure image, the network classifies which pressure bin the pixel should reside in. The network is trained with a cross-entropy loss. We tested various output representations and found that this outperformed a direct regression of a pressure scalar.

Images from the cameras are cropped to extend slightly past the edges of the pressure sensor. VPEC-Net is trained for 600k iterations using the Adam optimizer [47]. The learning rate is initially set at  $1e-3$ , which drops to  $1e-4$  after 100k

iterations. During training, several types of augmentations are used, including flips, random rotations, translations, scaling, brightness, and contrast changes.

#### IV. EVALUATION OF VISUAL PRESSURE ESTIMATION

To evaluate the performance of VPEC-Net, we perform evaluations on the held-out test set. We use a variety of evaluation metrics similar to [1] to quantify pressure estimation accuracy.

a) *Temporal Accuracy:* To evaluate the temporal accuracy of pressure estimates, if *any* pressure pixel is above a threshold of 1.0 kPa, the frame is marked as containing contact. Temporal Accuracy measures the consistency between the presence of ground truth and estimated contact.

b) *Contact IoU:* To determine the spatial and temporal accuracy of pressure estimates, binary contact images are generated by thresholding pressure at the same value used for *temporal accuracy*. The ground truth contact image and estimated contact image are compared to calculate intersection over union (IoU).

c) *Volumetric IoU:* To assess the magnitude of pressure estimates, we extend the Contact IoU to Volumetric IoU. This

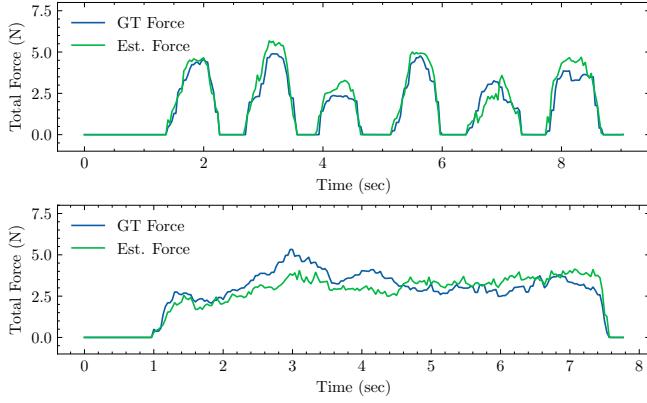


Fig. 4. Force estimates over time are visualized for *make contact* and *slide* sequences of the tendon-actuated gripper in the test set. While VPEC-Net displays some amount of error in the quantity of force exerted, it accurately captures the onset and termination of contact.

views pressure images as 3D volumes, with the height of the volume proportional to the quantity of pressure. The metric calculates the intersection over union of the two volumes and returns a percentage.

$$IoU_{vol} = \frac{\sum^{i,j} \min(P_{i,j}, \hat{P}_{i,j})}{\sum^{i,j} \max(P_{i,j}, \hat{P}_{i,j})} \quad (1)$$

*d) MAE:* To quantify the error in pressure in physical units, the mean absolute error is calculated across each pixel. As most of the pixels in the dataset contain zero pressure, the MAE is low compared to the peak pressure observed in the dataset.

#### A. Results

We train one network for each gripper and measure performance on the held-out test set. The results are reported in Table I. We also provide qualitative examples of the network pressure prediction in Figure 3.

Generally, VPEC-Net performs well at estimating pressure from a single image. Our approach can accurately estimate if the gripper is in contact with the surface or not, achieving a temporal accuracy  $> 95\%$  for both grippers.

We observe that the network trained on the tendon-actuated gripper outperforms the pneumatic gripper in all metrics. We hypothesize that this is because the shape of the pressure distribution created by the tendon-actuated gripper is often simpler (Figure 3). The tendon-actuated gripper also tends to visibly deform in a localized way, while deformation of the pneumatic gripper is less local and occurs across a wide area.

#### B. Limitations

While the network successfully reconstructs pressure on a flat surface, our dataset does not include objects with curved surfaces or unseen textures and only includes a limited set of action classes.

## V. ROBOTIC CONTROL OF PRESSURE

We evaluated VPEC-Net with robotic manipulation tasks involving pressure objectives and precision grasping. We first show that VPEC-Net can be used to modulate the pressure applied to a surface and increase the spatial accuracy of a compliant robot. We then show how our approach can be used to pick up small objects (penny, screw) that require precision manipulation. For all experiments, we used a Stretch RE1 mobile manipulator with its stock gripper.

#### A. Making Contact with a Desired Pressure

VPEC-Net can be used to regulate the amount of force a gripper exerts while in contact with a surface. We perform an experiment where the robot is commanded to exert a specified amount of normal force by lowering the gripper to make contact with a pressure sensor.

The pressure estimated by VPEC-Net is integrated with respect to area on the surface to acquire a total force estimate (Eqn. 2). The robot uses a simple bang-bang controller to modulate force with the surface by adjusting the height of the gripper using the Stretch RE1's lift joint.

$$\hat{F} = \int f(I)dA \quad (2)$$

Each trial begins with the robot placed 3-5 cm above the pressure sensor with VPEC-Net running on a single camera at a rate of 12 Hz. Once the pressure estimates indicate that the target force has been achieved, the ground truth force is measured with the pressure sensor. We conduct a total of 60 trials, with 10 trials being recorded for each force level ranging from 0 to 5N (Figure 5).

VPEC-Net can accurately estimate force at higher levels. However, it tends to underestimate forces in the range of 1 to 2N, near the boundary of contact. This may be due to

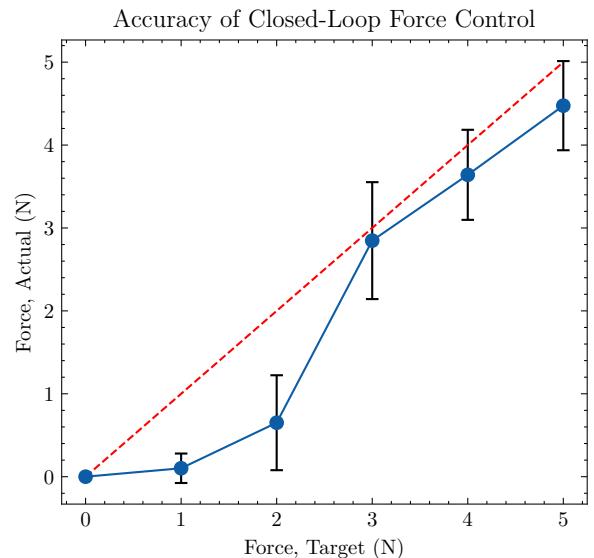


Fig. 5. We use a simple controller to achieve a target force, applying feedback from VPEC-Net's visual pressure estimation. The actual force is measured using the pressure sensor and matches the target value well at higher force levels.

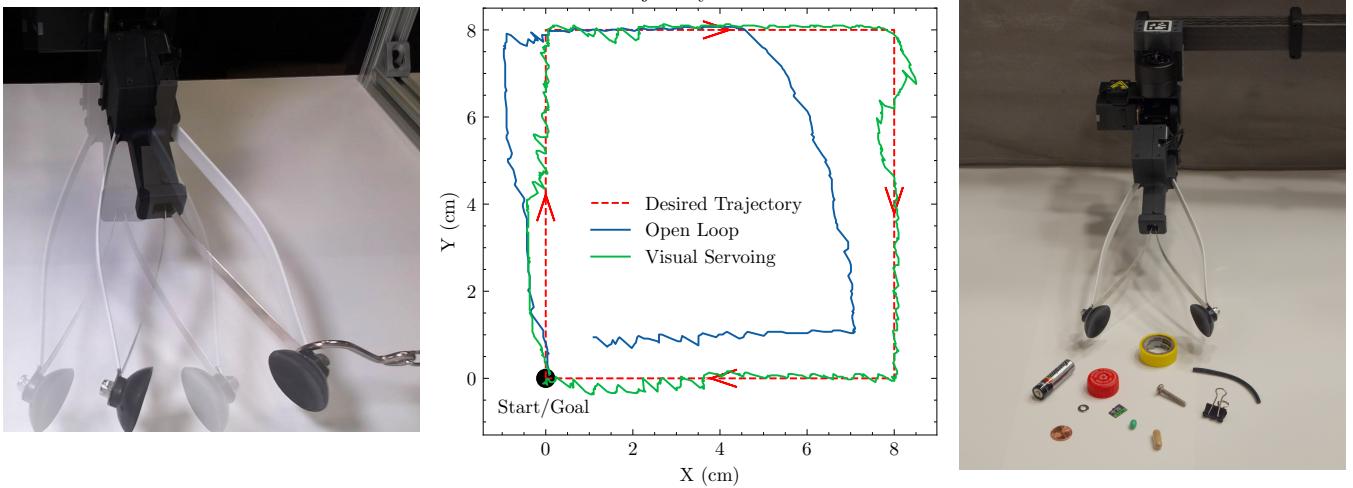


Fig. 6. **Left:** The fingertips of the tendon-actuated gripper deflect 4cm when subjected to 5N of lateral force due to deformation of the gripper. **Middle:** When commanded to trace a square path while in contact with a flat surface (red), a gripper using open-loop control accumulates significant error (blue). Feedback control using image space pressure estimates reduces tracking error (green). **Right:** In left to right order, this image shows the objects from our grasping evaluation: AA battery, penny, washer, bottle cap, microSD card, small green pill, tape roll, large pill, screw, cable segment, and binder clip.

differences between the manually operated gripper used for training and the robotic gripper used for testing.

### B. Following a Spatial Pressure Trajectory

Due to the inherent compliance of soft grippers, the precise pose of the gripper can be difficult to control. This is especially true when in contact with a surface, as deformation and friction with the surface can cause the gripper to stick and slip. Figure 6a shows that the gripper has significant deflection in response to an external disturbance. Additionally, to move the gripper laterally, the Stretch RE1 drives on a carpeted floor with its differential drive mobile base, which can result in movement variations and inaccurate positioning due to wheel slip and other phenomena.

We show that an open-loop controller accumulates significant error while executing a trajectory in contact (Figure 6b, blue). The robot gripper was rotated and lowered to a constant height such that one fingertip was in contact with the surface. The robot was then commanded to move in an 8cm square path. The true path of the gripper was measured by calculating the center of pressure detected by the pressure sensor.

To achieve a higher accuracy, we use an image-based visual servoing (IBVS) controller [28] that leverages the image space pressure estimates from VPEC-Net (Figure 6b, green). The error function  $E(t)$  uses the position of the maxima in the estimated pressure image,  $M(t)$ , and a desired target position in image space  $T$ .

$$E(t) = \begin{bmatrix} e_x \\ e_y \end{bmatrix} = \begin{bmatrix} T_x - M_x(t) \\ T_y - M_y(t) \end{bmatrix} \quad (3)$$

This error is transformed into robot actuator commands  $\dot{q}$  with the image Jacobian  $J$  and a gain  $\lambda$ :  $\dot{q}(t) = \lambda J^+ E(t)$ , where  $J^+$  is the pseudo-inverse. Because we observe both

the target and gripper contact location in the same image, our controller is ‘endpoint closed-loop’ [28] and robust to inaccuracies in  $J$ .

### C. Grasping Small Low-Profile Objects

To demonstrate the real-world value of VPEC, we perform grasping trials with a range of objects (Figure 6c), including very thin objects. Humans typically grasp these small objects by using their fingers to first make contact with the surface near the object, then slide their fingertips to close around the object. We take inspiration from this approach and design a robot control algorithm to grasp objects while maintaining contact with the surface.

The robot must autonomously approach the object, grasp it, and pick it up without dropping it for 5 seconds (Figure 7). Trials where any of the robot’s actuators exceed their torque limits are marked as failures. We remove the pressure sensor from our capture rig (Figure 2) during grasping experiments, providing the robot with a larger workspace and demonstrating that VPEC-Net can generalize to a new surface. We conduct 10 grasping attempts for each object. The object is reset to a random position and orientation after each trial.

Our system used a simple color thresholding algorithm to find the centroid of the object in the image. The robot starts with the gripper positioned above the surface and is lowered until pressure above a threshold is reported by VPEC-Net. The normal force exerted by the gripper is continuously controlled to maintain a set force (Sec V-A).

We then perform visual servoing (Sec V-B) to grasp the object. Our algorithm attempts to navigate the mean position of the two local pressure maxima produced by the gripper fingertips to the object centroid in image space. Once the average fingertip position is within a fixed radius of the object centroid, the gripper is closed and lifted.

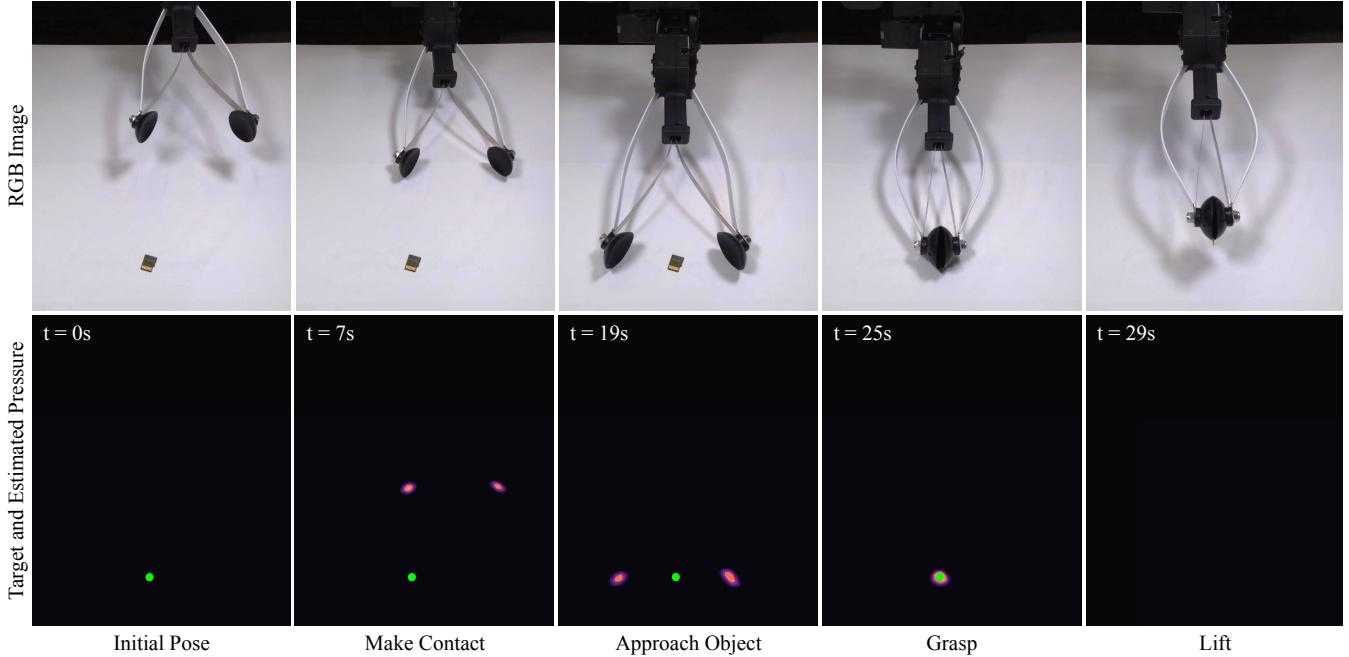


Fig. 7. **Left to Right:** Grasping a 1mm thick microSD card. **t=0s:** The centroid of the object shown as a green circle is estimated using the RGB input image. **t=7s:** The robot makes contact with the uninstrumented tabletop to achieve a desired pressure and estimate its location. Fingertip contact results in two ellipsoidal contact pressure regions. **t=19s:** The robot moves the estimated fingertip pressure regions to the centroid of the object. **t=25s and t=29s:** The gripper closes while maintaining a desired pressure on the surface to grasp and then pick up the microSD card.

#### D. Grasping Results

We find that VPEC allows the robot to accomplish precision grasping using images from a single RGB camera. The robot is able to grasp all 11 objects in our set (Table II), and achieves an average success rate of 93%. The robot is also able to maintain contact with the surface, allowing the visual servoing controller to accurately track the position of the robot in image space and modulate normal force to grasp thin objects on a flat surface.

Failures during grasping experiments can be attributed to a few causes. As our dataset was collected without distractor objects present, when objects are placed in the camera's field of view, the network occasionally estimates pressure near the object in image space. This extra pressure estimate may cause the gripper to lift off the surface. We also find that the network may occasionally overestimate the gripper pressure,

also causing the gripper to be in inconsistent contact with the surface. In very rare cases, pressure is underestimated, causing the gripper to be driven into the surface such that the motor torque limits are exceeded and the trial is stopped. We would expect additional training data to increase robustness and alleviate these issues.

#### VI. CONCLUSION

We present VPEC, a method to visually estimate pressure from changes in the appearance of a soft gripper. We demonstrate that a trained model can accurately estimate pressure for two designs: a tendon-actuated gripper and a pneumatic gripper. These pressure estimates can be used to perform closed-loop control of a robot to maintain a desired pressure, accurately trace a trajectory, and successfully manipulate small objects. Our results suggest that visual estimation of pressure is a promising approach for soft robotic grippers.

#### REFERENCES

- [1] P. Grady, C. Tang, S. Brahmhatt, C. D. Twigg, C. Wan, J. Hays, and C. C. Kemp, "PressureVision: estimating hand pressure from a single RGB image," *European Conference on Computer Vision (ECCV)*, 2022.
- [2] M. Kazemi, J.-S. Valois, J. A. Bagnell, and N. Pollard, "Human-inspired force compliant grasping primitives," *Autonomous Robots*, vol. 37, no. 2, pp. 209–225, 2014.
- [3] C. Eppner, R. Deimel, J. Alvarez-Ruiz, M. Maertens, and O. Brock, "Exploitation of environmental constraints in human and robotic grasping," *The International Journal of Robotics Research*, vol. 34, no. 7, pp. 1021–1038, 2015.
- [4] M. Ciocarlie, F. M. Hicks, R. Holmberg, J. Hawke, M. Schlicht, J. Gee, S. Stanford, and R. Bahadur, "The Velo gripper: A versatile single-actuator design for enveloping, parallel and fingertip grasps," *The International Journal of Robotics Research*, vol. 33, no. 5, pp. 753–767, 2014.

TABLE II  
OBJECT DIMENSIONS AND GRASPING RESULTS

Object	Dims. L×W×H	Grasp Successes/Trials
Washer	10×10×1 mm	9/10
Small Green Pill	10×6×6 mm	10/10
Large Pill	21×8×8 mm	9/10
MicroSD Card	15×11×1 mm	8/10
Cable Segment	82×4×4 mm	10/10
Penny	19×19×1.5 mm	9/10
Bottle Cap	30×30×13 mm	9/10
AA Battery	50×14×14 mm	9/10
Binder Clip	25×24×19 mm	9/10
Screw	32×9×9 mm	10/10
Tape Roll	36×36×13 mm	10/10

- [5] V. Babin and C. Gosselin, "Picking, grasping, or scooping small objects lying on flat surfaces: A design approach," *The International Journal of Robotics Research*, vol. 37, no. 12, 2018.
- [6] D. Yoon and Y. Choi, "Analysis of fingertip force vector for pinch-lifting gripper with robust adaptation to environments," *IEEE Transactions on Robotics*, vol. 37, no. 4, pp. 1127–1143, 2021.
- [7] C. Eppner and O. Brock, "Visual detection of opportunities to exploit contact in grasping using contextual multi-armed bandits," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 273–278.
- [8] M. Pozzi, S. Marullo, G. Salvietti, J. Bimbo, M. Malvezzi, and D. Prattichizzo, "Hand closure model for planning top grasps with soft robotic hands," *The International Journal of Robotics Research*, vol. 39, no. 14, pp. 1706–1723, 2020.
- [9] A. Gupta, C. Eppner, S. Levine, and P. Abbeel, "Learning dexterous manipulation for a soft robotic hand from human demonstrations," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 3786–3793.
- [10] S. Begej, "Planar and finger-shaped optical tactile sensors for robotic applications," *IEEE Journal on Robotics and Automation*, vol. 4, no. 5, pp. 472–484, 1988.
- [11] R. Li, R. Platt, W. Yuan, A. ten Pas, N. Roscup, M. A. Srinivasan, and E. Adelson, "Localization and manipulation of small parts using gelsight tactile sensing," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2014, pp. 3988–3993.
- [12] A. Yamaguchi and C. G. Atkeson, "Combining finger vision and optical tactile sensing: Reducing and handling errors while cutting vegetables," in *2016 IEEE-RAS 16th International Conference on Humanoid Robots*. IEEE, 2016, pp. 1045–1051.
- [13] N. Kuppuswamy, A. Alspach, A. Uttamchandani, S. Creasey, T. Ikeda, and R. Tedrake, "Soft-bubble grippers for robust and perceptive manipulation," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 9917–9924.
- [14] N. F. Lepora, "Soft biomimetic optical tactile sensing with the TacTip: A review," *IEEE Sensors Journal*, 2021.
- [15] W. Kuang, M. Yip, and J. Zhang, "Vibration-based multi-axis force sensing: Design, characterization, and modeling," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 3082–3089, 2020.
- [16] G. S. Koonjul, G. J. Zeglin, and N. S. Pollard, "Measuring contact points from displacements with a compliant, articulated robot hand," in *2011 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2011, pp. 489–495.
- [17] S. Wang, A. Bhatia, M. T. Mason, and A. M. Johnson, "Contact localization using velocity constraints," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 7351–7358.
- [18] A. A. Nazari, F. Janabi-Sharifi, and K. Zareinia, "Image-based force estimation in medical applications: A review," *IEEE Sensors Journal*, vol. 21, no. 7, pp. 8805–8830, 2021.
- [19] C. W. Kennedy and J. P. Desai, "A vision-based approach for estimating contact forces: Applications to robot-assisted surgery," *Applied Bionics and Biomechanics*, vol. 2, no. 1, pp. 53–60, 2005.
- [20] E. Noohi, S. Parastegari, and M. Žefran, "Using monocular images to estimate interaction forces during minimally invasive surgery," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2014, pp. 4297–4302.
- [21] D. Kim, H. Cho, H. Shin, S.-C. Lim, and W. Hwang, "An efficient three-dimensional convolutional neural network for inferring physical interaction force from video," *Sensors*, vol. 19, no. 16, p. 3579, 2019.
- [22] A. Marban, V. Srinivasan, W. Samek, J. Fernández, and A. Casals, "A recurrent convolutional neural network approach for sensorless force estimation in robotic surgery," *Biomedical Signal Processing and Control*, vol. 50, pp. 134–150, 2019.
- [23] Z. Chua, A. M. Jarc, and A. M. Okamura, "Toward force estimation in robot-assisted surgery using deep learning with vision and robot state," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 12 335–12 341.
- [24] M. A. Greminger and B. J. Nelson, "Modeling elastic objects with neural networks for vision-based force measurement," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, vol. 2. IEEE, 2003, pp. 1278–1283.
- [25] Y. Li, J.-Y. Zhu, R. Tedrake, and A. Torralba, "Connecting touch and vision via cross-modal prediction," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 10 609–10 618.
- [26] B. S. Zapata-Impata, P. Gil, Y. Mezouar, and F. Torres, "Generation of tactile data from 3d vision and target robotic grasps," *IEEE Transactions on Haptics*, vol. 14, no. 1, pp. 57–67, 2020.
- [27] K. Patel, S. Iba, and N. Jamali, "Deep tactile experience: Estimating tactile sensor output from depth sensor data," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 9846–9853.
- [28] S. Hutchinson, G. D. Hager, and P. I. Corke, "A tutorial on visual servo control," *IEEE Transactions on Robotics and Automation*, vol. 12, no. 5, pp. 651–670, 1996.
- [29] P. M. Fitzpatrick and E. R. Torres-Jara, "The power of the dark side: using cast shadows for visually-guided touching," in *4th IEEE/RAS International Conference on Humanoid Robots, 2004.*, vol. 1. IEEE, 2004, pp. 437–449.
- [30] B. Calli and A. M. Dollar, "Robust precision manipulation with simple process models using visual servoing techniques with disturbance rejection," *IEEE Transactions on Automation Science and Engineering*, vol. 16, no. 1, pp. 406–419, 2018.
- [31] A. S. Morgan, B. Wen, J. Liang, A. Bouliarias, A. M. Dollar, and K. Bekris, "Vision-driven compliant manipulation for reliable, high-precision assembly tasks," *Robotics: Science and Systems, (RSS)*, 2021.
- [32] P. Sikka, H. Zhang, and S. Suphen, "Tactile servo: Control of touch-driven robot motion," in *Experimental Robotics III*. Springer, 1994, pp. 219–233.
- [33] N. Chen, H. Zhang, and R. Rink, "Edge tracking using tactile servo," in *Proceedings 1995 IEEE/RSJ International Conference on Intelligent Robots and Systems. Human Robot Interaction and Cooperative Robots*, vol. 2. IEEE, 1995, pp. 84–89.
- [34] Q. Li, C. Schürmann, R. Haschke, and H. J. Ritter, "A control framework for tactile servoing," in *Robotics: Science and Systems, (RSS)*. Citeseer, 2013.
- [35] C.-T. Wen, S. Arai, J. Kinugawa, and K. Kosuge, "Tactile servoing based pressure distribution control of a manipulator using a convolutional neural network," *IEEE Access*, vol. 9, pp. 117 132–117 139, 2021.
- [36] N. F. Lepora and J. Lloyd, "Pose-based tactile servoing: Controlled soft touch using deep learning," *IEEE Robotics & Automation Magazine*, vol. 28, no. 4, pp. 43–55, 2021.
- [37] C. C. Kemp, A. Edsinger, H. M. Clever, and B. Matulevich, "The design of Stretch: A compact, lightweight mobile manipulator for indoor human environments," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2022.
- [38] SoftGripping by Wegard GmbH. (2022) SoftGripping, the modular design system for flexible gripping. [Online]. Available: <https://soft-gripping.com/>
- [39] Sensel, "Sensel Morph haptic sensing tablet," <https://morph.sensel.com/>, Last accessed on 2022-02-22.
- [40] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and M. J. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognition*, vol. 47, no. 6, pp. 2280–2292, 2014.
- [41] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition, (CVPR)*, 2016, pp. 770–778.
- [42] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *2018 IEEE Conference on Computer Vision and Pattern Recognition, (CVPR)*, 2018.
- [43] S. Xie, R. B. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition, (CVPR)*, 2017.
- [44] P. Yakubovskiy, "Segmentation models pytorch," [https://github.com/qubvel/segmentation\\_models.pytorch](https://github.com/qubvel/segmentation_models.pytorch), 2020.
- [45] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition, (CVPR)*. IEEE, 2009, pp. 248–255.
- [46] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *IEEE Conference on Computer Vision and Pattern Recognition, (CVPR)*, 2017, pp. 2117–2125.
- [47] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, (ICLR) 2015*, Y. Bengio and Y. LeCun, Eds., 2015.