IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

9DTact: A Compact Vision-Based Tactile Sensor for Accurate 3D Shape Reconstruction and Generalizable 6D Force Estimation

Changyi Lin, Han Zhang, Jikai Xu, Lei Wu and Huazhe Xu

***Abstract*—The advancements in vision-based tactile sensors have boosted the aptitude of robots to perform contact-rich manipulation, particularly when precise positioning and contact state of the manipulated objects are crucial for successful execution. In this work, we present 9DTact, a straightforward yet versatile tactile sensor that offers 3D shape reconstruction and 6D force estimation capabilities. Conceptually, 9DTact is designed to be highly compact, robust, and adaptable to various robotic platforms. Moreover, it is low-cost and easy-to-fabricate, requiring minimal assembly skills. Functionally, 9DTact builds upon the optical principles of DTact and is optimized to achieve 3D shape reconstruction with enhanced accuracy and efficiency. Remarkably, we leverage the optical and deformable properties of the translucent gel so that 9DTact can perform 6D force estimation without the participation of auxiliary markers or patterns on the gel surface. More specifically, we collect a dataset consisting of approximately 100,000 image-force pairs from 175 complex objects and train a neural network to regress the 6D force, which can generalize to unseen objects. To promote the development and applications of vision-based tactile sensors, we open-source both the hardware and software of 9DTact, along with a comprehensive video tutorial, all of which are available at https://linchangyi1.github.io/9DTact.**

***Index Terms*—Force and Tactile Sensing; Perception for Grasping and Manipulation**

I. Introduction

TACTILE sensing, which provides physical properties and spatial state of contact objects, is vital for robots to interact with the real world. With the help of cameras, vision-based tactile sensors [1, 2, 3, 4, 5] are able to sense the deformation of the gel surface approaching human-scale resolution. The acquired high-resolution information enables robots to perform stable and accurate robotic manipulation tasks such as object insertion [6, 7] and cable manipulation [8, 9]. Despite the potential benefits of these sensors, their adoption within the robotics community has been limited due to various factors such as the lack of compactness, the complexity of fabrication, high cost of acquisition, fragility and instability during use, and deficient functional capabilities.

In this work, we present 9DTact, a vision-based tactile sensor equipped with the following merits, aiming to overcome the drawbacks of previous counterparts.

- **Hardware.** Our sensor not only excels in specifications, it can also be fabricated in a convenient manner. With iterations of optimization in illumination, structure, crafts, and materials, 9DTact is designed to be compact, robust, and adaptable to various robotic platforms. Furthermore, 9DTact is an affordable and easily assembled sensor that only requires accessible components, standard machining processes, and minimal assembly skills.
- **Software.** 9DTact is a versatile sensor capable of both accurate 3D shape reconstruction and generalizable 6D (1D normal, 2D shear, and 3D torque) force estimation. The new design and simple calibration method improve the accuracy and efficiency of 3D shape reconstruction. Inspired by the principles of DTact [1] that pixels corresponding to thinner areas become darker, we observe another interesting phenomenon that pixels corresponding to bulging areas become brighter, and the in-plane motion of the contact object induces accompanying movement of the brighter pixels. Based on this finding, we extract a dense deformation representation from the original tactile image for force estimation, where no auxiliary markers or patterns are needed. Making use of a neural network trained on approximately 100,000 pairs of deformation representation and 6D force sampled from 175 objects, 9DTact could estimate accurate 6D force with generalization to unseen geometries and objects.
- **Open-Source.** We would like to clear the obstacles as possible for building and utilizing tactile sensors such as 9DTact in the robotics community. Hence, we open-source everything about 9DTact including its design files, codes, datasets, and pre-trained models. Furthermore, we also provide a comprehensive video tutorial that documents the entire process of replicating a 9DTact sensor, including a bunch of experiences for simplifying and improving the manufacturing processes.

The remainder of this paper is organized as follows. We introduce related work on hardware designs and force estimation methods of vision-based tactile sensors in Section II. The details of 9DTact design are described in Section III. We then introduce the improvements of 3D shape reconstruction in Section IV. Next, we present the principle, implementation, and performance of 6D force estimation in Section V. Finally, the conclusion is summarized in Section VI.

Manuscript received: July, 24, 2023; Revised October, 11, 2023; Accepted November, 25, 2023. This paper was recommended for publication by Editor A. Banerjee upon evaluation of the Associate Editor and Reviewers' comments. (*Corresponding author: Huazhe Xu*)

Changyi Lin is with Shanghai Qi Zhi Institute, Shanghai 200030, China, and also with Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing 100084, China. (e-mail: changyil@andrew.cmu.edu)

Han Zhang is with Department of Electronic Engineering, Tsinghua University, Beijing 100084, China, and also with Shanghai Qi Zhi Institute, Shanghai 200030, China. (e-mail: zhanghan14@mails.tsinghua.edu.cn)

Jikai Xu and Lei Wu are with School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China, and also with Shanghai Qi Zhi Institute, Shanghai 200030, China. (e-mail: jikai_xu@hust.edu.cn; lei_wu@hust.edu.cn)

Huazhe Xu is with the Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing 100084, China, and also with Shanghai Qi Zhi Institute, Shanghai 200030, China, as well as with Shanghai Artificial Intelligence Laboratory, Shanghai 200032, China. (e-mail: huazhe_xu@mail.tsinghua.edu.cn)

Digital Object Identifier (DOI): see top of this page.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

II. Related Work

A. Compact Vision-Based Tactile Sensors for 3D Shape Reconstruction

Vision-based tactile sensor GelSight [2] leverages the photometric stereo technique [18] to achieve 3D shape reconstruction of its sensing surface. Following, many GelSight-like sensors [10, 19, 3] improve the designs to be more compact to mount them in grippers or dexterous hands. However, their reliance on the photometric stereo technique, which is highly dependent on the uniformity and reflection of the internal illumination, makes them challenging to replicate due to strict requirements on material preparation, fabrication processes, and assembly skills. Consequently, researchers lacking hardware experience must purchase commercial products [4, 5].

To address the manufacturing challenges mentioned above, DTact [1] leverages the reflection property of translucent elastomer for 3D shape reconstruction, which has demonstrated comparable accuracy, superior robustness and surface shape extensibility. Based on this promising principle, 9DTact is carefully designed to simultaneously possess exceptional physical characteristics including compactness, robustness, and affordability as highlighted in Table I. Moreover, 9DTact is easy-to-fabricate and open-sourced, which expands its potential as a general vision-based tactile sensor.

B. Deformation Representation for 6D Force Estimation

On vision-based tactile sensors with gel surfaces, an applied force induces deformation. Therefore, the methods for force estimation generally consist of the following three key elements.

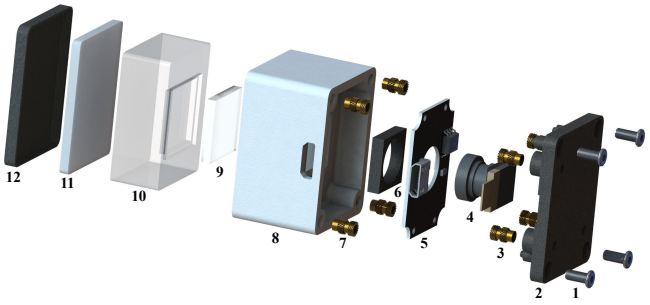
- **Deformation visualizer:** the physical medium for visualizing the full-dimensional deformation of the gel into visual features for the camera.
- **Deformation representation:** the information extracted from the visual features.
- **Inferring method:** the method for decoupling force from the deformation representation.

As outlined in Table II, tactile sensors capable of 3D shape reconstruction often incorporate a marker array or a pattern on their sensing surface as a common method for force estimation. GelSight [2] employs Convolutional Neural Networks (CNNs) [11] to predict force from tactile images, while GelSlim [12] utilizes the inverse Finite Element Method (iFEM) [13] to infer force based on the 3D motions of markers. However, the sparsity of the marker array limits its capacity to fully capture deformation across the entire gel surface, leading to validations on only a few selected objects with simple geometries. Similar deformation representation is used in DelTact [14] but with more data for computing the coefficient matrix. To enrich the information contained in the deformation representation, DenseTact 2.0 [17] develops a special fabrication process to paint a continuous pattern, and employs CNNs for predicting 6D force.

Although 9DTact could potentially adopt the method of painting pattern for force estimation, there are inherent drawbacks with this method. Specifically, the painted pattern, being non-reflective to directional lights, compromises the sensor's 3D shape reconstruction capability within the regions it covers [19], meaning that there is a trade-off between the two functionalities. Moreover, crafting the pattern necessitates specialized equipment and expertise. Consequently, the method of painting pattern is incompatible with our objectives of maintaining the accuracy of 3D shape reconstruction and simplifying the sensor fabrication.

Surprisingly, we find that 9DTact, without any painted pattern, naturally possesses a dense gel flow. Such flow is innate from the optical and deformable properties of the translucent gel. Moreover, the flow reflects the full-dimensional deformation of the gel. This phenomenon not only helps extract dense deformation representations to perform 6D force estimation with generalization to unseen objects, but also preserves the 3D shape reconstruction quality, and simplifies the fabrication.

(a) Exploded View:



(b) Schematic Diagram:

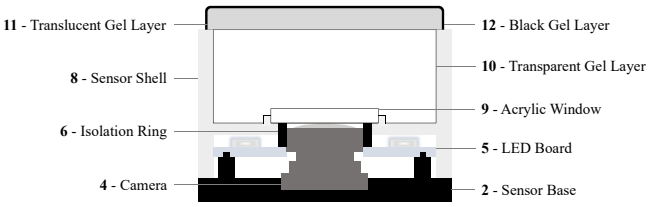


Figure 1: Design of 9DTact. (a) The exploded view of 9DTact. The components labeled as 1 are screws for connecting the sensor base and the sensor shell. The components labeled as 3 and 7 are both heat-set threaded inserts. (b) The schematic diagram of 9DTact.

III. 9DTact Sensor Design

A. Design Goals

Throughout the fabrication, installation, and utilization of the sensor, we aim to achieve the following objectives. To lower the barrier of fabrication, the components should be readily accessible and the assembly process should require minimal professional expertise. For installation, the sensor should possess a compact structural configuration that allows installation in constrained spaces, such as within the fingertips of dexterous hands. Regarding utilization, the sensor should exhibit robustness in both physical and functional aspects, as well as adaptability to various computing platforms.

In the following section, we will demonstrate how we achieve these goals by providing a detailed description of each component shown in Fig. 1 (b).

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

Table I: Comparison of GelSight, GelSlim 3.0, DIGIT, GelSight-Mini, DTact, and 9DTact. (*Manufacturing of 1000 pieces. †Commodity price.)

Sensor	Dimension $[mm^3]\downarrow$	Sensing Area $[mm^2]\uparrow$	D/A Ratio $[mm]\downarrow$	Weight $[g]\downarrow$	FPS $\uparrow$	Cost[$] $\downarrow$
GelSight [10]	$80 \times 40 \times 40 = 128000$	252	508	NA	**90**	30
GelSlim 3.0 [3]	$80 \times 37 \times 20 = 59200$	**675**	88	45	**90**	25*
DIGIT [4]	$36 \times 26 \times 33 = 30888$	$19 \times 16 = 304$	102	**20**	60	15* / 300†
GelSight-Mini [5]	$32 \times 28.5 \times 28 = 25536$	$19 \times 15 = 285$	90	20.8	25	499†
DTact [1]	$45 \times 45 \times 47 = 95175$	$24 \times 24 = 576$	165	78	60	34
9DTact (Ours)	$32.5 \times 25.5 \times 25.5 =$ **21133**	$24 \times 18 = 432$	**49**	**20**	**90**	**15**

Table II: Comparison of the methods and configurations utilized for force estimation by GelSlim 2.0, DelTact, DenseTact 2.0, and 9DTact.

Sensor	Deformation Visualizer	Deformation Representation	Infer Method	Validation Objects	Collection State
GelSight [2]	Black marker array	Marker pixels in raw image	CNNs [11]	3 simple objects	Dynamic
GelSlim 2.0 [12]	Black marker array	3D motions of markers	iFEM [13]	1 sphere	Static
DelTact [14, 15]	Colorful dense pattern	Vectors field from optical flow	NHHD [16]	5 spheres	Static
DenseTact 2.0 [17]	Black randomized pattern	Pattern pixels in raw image	CNNs	10 simple objects	Static
9DTact (ours)	Only original gel	Dense gel flow image	CNNs	175 complex objects	Dynamic

B. Details of the Components

Camera. In order to capture the sensor's contact surface as comprehensively as possible, we select an OV5647 camera with a wide Field Of View (FOV) of 160 degrees and attach it to the sensor base using 3M glue. The camera occupies a small space and is also adaptable to various computing platforms. It can be connected to the Camera Serial Interface (CSI) port of Raspberry Pi Zero directly or to the Universal Serial Bus (USB) port with an off-the-shelf CSI-to-USB transformation board. In this work, we choose the latter connection format to use the camera with a desktop.

LED board. In the previous version, DTact sensor [1], the utilization of an LED ring may produce uneven illumination, characterized by a brighter light intensity at the center than that at the periphery. To mitigate this problem, we design a compact LED board with eight LEDs evenly arranged in a rectangular shape on it as shown in Fig. 2 (a). Furthermore, we incorporate a CN5711 integrated circuit to regulate the current inputs for the LEDs, which helps to provide stable and consistent illumination for the sensor. The LED board is secured to the sensor base by means of four locating holes and is powered by a 5V USB port.

Sensor base. The sensor base, which is used to secure the camera and the LED board, is 3D printed with black nylon material (HP3DHR-PA12) that has high strength and toughness. Furthermore, four M2 heat-set threaded inserts (labeled as 3 in Fig. 1 (a)) are installed in the sensor base, serving as connectors between the 9DTact sensor and other platforms such as robot grippers and dexterous hand fingers.

Sensor shell. Attaching with the acrylic window that provides a clear window for the camera, the sensor shell serves as a container for the transparent gel layer. It is 3D printed with white nylon (FS3300PA) material which exhibits superior durability. Since the white nylon material is not opaque, the inner base layer of the sensor shell allows light to transmit from the LED board and also helps to diffuse the light. Four M2 heat-set threaded inserts (labeled as 7 in Fig. 1 (a)) are also mounted in the bottom of the sensor shell so that it can be connected to the sensor base with four M2 screws.

Transparent gel layer. The transparent gel layer not only facilitates the diffusion of light to a more uniform distribution, but also serves as a transitional propagation medium with optical properties similar to those of the translucent gel layer, which helps to mitigate the issue of excessive reflection that can occur when light passes through air or other media. Compared to ELASTOSIL® RT 601 silicone used in [1], Hongye Jie® 9345 silicone (mixing ratio 1:1, shore A hardness 45) is easier to remove air bubbles with a vacuum pump. Therefore, we choose it as the material of the transparent gel layer. The mixed bubble-free silicone is poured into the sensor shell until filled. Due to the inadequate levelness of the base surface of the thermostatic oven, the sensor shell is initially placed on a horizontal optical platform at a room temperature of 25 ℃ for 4 hours to allow the silicone to solidify. Subsequently, the sensor shell is transferred to the thermostatic oven maintained at 50 ℃ for a duration of 6 hours to ensure complete hardening of the silicone.

Acrylic window. In the DTact sensor [1], the acrylic window is designed to align with the inner dimensions of the sensor shell. As a result, an air gap often exists between the transparent gel layer and the acrylic window after the transparent silicone cures, as Fig. 2 (b) illustrates. This is because the rough inner walls of the sensor shell have much higher adsorption capacity than that of the smooth surface of the acrylic window. The air gap is squeezed out when the sensor comes into contact with objects; thus the tactile image brightens overall because the light from the translucent gel layer transmits to the camera without decaying through the air gap. To eliminate this air gap, we reduce the size of the acrylic window, so that the rough inner base of the sensor shell increases the downward adsorption force to the transparent gel layer. The acrylic window is securely attached to the nested frame of the sensor shell with waterproof glue.

Translucent gel layer. The translucent gel layer is used to reflect light, which forms the fundamental principle of the 3D shape reconstruction function of 9DTact. In the previous version as described in [1], a mold is attached to the sensor shell, and the translucent silicone is poured into the mold and left to cure to the translucent gel layer. Although this method is relatively easy to fabricate and install, it has some

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

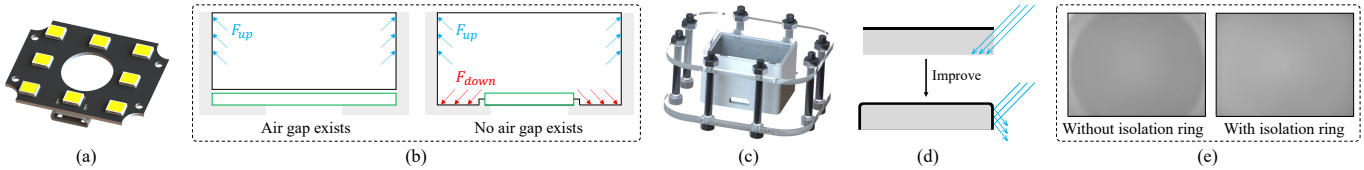


Figure 2: Design improvements of 9DTact. (a) Eight LEDs distribute in a rectangular shape on the LED board. (b) The acrylic window is resized to avoid the generation of an air gap. (c) The mold used for fabricating the translucent gel layer is secured with eight screws. (d) The four sides of the translucent gel layer are also coated with black gel to prevent ambient light from entering. (e) Two reference images captured without and with installing the isolation ring.

drawbacks. For instance, the thickness of the sensor shell is increased due to the requirement of four holes to install the mold. Furthermore, the success rate of curing the translucent silicone is reduced, since only four screws can not provide sufficient force to secure the mold in place. To address these issues, in 9DTact, we develop a base board and attach it to the sensor shell through the four heat-set threaded inserts. The acrylic mold for the translucent gel layer is then connected to the base board using eight pairs of M3 screws and nuts, as shown in Fig. 2 (c). The tightness resulting from this fastening mechanism ensures that the mold remains stable throughout fabrication, thus strongly guaranteeing successful outcomes. In addition, to augment the adhesive force of the translucent gel layer, its size is intentionally designed to exceed the inner dimensions of the sensor shell as Fig. 1 (b) shows. The material for the translucent gel layer is translucent Posilicone® DRSGJ02 silicone (mixing ratio: 1: 1, shore A hardness 5). After pouring the mixed bubble-free silicone into the mold, it takes 6 hours for the silicone to cure at 25 ℃.

Black gel layer. The black gel layer serves two purposes. First, it absorbs the inner light that transmits through the translucent gel layer. Second, it prevents ambient light from entering the translucent gel layer. However, in the case of the DTact sensor [1], only the upper surface of the translucent gel layer is coated with a thin black gel layer. This inevitably results in ambient light transmitting into the translucent gel layer through its four sides, leading to fluctuations in the brightness of tactile images. Therefore, we replace the mold for the translucent gel layer after it cures with one that has larger thickness and inner frame size, which ensures that the four sides of the translucent gel layer are also covered with black gel, as shown in Fig. 2 (d). Instead of using the same silicone as the translucent gel layer for the black gel layer, we opt for Smooth-On® Ecoflex 00-30 silicone (mixing ratio: 1: 1, shore 00 hardness 30) for its superior durability. To give it a black color, we add some black silicone pigment to the mixture. With the LED board lighting from below, we apply the black silicone onto the translucent gel layer until it completely blocks the light.

Isolation ring. The surface of the acrylic window reflects light from the vertical surfaces of the inner base layer of the sensor shell, which can interfere with the light transmission in the peripheral areas of the translucent gel layer. This is because the light from the inner base layer is much stronger than that from the translucent gel layer. As a result, the camera loses its ability to sense changes in light from these areas of the contact surface as the left image in Fig. 2 (e) shows. To this end, we add the isolation ring, 3D printed with black nylon material (HP3DHR-PA12), to prevent such strong light from transmitting from the vertical surfaces of the inner base layer to the acrylic window. The right image in Fig. 2 (e) shows the corrected sensed tactile surface.

Table III: The bill of materials (BOM) for fabricating a 9DTact sensor.

Component	Description	Process
Glue	YLG-YKL500	Off-the-shelf
Nuts	8 M2-3-5, 8 M3	
Screws	4 M2-6, 8 M3-30	
Camera	Frank-S15-V1.0-160°	
LED Board	$28 \times 21 \times 4mm$	PCB soldering
Isolation Ring	Black nylon (HP3DHR-PA12)	3D printing
Sensor Base	Black nylon (HP3DHR-PA12)	
Sensor Shell	White nylon (FS3300PA)	
Acrylic Window	$2mm$ thick acrylic board	Laser cutting
Base Board	$3mm$ thick acrylic board	
The First Mold	$2.5mm$ thick acrylic board	
The Second Mold	$2.8mm$ thick acrylic board	
Transparent Gel	Hongye Jie® 9345	Silicone processing
Translucent Gel	Posilicone® DRSGJ02	
Black Gel	Smooth-On® Ecoflex 00-30	

C. Conclusion and Comparison of the Sensor Design

Table I compares 9DTact with existing flattened compact vision-based sensors, while Table III summarizes some detailed information for each component. Here, we summarize the outstanding characteristics of 9DTact:

- **Compact.** The 9DTact sensor is remarkably compact, with the size of only $32.5\text{mm} \times 25.5\text{mm} \times 25.5\text{mm}$, which is approximately 22% the size of the DTact sensor [1]. With the smallest volume among the sensors listed in Table I, 9DTact is adaptable to be installed in a wide range of robotic platforms, from grippers to dexterous hands. Moreover, 9DTact features a relatively large sensing area, resulting in the smallest dimension-to-sensing area (D/A) ratio of all the sensors in Table I.
- **Robust.** Our improvement in the black gel layer makes 9DTact robust to dynamic ambient light. Furthermore, in Section V-B, we press the objects with sharp geometries against a single 9DTact sensor to collect over $100,000$ images. Remarkably, the sensor factors, such as imaging and illumination, remain stable throughout the experiment, and the contact surface shows no visible signs of damage.
- **Easy-to-fabricate and low-cost.** As summarized in Table III, the components are fabricated with minimal effort,

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

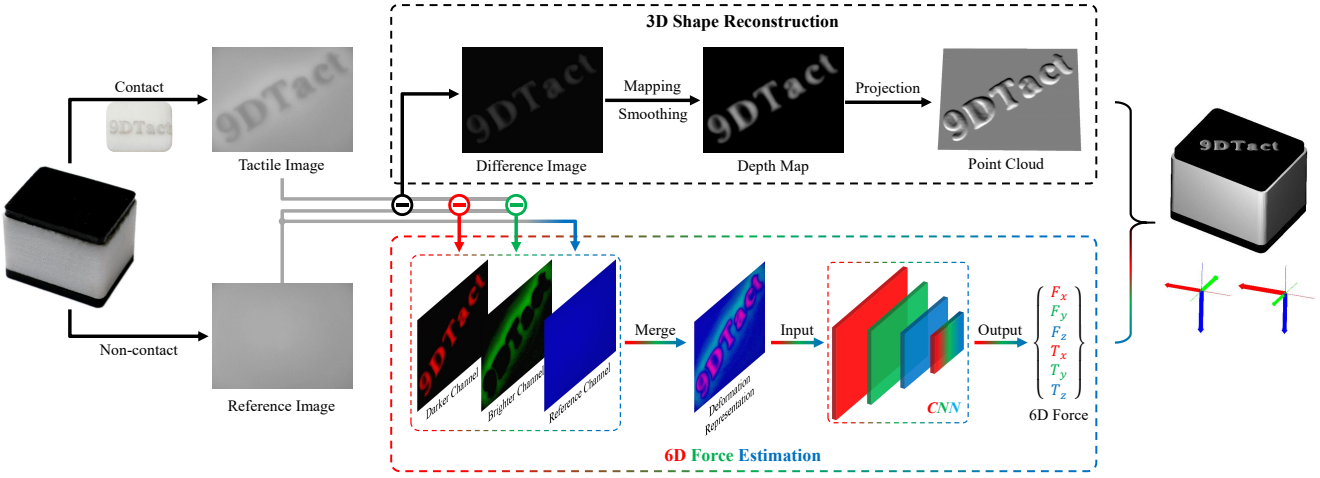


Figure 3: Pipeline for 9DTact's two key functions. 9DTact utilizes a modeling-based method for 3D shape reconstruction and a learning-based method for 6D force estimation. The reference image is captured when there is no contact on 9DTact, while the tactile image is captured when a badge with the text "9DTact" is pressed on 9DTact.

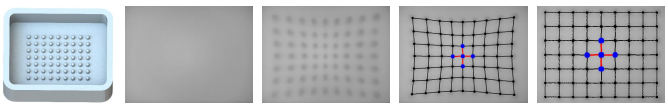


Figure 4: Camera calibration. (a) The calibration board features a cylinder array. (b) The reference image. (c) The tactile image is captured when pressing the calibration board on 9DTact. (d) Imprints from the calibration board are detected. (e) The result of rectifying and cropping the image in (d).

utilizing highly commercialized conventional machining processes such as 3D printing, laser cutting, and printed circuit board (PCB) soldering. Besides, each component is subject to strict mechanical positional constraints during assembly, thereby minimizing the need for additional adjustments and reducing performance differences in sensors caused by variations in assembly. To further support the replication of our 9DTact sensor, we also provide a step-by-step video tutorial that details the entire fabrication and assembly process, enabling researchers with varying levels of experience to easily reproduce the sensor on their own. Furthermore, building a 9DTact only costs about $15, which includes two reusable molds.

IV. 3D Shape Reconstruction

A. Reconstruction Pipeline

9DTact inherits the 3D shape reconstruction method of DTact [1], the pipeline of which is illustrated in Fig. 3. Both the tactile and reference images are converted to grayscale images because the reconstruction process only relies on pixel luminance. The difference image, calculated by subtracting the reference image from the tactile image, is mapped to a depth map with a calibrated mapping list. This mapping list is calibrated using the "single image" calibration approach proposed in [1], which is efficient for requiring only a single image for calibration. In addition, we apply two continuous Gaussian filters to denoise the depth map. Finally, the depth map is converted to point clouds to render and visualize the sensor surface.

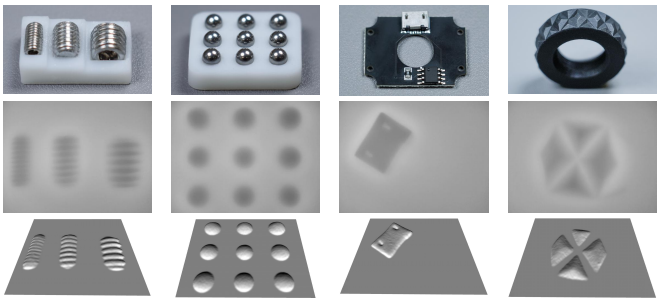


Figure 5: 3D shape reconstruction results of three set screws (M4, M6, and M8), a ball array, a micro USB port, and a wheel hub model. They are visual images, tactile images, and reconstructed point clouds from top row to bottom.

B. Camera Calibration

In order to obtain accurate tactile images, we need to rectify the optical distortion introduced by the wide-angle camera and the three gel layers. The marker-based image rectification method introduced by GelSlim3.0 [3] has the potential to correct distortion caused by multiple factors. However, markers are not painted on 9DTact for the reasons described in II-B. Therefore, a calibration board, as shown in Fig. 4 (a), is designed with a cylinder array and a frame that fits into the sensor shell of 9DTact. When the calibration board is pressed on the sensor surface, the cylinders imprint a grid array of virtual markers as shown in Fig. 4 (c). Fig. 4 (d) shows that markers are detected using algorithms from OpenCV [20].

In contrast to GelSlim3.0 [3] that regards the outermost points as anchor points, we choose the five relatively central points in blue because of their minimal aberration. The rectified positions of other points in the image frame can be computed by extending the five anchor points to equidistant grids. With the detected positions and rectified positions of all markers, we can compute the mapping array to rectify image. Finally, we crop the rectified image from 640×480 resolution to 460×345 as shown in Fig. 4 (e). In summary, the virtual markers-based image rectification method employed by 9DTact enables simultaneous calibration of these parameters:

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

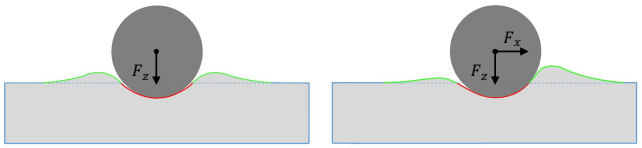


Figure 6: Deformable property of the translucent gel. The dotted line represents the original thickness of the gel. The areas that are thinner than the original thickness are colored in red, while the thicker areas are in green. (a) The object is pressed on the gel. (b) The object is dragged to the right after being pressed.

1) The lens distortion;
2) The pixel position on the tactile image that corresponds to the sensor surface's central position;
3) The physical length on the sensor surface corresponding to one pixel of the tactile image.

C. Improvements of 3D Shape Reconstruction

Although DTact [1] demonstrates excellent performance in shape reconstruction, it is limited to reconstructing only the central area of its surface due to uneven illumination and significant disturbance from environmental light in the peripheral area, as discussed in III-B. In contrast, 9DTact, through its refined design and advanced calibration technique, achieves comprehensive surface reconstruction with an incremental yet significant improvement in precision.

To validate its performance, 9DTact employs the same approach as proposed in [1]. This involves pressing a metal ball, different in radius from that used in the calibration phase, at various positions on the sensor surface to capture 20 distinct images. These images are then processed to compute actual depth maps, using circle detection algorithms in OpenCV [20]. The quantitative results of 9DTact's reconstruction precision yield a mean absolute error (MAE) of 0.0462mm and a standard deviation (Std) of 0.0304mm. While DTact records an MAE of 0.0476mm and an Std of 0.0352mm.

Fig. 5 showcases several reconstruction examples, highlighting the intricate geometric details captured by 9DTact.

V. 6D Force Estimation

A. Dense Deformation Representation

As mentioned in IV, the pixels within the contact areas become darker and are utilized to compute the 3D contact geometry. Surprisingly, we also observe a concurrent increase in the luminosity of pixels surrounding the contact areas. To elucidate this phenomenon, it is necessary to consider the flow properties of the gel.

As a hyper-elastic material, the pressed gel tends to flow outward to its neighboring regions, resulting in an increase in thickness in the surrounding areas, as shown in Fig. 6 (a). Similarly, when the contact object applies shear force or twist force, the surrounding gel will flow to accumulate along the moving direction, as shown in Fig. 6 (b). Furthermore, according to the principle that thinner contact areas result in darker pixels, the thicker surrounding regions induce brighter pixels. Eventually, as the visualization images illustrated in Fig. 7 with darker areas colored red and brighter areas green, the green pixels turn to be much brighter in the object's moving direction. Therefore, with the darker areas extracting the concave deformation information and the brighter areas extracting both the convex deformation and shear deformation information, deformation in all directions can be encoded in such dense deformation representation as the visualization images in Fig. 7 show.

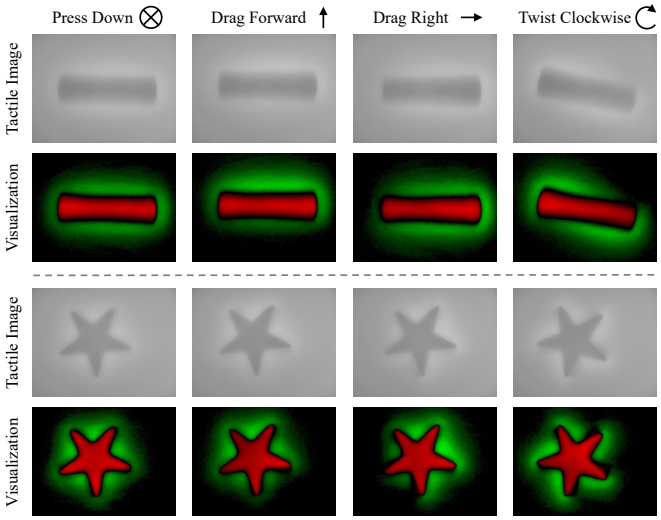


Figure 7: A cylinder and a star-shape object contact with 9DTact respectively. The object is firstly pressed onto 9DTact, and then dragged forward, dragged to the right, and twisted clockwise while maintaining contact. The tactile images are captured by the camera, and they are visualized as the visualization images with the darker areas in red and the brighter areas in green.

Our goal is to reconstruct the 6D force from the described dense deformation representation. It is unlikely to use numerical methods such as iFEM because the physical thickness of the bulging areas can not be acquired. Therefore, we leverage deep convolutional neural networks, which have shown great potential in image information extraction [21]. Following, we will introduce physical configurations to collect dataset, details on model training, and results of 6D force estimation.

B. Data Collection and Splitting

Previous tactile image datasets [17, 22, 23, 24] are mainly collected by mounting several objects on an autonomous machine, resulting in limited flexibility due to the inconvenience of object swapping, and thus, severely restricting the diversity of contact geometry. Furthermore, the process of force label collection dose not require recording the pose of the contact object. Therefore, we opt to manually press objects to increase the diversity of objects and the flexibility of pressing and swapping them.

We handpick 175 CAD models with various geometric shapes from the Thingi10K [25] dataset, and 3d-print them with black resin material, as shown in Fig. 8 (a). For the hardware setup, as Fig. 8 (b) shows, 9DTact is fastened on a BOTA MiniONE Pro 6-axis F/T sensor which provides precise 6D force labels. To efficiently collect data, we develop a program to autonomously sample image-force pairs so that we only need to press the objects on 9DTact in different

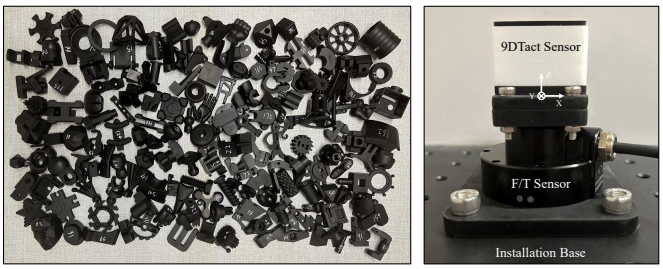
IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

Table IV: Configuration and hyper-parameters for force estimation deep model training.

Model	Batch size	Learning rate	Pretrained	Optimizer	Loss Function	Total epoch	GPU
Densenet-169	64	5×10^{-4}	No	Adam	MAE (Sum)	200	Nvidia A40

Table V: Validation results of 6D force estimation on two test sets (force in N, torque in Nm).

Splitting Method	Training set	Test set	Selected epoch	Mean absolute error (MAE)	Standard deviation (Std)
Standard	90417	10000	153	[0.30, 0.35, 0.28, 0.009, 0.008, 0.001]	[0.26, 0.30, 0.32, 0.009, 0.008, 0.001]
Object-based	89995	10422	189	[0.35, 0.40, 0.41, 0.011, 0.010, 0.002]	[0.31, 0.36, 0.44, 0.015, 0.014, 0.003]



(a) (b)

Figure 8: Physical configurations for dataset collection. (a) 175 objects with various geometric shapes are 3D-printed for collecting data. (b) The 9DTact sensor is installed on a BOTA MiniONE Pro 6-axis F/T sensor to collect 6D force labels.

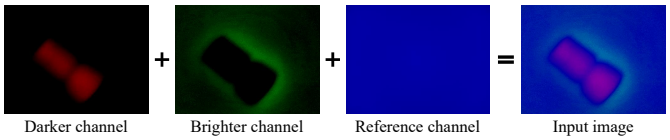


Figure 9: Input image for predicting 6D force. It consists of three channels: the darker channel, the brighter channel, and the reference channel.

orientations evenly. More specifically, a data pair is saved only when contact is detected and the 6d force is significantly different from all the forces saved during the same continuous contact period.

As introduced in Section V-A, 9DTact is able to extract pressing, dragging, and twisting motions of the contact geometry. Therefore, we not only press the objects but also drag and twist them on 9DTact to collect data with various contact status. Finally, we collect 100,417 image-force pairs totally from a single 9DTact sensor, taking about 10 hours cumulatively.

Leaving the remaining data as the training set, the test set is selected based on two splitting methods.

- **Standard splitting.** 10000 pairs of data are randomly selected from all data as test set. This splitting strategy is generally used to test the model's in-distribution generalization capability.
- **Object-based splitting.** 18 of the 175 objects are randomly selected as test objects, and all data sampling from these objects serves as test set, which consists of 10422 pairs of data in our case. This splitting strategy aims to test the model's ability to generalize to unseen objects.

C. Details of the Neural Network

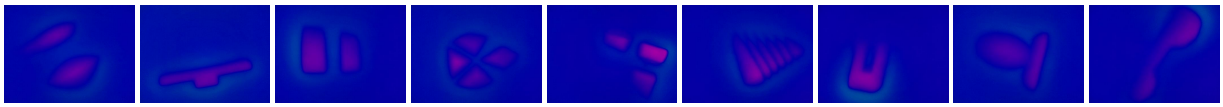
Input images. In order to provide sufficient and well-defined physical information for the neural network, the input image's three channels are replaced by the darker image, the brighter image, and the grayscale reference image, as shown in Fig. 9. The darker image, which contains information about the contact geometry, is generated by subtracting the tactile image from the reference image. It is the same as the difference image used for shape reconstruction as illustrated in Fig. 3. The brighter image reveals information about the dragging and twisting motions of the contact geometry, and it is obtained in a similar way as the darker image, but with the objects of subtraction reversed. The reference image is captured each time the sensor is initiated, which differs slightly for different continuous usage periods and thus provides information of the sensor's initial state.

Neural Networks. We select Densenet [26] as the neural network for predicting the 6D force. We utilize the implementation of Densenet from PyTorch library [27], and modify the output channel of the fully connected layer to 6. Two models are trained on datasets split using the two splitting methods, with the same training configuration depicted in Table IV. We also train Resnet [28] and ViT [29], but they perform worse than Densenet. Therefore, we only present the details and following results of Densenet, as our focus in this paper is to validate that our proposed dense deformation representation is able to provide comprehensive deformation information.

D. Results of 6D Force Estimation

As the quantitative validation results in Table V shows, the models are able to estimate accurate 6D force and generalize to both unseen contact status and objects. For standard validation, the absolute mean errors are $0.307N$ for forces and $0.006Nm$ for torques. For object-based validation, the errors are $0.370N$ and $0.0077Nm$ respectively. While these quantitative metrics are informative, direct comparisons of 9DTact with other sensors remains complex for several reasons. First, in contrast to other methods listed in Table II that validate their methods with a small set of simple objects, our test set comprises images with intricate geometries and in-plane motions as shown in Fig. 10. Second, different studies report varied metrics for normal and tangential force estimation such as an RMSE of $0.62N$ in [2] and an MAE of $0.41N$ in [17]. Lastly, inherent challenges in hardware comparisons arise due to factors such as resource limitations for sensor replication and performance variations attributed to the fabrication process.

Drawing from the comprehensive analysis of methodology, dataset, and results, our method for 6D force estimation



Input Image									
Force Error (N)	0.021	0.044	0.052	0.059	0.073	0.076	0.089	0.094	0.091
Torque Error (Nm)	0.0021	0.0052	0.0014	0.0029	0.0028	0.0012	0.0046	0.0008	0.0047

Figure 10: Results from the test set, using the standard splitting method, highlight our dataset's complexity in both geometry and motion dynamics. The average force and torque errors demonstrate the generalization capability of the trained deep model.

exhibits superiority in physical simplicity due to the integration of the optical and deformable properties of the translucent gel, in accuracy with the help of the dense information included in the proposed deformation representation, and in generalization capability by learning from the large and resourceful dataset.

VI. Conclusion

In this work, we present 9DTact, a general vision-based tactile sensor capable of 3D shape reconstruction and 6D force estimation. Specifically, we meticulously select and design each component of 9DTact to make it compact for installation, robust for illumination, durable for long-term usage, and simple for reproduction. We also improve 3D shape reconstruction to have a larger field ratio for reconstruction, more effective calibration procedures, and higher accuracy. Furthermore, unlike conventional methods for force estimation using painted markers, we extract a dense deformation representation from the raw tactile image by integrating the optical and deformable properties of the translucent gel. Finally, we train a force estimation neural network on a large dataset sampling from various objects with complex geometry. Empirical results show that it not only can estimate the 6D force accurately, but also can generalize to unseen geometries and objects.

References

[1] C. Lin, Z. Lin, S. Wang, and H. Xu, "Dtact: A vision-based tactile sensor that measures high-resolution 3d geometry directly from darkness," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 10 359–10 366.

[2] W. Yuan, S. Dong, and E. H. Adelson, "Gelsight: High-resolution robot tactile sensors for estimating geometry and force," *Sensors*, vol. 17, no. 12, p. 2762, 2017.

[3] I. H. Taylor, S. Dong, and A. Rodriguez, "Gelslim 3.0: High-resolution measurement of shape, force and slip in a compact tactile-sensing finger," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 10 781–10 787.

[4] M. Lambeta, P.-W. Chou, S. Tian, B. Yang, B. Maloon, V. R. Most, D. Stroud, R. Santos, A. Byagowi, G. Kammerer *et al.*, "Digit: A novel design for a low-cost compact high-resolution tactile sensor with application to in-hand manipulation," *IEEE Robotics and Automation Letters*, vol. 5, no. 3, pp. 3838–3845, 2020.

[5] GelSight. (2022) Gelsight-mini. [Online]. Available: https://www.gelsight.com/gelsightmini/

[6] 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*. IEEE, 2014, pp. 3988–3993.

[7] S. Dong, D. K. Jha, D. Romeres, S. Kim, D. Nikovski, and A. Rodriguez, "Tactile-rl for insertion: Generalization to objects of unknown geometry," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 6437–6443.

[8] Y. She, S. Wang, S. Dong, N. Sunil, A. Rodriguez, and E. Adelson, "Cable manipulation with a tactile-reactive gripper," *The International Journal of Robotics Research*, vol. 40, no. 12-14, pp. 1385–1401, 2021.

[9] A. Wilson, H. Jiang, W. Lian, and W. Yuan, "Cable routing and assembly using tactile-driven motion primitives," *arXiv preprint arXiv:2303.11765*, 2023.

[10] S. Dong, W. Yuan, and E. H. Adelson, "Improved gelsight tactile sensor for measuring geometry and slip," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 137–144.

[11] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[12] D. Ma, E. Donlon, S. Dong, and A. Rodriguez, "Dense tactile force estimation using gelslim and inverse fem," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 5418–5424.

[13] K.-J. Bathe, *Finite element procedures*. Klaus-Jurgen Bathe, 2006.

[14] G. Zhang, Y. Du, H. Yu, and M. Y. Wang, "Deltact: A vision-based tactile sensor using a dense color pattern," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 10 778–10 785, 2022.

[15] Y. Zhang, Z. Kan, Y. Yang, Y. A. Tse, and M. Y. Wang, "Effective estimation of contact force and torque for vision-based tactile sensors with helmholtz–hodge decomposition," *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 4094–4101, 2019.

[16] H. Bhatia, V. Pascucci, and P.-T. Bremer, "The natural helmholtz-hodge decomposition for open-boundary flow analysis," *IEEE transactions on visualization and computer graphics*, vol. 20, no. 11, pp. 1566–1578, 2014.

[17] W. K. Do, B. Jurewicz, and M. Kennedy, "Densetact 2.0: Optical tactile sensor for shape and force reconstruction," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 12 549–12 555.

[18] 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*. IEEE, 2009, pp. 1070–1077.

[19] S. Wang, Y. She, B. Romero, and E. Adelson, "Gelsight wedge: Measuring high-resolution 3d contact geometry with a compact robot finger," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 6468–6475.

[20] G. Bradski, "The opencv library." *Dr. Dobb's Journal: Software Tools for the Professional Programmer*, vol. 25, no. 11, pp. 120–123, 2000.

[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, 2012.

[22] D. F. Gomes, P. Paoletti, and S. Luo, "Generation of gelsight tactile images for sim2real learning," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 4177–4184, 2021.

[23] A. Church, J. Lloyd, N. F. Lepora *et al.*, "Tactile sim-to-real policy transfer via real-to-sim image translation," in *Conference on Robot Learning*. PMLR, 2022, pp. 1645–1654.

[24] W. Chen, Y. Xu, Z. Chen, P. Zeng, R. Dang, R. Chen, and J. Xu, "Bidirectional sim-to-real transfer for gelsight tactile sensors with cyclegan," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 6187–6194, 2022.

[25] Q. Zhou and A. Jacobson, "Thingi10k: A dataset of 10,000 3d-printing models," *arXiv preprint arXiv:1605.04797*, 2016.

[26] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[27] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, vol. 32, 2019.

[28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[29] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.