2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

TranTac: Leveraging Transient Tactile Signals for Contact-Rich Robotic Manipulation

Yinghao Wu[1], Shuhong Hou[1], Haowen Zheng[1], Yichen Li[1], Weiyi Lu[1], Xun Zhou[1], Yitian Shao[1,2]

***Abstract*— Robotic manipulation tasks such as inserting a key into a lock or plugging a USB device into a port can fail when visual perception is insufficient to detect misalignment. In these situations, touch sensing is crucial for the robot to monitor the task's states and make precise, timely adjustments. Current touch sensing solutions are either insensitive to detect subtle changes or demand excessive sensor data. Here, we introduce TranTac, a data-efficient and low-cost tactile sensing and control framework that integrates a single contact-sensitive 6-axis inertial measurement unit within the elastomeric tips of a robotic gripper for completing fine insertion tasks. Our customized sensing system can detect dynamic translational and torsional deformations at the micrometer scale, enabling the tracking of visually imperceptible pose changes of the grasped object. By leveraging transformer-based encoders and diffusion policy, TranTac can imitate human insertion behaviors using transient tactile cues detected at the gripper's tip during insertion processes. These cues enable the robot to dynamically control and correct the 6-DoF pose of the grasped object. When combined with vision, TranTac achieves an average success rate of 79% on object grasping and insertion tasks, outperforming both vision-only policy and the one augmented with end-effector 6D force/torque sensing. Additionally, TranTac's contact localization performance is validated through tactile-only insertion tasks, where the inserted object and slot are initially misaligned by 1 to 3 mm, achieving an average success rate of 88%. We assess the generalizability by training TranTac on a single prism-slot pair and testing it on unseen data, including a USB plug and a metal key, and find that the insertion tasks can still be completed with an average success rate of nearly 70%. The proposed framework may inspire new robotic tactile sensing systems for delicate manipulation tasks. Project page: `https://wusdream.github.io/TranTac/`**

I. INTRODUCTION

Recent advancements in vision-language-action models have sparked rapid development of robotic technologies, allowing robots to learn various manipulation skills, such as rearranging tools and delivering objects [3], [45]. However, delicate manipulation remains challenging because vision alone often fails to detect transient, subtle changes in objects. In addition, visual sensing can be obstructed when handling small objects or operating in complex environments. In these situations, touch sensing is highly desirable.

Many tactile sensing strategies have been proposed [23], with visuo-tactile sensing emerging as one of the most popular choices due to its high spatial resolution and exceptional reliability [43], [20], [33]. However, because these sensors prioritize the capture of detailed spatial information, they suffer from limited responsiveness due to their relatively low sampling rate. This prevents robots from reacting quickly in dynamically changing, contact-rich environments. To address this, some designs have explored the use of contact microphones [8], [22], [37], [29], [25] or event cameras [31], [19], [39], [10] to detect transient touch events. However, event cameras are costly, while contact microphones lack directional information.

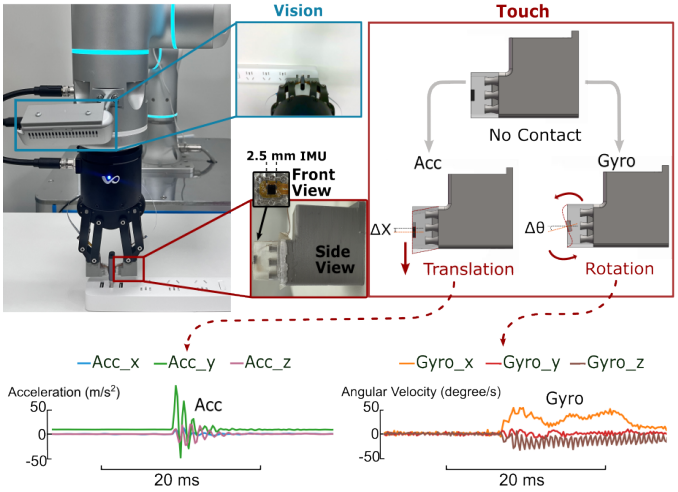

Fig. 1: Overview of TranTac, a tactile sensing and control framework for fine robotic manipulation that integrates visual input from a wrist-mounted camera with high-frequency tactile feedback from fingertip-embedded inertial measurement units (IMUs). When the grasped objects interacts with the environment, the embedded IMUs in the gripper's tip capture the subtle translation and torsion of the elastomeric tip, which produces 3-axis acceleration (ACC) and angular velocity (Gyro) signals. These signals provide fine-grained feedback for precise 6-DoF control of the robot end-effector.

In this paper, we draw inspiration from the dexterity of the human hand, which utilizes dynamic tactile sensing to efficiently track the pose of grasped objects [14]. In particular, a light touch contact on the human fingertip can be readily captured by the tactile sensors embedded in the skin, efficiently encoding contact information through sparse temporal neural coding [32]. Such encoding mechanism is characterized by high temporal fidelity, informing the importance of a wide bandwidth tactile sensing system. Here, our approach focuses on the design of gripper tips, where contact interactions most frequently occur during manipulation.

This work was supported by the Shenzhen Science and Technology Innovation Program under Grant 20240148 and National Natural Science Foundation of China under Grant 62576119.

[1]School of Computer Science and Technology, Harbin Institute of Technology, Shenzhen, Shenzhen, China (correspondence: shaoyitian@hit.edu.cn).

[2]State Key Laboratory of Smart Farm Technologies and Systems, Harbin, China.

TABLE I: Comparison of TranTac (ours) with existing tactile sensing methods previously applied to object insertion tasks. Performance indicators are marked to show whether ↓ low values are better or ↑ high values are better, with the best value highlighted in bold.

Sensor	TranTac (Ours)	GelSlim 3.0 [35]	9DTact [24]	DIGIT [21]	ReSkin/AnySkin [1]	3D-ViTac [13]
↓ Size [mm]	**11×11×8**	37×80×20	33×26×26	20×27×18	20×20×2	48×48×3
↓ Cost [$]	**5**	25	6	15	30	20
↓ Channels/Pixels	**6**	640×480	640×480	640×480	15	16×16×4
↓ Data Efficiency (Stream Volume) [KB/s]	42	27648	27648	55296	**12**	33
↑ Sensing Bandwidth [Hz]	**3500**	30	30	60	400	32.2

Given the need for timely tracking and control of the 6-DoF pose of a grasped object, we consider that a sensor capable of detecting both translational and torsional deformations of the gripper is sufficient. We designed TranTac, a novel robotic sensing and control framework that leverages contact-sensitive, wide-bandwidth, and low-cost 6-axis IMU sensors to capture subtle and transient deformations of the gripper tip induced by pose variations of the grasped object. Compared to existing methods, our hardware design is both data-efficient and cost-effective, while retaining the capability to capture dynamic tactile information with high temporal accuracy (Table I). It learns a visuotactile policy through diffusion of action, enabling fine manipulation. Temporal features are extracted from instantaneous IMU data, encoded via transformers, fused with vision features, and fed into a diffusion model to generate 6-DoF poses for subsequent movement steps. We validate the effectiveness of TranTac through physical experiments, in which the robot grasps and inserts objects of various shapes and materials.

The key contributions of this paper are as follows: 1) A novel tactile sensing mechanism for robots that leverages temporal tactile information to efficiently detect transient, subtle pose changes in grasped objects. 2) A compact and readily reproducible design of the tactile sensing hardware that utilizes low-cost, off-the-shelf IMU chip and silicone molding. 3) A robot imitation learning framework based on action diffusion, which integrates IMU-based tactile data with spatial visual information to perform insertion tasks using a diverse set of objects—from plastic prisms to everyday items like USB plugs and metal keys.

II. Related Work

Tactile sensing for robot manipulation: Various tactile sensing designs have been developed for robot grippers. While force/torque sensors are common in industrial applications [4], their bulkiness often precludes installation at the gripper tip, limiting their effectiveness for localized tactile sensing. Tactile sensors that utilize piezoresistive, piezoelectric, or capacitive components can be embedded in robot skin with slim profiles [27], [9], [13], [16]. However, these sensors often require complex fabrication processes and acquisition electronics, and they are susceptible to environmental noise and crosstalk. Visuo-tactile sensors [43], [33] have gained increasing attention due to their excellent spatial resolution and reliability. The deformation of a soft robot gripper tip can be captured by an embedded camera, capturing high-precision data on the contact location and surface texture of the grasped object. However, the camera and required optical path space enlarge the end effector equipped with such sensors. Moreover, the temporal resolution of visuo-tactile sensing is constrained by the camera frame rate, making it difficult to capture high-frequency transient tactile signals at the kilohertz level. Acoustic sensors such as piezoelectric contact microphones have been integrated into robotic grippers or affixed to experimental objects to compensate for the response delay of vision-based tactile sensors in detecting contact events, modes and object states [8], [22], [37], [29], [25]. However, bulky contact microphones are hard to be integrated into gripper tips, and their signals lack directional information and are highly susceptible to ambient acoustic noise [25].

Peg-in-hole insertion via touch: Peg-in-hole problems are classic but challenging benchmarks for robotic manipulation. Early works relied primarily on visual sensing for object localization and alignment, but vision alone often struggles with occlusion and limited accuracy in fine insertion tasks [6], [15]. Recent research has increasingly focused on the integration of visuo-tactile sensors to provide high-resolution local contact information for precise insertion tasks. A constraint-based estimation framework has been developed to localize extrinsic contacts using distributed tactile sensing [28]. Streaming tactile imprints can be leveraged to estimate and correct object pose errors, enabling robust insertion of unknown objects without prior geometric knowledge [7]. Tactile sensing can also support closed-loop control in complex tool-using manipulation tasks by providing accurate real-time pose estimation [34]. Learning-based approaches have been developed to recognize contact states using force and torque data, enabling robots to perform assembly tasks through adaptive impedance control strategies [42]. More recently, diffusion models have been applied to generate force-domain policies for high-precision tactile insertion, achieving zero-shot transfer across novel tasks while addressing the frequency misalignment between policy inference and real-time control [40]. Visuo-tactile sensing typically relies on tactile image sequences, where object pose and extrinsic contact locations are implicitly encoded. This approach demands extensive spatiotemporal data, often leading to delayed pose adjustments. To mitigate this and better capture transient cues, active exploration has been integrated to improve contact location estimation [18].

Imitation learning for robot manipulation: Imitation learning enables robots to acquire complex manipulation skills through demonstrations. Recent approaches employ

deep-generative models to preserve the variability of imitated trajectories, allowing for generalization to new scenarios [17], [36], [2]. End-to-end imitation learning has been shown to enable low-cost hardware to perform fine-grained bimanual manipulation tasks based on real-world demonstrations [44]. Diffusion-based generative models have demonstrated potential in visuomotor policy learning, enabling robots to generate diverse and adaptive motion plans [5].

III. Bio-inspired Dynamic Sensing for Subtle Touch

Humans can perform blind insertion tasks by perceiving subtle contact between a grasped object and its surroundings. Tactile sensors in the hand play a pivotal role, as they are extremely sensitive to transient skin deformations [14] and can efficiently capture tactile information through sparse but highly precise temporal encoding [38], [32]. Inspired by this, we designed our TranTac sensing system using a single 6-axis IMU with a temporal resolution as high as 0.28 ms.

This design aims to responsively and efficiently capture the 3-axis translation and 3-axis torsional deformation that occur in the contact region of the gripper tip (Fig. 1). The sensing efficiency of our design can be compared to emerging visuo-tactile sensing techniques, where the status of grasped object is encoded by high-dimensional spatio-temporal tactile signals recorded by an integrated camera. For example, a 30 fps, 480p camera produces at least 9000 kilobytes of data per second ($30\times640\times480$), not to mention that many existing studies utilize cameras with even higher spatial and temporal resolutions. In contrast, our design features a data volume of just 42 kilobytes per second. The detailed comparison between different sensors is provided in Table I.

A. Gripper Tip with 6D Dynamic Tactile Sensing

TranTac employs a soft elastomeric fingertip with an embedded 6-axis IMU (ST LSM6DSR iNEMO) to capture subtle translational and torsional deformations of the elastomeric tips. The size of the IMU chip is only 2.5×3 mm, compact enough to be integrated into gripper tips. The sensor measures accelerations up to ±16 g and angular velocities up to ±4000 dps. IMU data are streamed via a flat flex cable at 3500 Hz to a compact computing unit (Raspberry Pi 5), which transmits the data in real-time to a circular buffer on the PC for storage and processing by the action generation module.

The elastomeric tip is fabricated from polydimethylsiloxane (PDMS, Dow SYLGARD 184) mixed at the standard 10:1 base-to-curing-agent ratio and is molded onto the protrusion of the 3D-printed fingertip support (Fig. 2). First, the IMU board is placed in the mold's bottom slot, filled with PDMS, and joined with tip base; the assembly is then cured at 60 °C for 120 minutes, cooled, and demolded. Next, a second mold is fitted over the tip to cast an additional thin PDMS layer over the exposed IMU. Following another curing cycle at 60 °C for 120 minutes, the mold is removed to complete the sensor.

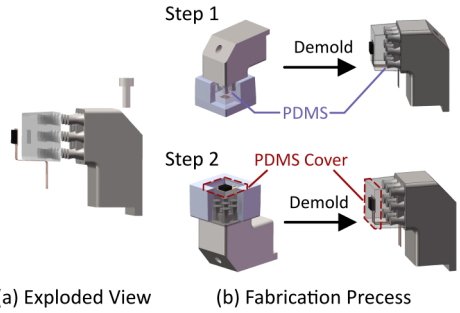

Fig. 2: Design of TranTac tactile sensor: a) exploded view and b) fabrication process.

B. Tactile Sensing for Contact Localization

Fig. 3 shows four scenarios where an object grasped by the parallel gripper moves vertically downward to contact an edge in the environment and subsequently lifts off. In each scenario, we visualize the 20 ms collision signals, which capture the immediate response at contact.

As demonstrated in the figure, different collision directions produce distinct signal patterns. In the first row, the gripper contacts an edge that is parallel to the sensor plane, resulting in a difference in the magnitude of the left and right acceleration signals along the y-axis and relatively large changes in the gyroscope (Gyro) x-axis signals. In the second row, the gripper contacts an edge perpendicular to the sensor plane, causing the object to rotate within the sensor plane and leading to a significant change in the gyroscope signals along the z-axis.

C. Policy Learning

Human hands can adjust the pose during tool use by perceiving dynamic translational and torsional deformations

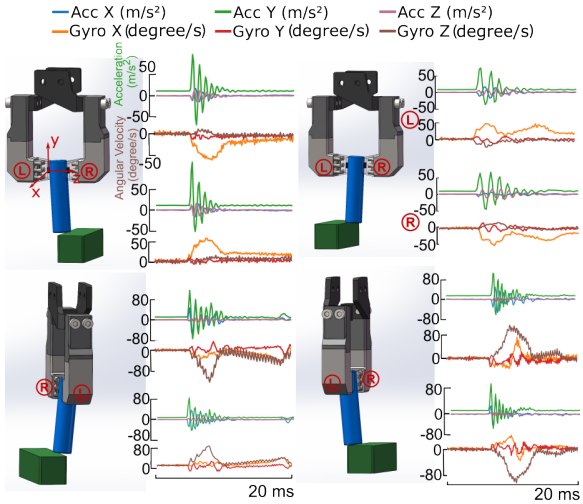

Fig. 3: Examples of measured TranTac tactile signals for four contact directions. The top two cases represent edge contact along the gripper's x-axis, while the bottom two indicate contact along the z-axis. The contact information is encoded via magnitude differences in the accelerometer (Acc) signals and phase shifts in the gyroscope (Gyro) signals.

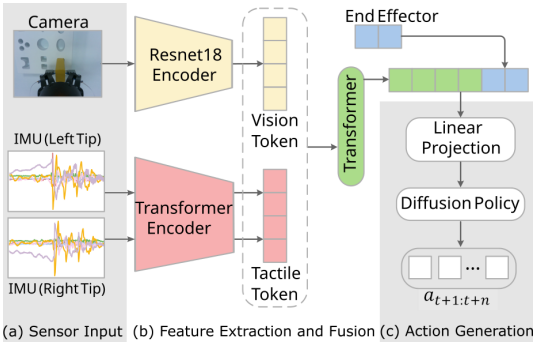

Fig. 4: Network architecture of the proposed model. It consists of three parts: a) Sensor input, which contains camera images and two IMU tip signals. b) Feature extraction and fusion, including the concatenation of vision and tactile tokens and a Transformer for multi-modal and temporal feature integration together with the current end-effector state. c) A conditional diffusion strategy to generate robot end-effector poses.

of the skin at fingertips [14]. To enable robots to imitate this human capability, we propose an end-to-end visuotactile policy that maps visual and tactile observations to actions, as illustrated in Fig. 4. Our method consists of three critical parts. Fig. 4(a) shows the sensor inputs, including one wrist-mounted camera and two 6-axis IMU sensors at the fingertips. The camera sampling frequency is 24 Hz, and the IMU sampling rate is 3500 Hz. Fig. 4(b) illustrates the feature extraction and fusion process from the image and imu signals, and Fig. 4(c) depicts the policy learning process for action generation, which leverages the diffusion policy [5] conditioned on our tactile representation to generate 6-DoF actions for the robot end-effector.

1) Visuotactile Encoding: We use camera images as visual input and adopt ResNet-18 [11] as the visual encoder. Within each visual frame interval, the 146-frame sequence of 6-axis IMU signals is used as input to the tactile module. Specifically, the IMU signals are projected to 64 dimensions through a single MLP layer, then processed by a transformer to extract features. We use the same IMU encoder to extract left tip tokens and right tip tokens, which are then concatenated and fused with visual tokens using a transformer. Finally, the fused tokens are concatenated and projected to 512 dimensions, then concatenated with the robot proprioception features to serve as input to the action head. Note that we use the last two time steps' sensor signals as observations.

2) Action Generation: Our action head is a diffusion policy D_{policy} [5]. As illustrated in Fig. 4, the policy network learns to denoise actions by predicting ε_θ from multimodal visual, tactile, and proprioceptive observations O. We employ the conventional DDPM [12] loss function:

$$L = \mathbb{E}_{(O,A_0)\in D_{policy},k,\varepsilon_k} \left\| \varepsilon_k - \varepsilon_\theta(O, A_0 + \varepsilon_k, k) \right\|^2, \quad (1)$$

where k represents the noise schedule step, ε_k is the ground-truth noise, and A_0 represents the ground-truth 16-step future robot trajectories. We predict the noise ε_θ through a CNN-based diffusion model with FiLM conditioning [30].

IV. Experiments

In this section, we investigate extensive contact-rich insertion experiments to demonstrate the effectiveness and generalization of using tactile cues from the insertion process to help policy learning. These experiments are designed to answer the following questions:

- Does TranTac improve policy performance in contact-rich tasks compared to both pure vision-based policies and robot end-effector 6D force sensor? (Experiment 1)
- Can TranTac correctly identify the contact location and the direction of adjustment? (Experiment 2)
- Can tactile cues captured by TranTac generalize to different objects? (Experiment 2)

A. Environment setup

The hardware used in the experiment is shown in Fig. 1, including a 7-DoF robot arm (Flexiv Rizon 4s) integrated with a 6D force/torque sensor and equipped with a gripper (Flexiv Grav), two TranTac sensors mounted on the gripper tips, and a RGB-D camera (RealSense D435i) mounted on the robot arm for wrist view. For each task, demonstrations are collected at 24 Hz via a teleoperation system employing a see-through VR headset [41]. Learned policies are trained for 200 epochs with a batch size of 32 using the AdamW optimizer, then deployed at a 12 Hz frequency. We use action chunking with exponential temporal averaging [44] to produce smoother behavior.

B. Experiment 1: Policy Comparison

1) Task Description: Fig. 5 shows the experimental objects used in the insertion experiments. We use 3D printable objects from [26], including a rectangle and a circle-square prism with insertion slot clearances of only 0.5 mm. Additionally, we test with two everyday objects, a USB connector-hub pair and a key-lock pair.

Plastic prism insertion tasks: This task requires the robot to insert a 3D-printed object—either a simple rectangular shape or a composite form consisting of a circle and a square—into its designated slot. The arm starts at a random position near the target slot and places the object directly below the gripper. The training dataset consists of 40 demonstration trajectories.

USB insertion task: This task requires the robot to plug a USB connector into a specific port on a power hub. The arm starts at a random position near the power hub and places the object directly below the gripper. The training dataset consists of 40 demonstrations.

Key insertion task: This task requires the robot to insert a metal key into its corresponding lock. The robotic arm begins from a random position to the left of the lock, with the key positioned directly beneath the gripper. The training dataset comprises 40 demonstrations.

Note that during testing, we randomly initialize the gripper location using the same distribution as in training to prevent the network from memorizing the initial robot pose, while maintaining the same relative position between the object and the slot.

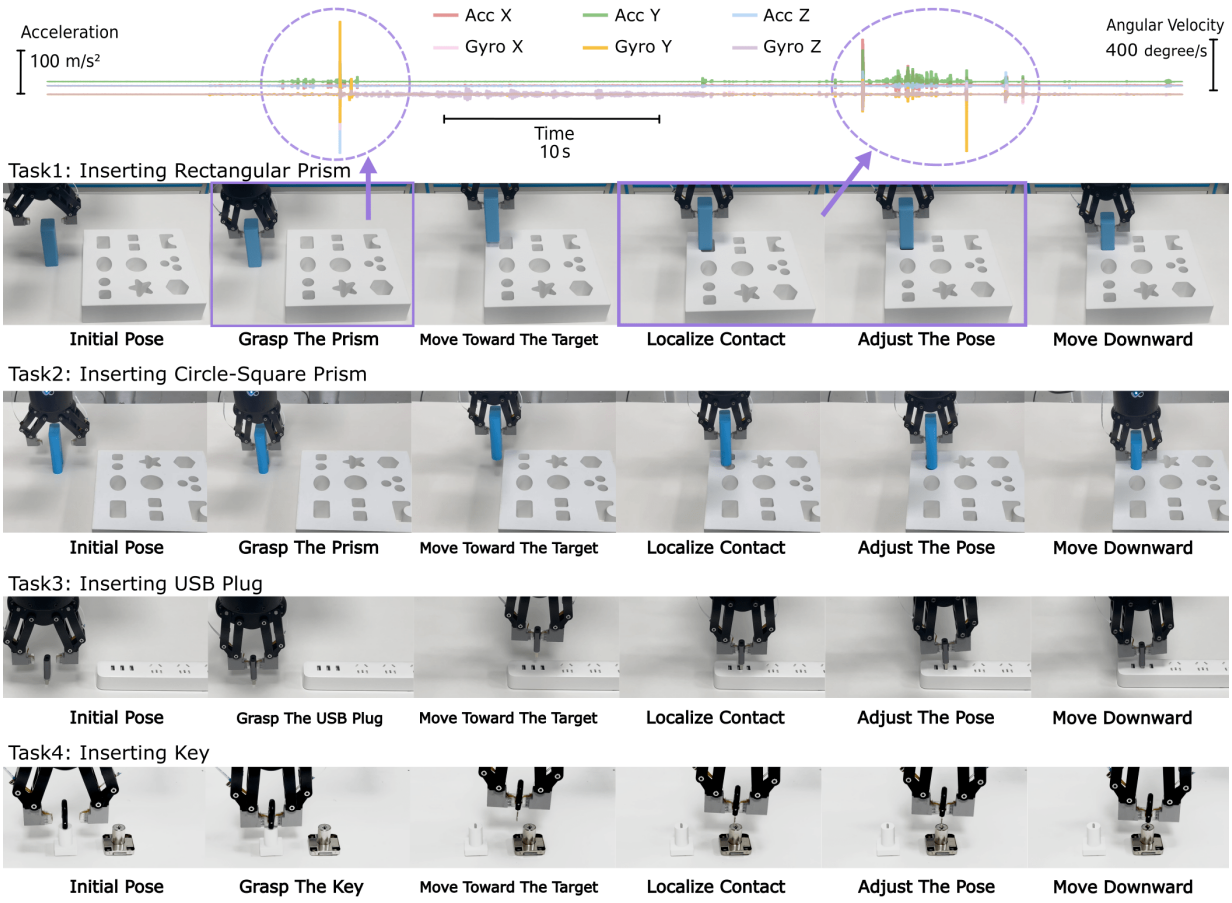

Fig. 5: Experimental objects used in the insertion experiments, including 3D printed plastic prisms and insertion board, USB connector-hub pair, and key-lock pair. The top panel displays the transient tactile signals captured by TranTac, highlighting key events during the grasping and insertion process.

Based on the four tasks above, we evaluated and compared the following three policies. Each policy was tested using 20 rollouts per task to ensure reliable assessment.

- **Vision Only:** ResNet-18 encoder extracts 512-dimensional features from RGB inputs.
- **Vision with Force:** 6D force/torque data is projected to 512 dimensions via MLP and fused with ResNet-18 visual features using Transformer attention.
- **Vision with TranTac:** Tactile inputs (TranTac) are processed through a 4-layer Transformer encoder with 8 attention heads, where the `[CLS]` token is used to extract temporal features and projected into a 512 dimensions and fused with ResNet-18 visual features using Transformer attention. No force/torque data is used.

2) Qualitative Analysis: Table II shows the experimental success rates. Our results show that TranTac policies outperform both vision-only approaches and vision-based methods augmented with 6D force/torque sensing. Additionally, we observe that TranTac policies can fit the demonstration datasets more precisely, particularly for contact strategies at hole edges, enabling accurate contact localization and adjustment for more precise insertion operations. In contrast, force-based policies suffer from low signal-to-noise ratios in the robot's end-effector 6D force/torque measurements due to their large sensing range, which adversely affects performance in fine manipulation tasks. The vision-only policy struggles with occlusion issues, especially in circle-square insertion tasks where the inserted object blocks the view of the square hole, frequently causing the robot to get stuck near the hole entrance and resulting in lower success rates for pure vision approaches.

TABLE II: Success rates (out of 20) of four insertion tasks for comparing three policies.

Policy	Rectangle Insertion	Circle-square Insertion	USB Insertion	Key Insertion
Vision Only	75%	60%	30%	80%
Vision w. Force	50%	70%	40%	40%
Vision w. TranTac	**80%**	**80%**	**65%**	**90%**

C. Experiment 2: Generalizability of Tactile-Only Policy

To validate TranTac's contact localization capability and the generalizability of tactile features, we designed tactile-only policy experiments.

1) Task Description: Fig. 6 shows the training and testing objects used in the insertion experiments. Six 3D-printed plastic objects in the shapes of prism, cylinder, and elliptical cylinder are used, each with lengths of 40 mm and 60 mm.

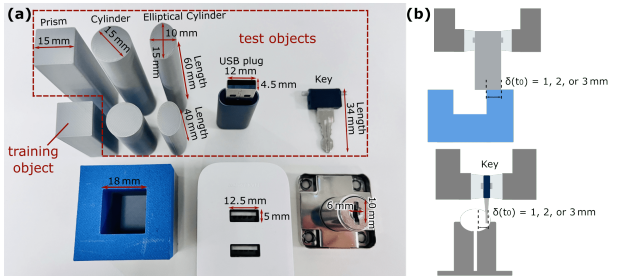

Fig. 6: The experiment involves inserting objects into a rectangular slot that is 3 mm wider than the objects. A 40 mm prism was used as the training object, while the remaining ones are testing objects. The tasks include inserting 40 mm and 60 mm long prism, round and elliptical cylinders into the rectangular slot, inserting a USB plug into a hub, and inserting a metal key into a keyhole.

In addition, the testing is generalized to everyday objects, including a USB connector-hub pair and a key-lock pair.

The robot is trained to perform the insertion tasks imitating human behaviors through four distinct stages (Fig. 7).

Move vertically downward (Stage 1): The gripper holding the object is positioned above the slot and moved vertically downward until contact is detected. This stage demonstrates the response speed of TranTac in contact detection.

Localize contact and slide along the opening (Stage 2): The robot localizes external contact by sensing subtle translational and torsional deformations within the fingertip, allowing it to determine the location of the slot opening. As the deviation level increases, localizing the contact becomes more challenging due to the reduction of torsional deformations and translational differences between the two fingertips. The robot then slides the object along the opening using the TranTac sensors, and upon detecting a signal indicating departure from the edge, it incorporates a downward movement into its trajectory. Admittance control is applied to the robotic arm to ensure that contact forces remain within a safe threshold of 5 N, using a proportional integral controller.

Detect contact with inner wall (Stage 3): TranTac can responsively track the dynamic 6-DoF poses, enabling it to promptly identify subtle contact with the inner wall and distinguish it from external contact outside the slot, as is common during stage 2. This stage demonstrates its ability to sense different contact modes.

Resume downward movement (Stage 4): This stage is relatively straightforward for the plastic prism-slot pair because the slot is 3 mm wider than the plastic prism, allowing for greater angular misalignment during insertion. However, in real-world peg-in-hole insertion tasks, such as USB plug or key insertion, the clearance between the peg and the hole is much smaller, as shown in Fig. 6. Therefore, it is necessary for TranTac to sense the direction of obstruction and make appropriate adjustments.

To successfully accomplish this task, the robot must be able to perceive various contact modes and directions and take appropriate actions accordingly. We collect 40 demonstrations in total, with 10 demonstrations for each edge of the slot. Note that only the 40 mm rectangular prism-slot pair is used for training, while the others are reserved for testing.

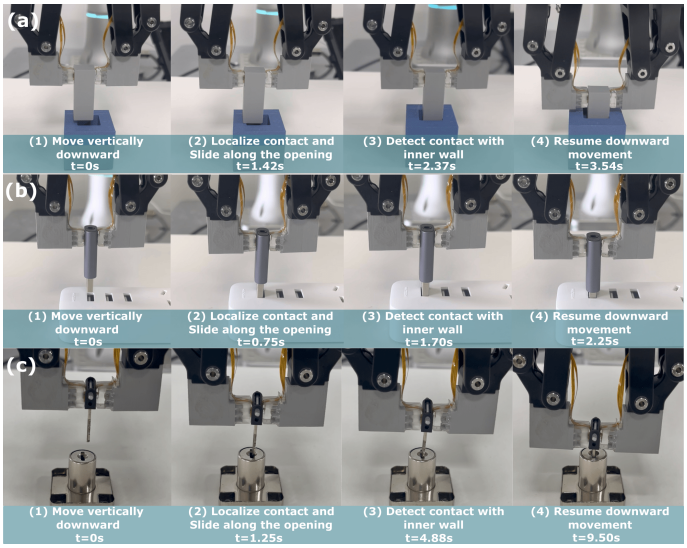

Fig. 7: Physical tests with three types of insertion tasks without visual sensing. The initial position of the objects is randomized. Four insertion stages are shown.

2) Evaluation Metrics: To ensure a fair comparison, the inserted object is initially offset from the target slot by 1 to 3 mm in each of the four lateral directions. For each deviation level, four lateral directions were tested and two rollouts were conducted per direction, resulting in a total of 24 rollouts per object. The task is considered successful if the object is adjusted towards the slot opening and successfully inserted into the slot.

3) Qualitative Analysis: The quantitative results are presented in Table III, showing the success rates for inserting 3D printed objects, a USB plug and a metal key at different deviation levels. The 40 mm rectangular prism was the only object used for training, achieving an average success rate of 88% during testing. All other objects, used for generalization evaluation, still achieved an average success rate of nearly 70% in insertion tasks. Our experimental results reveal

TABLE III: Success rates (out of 24) of insertion tasks using TranTac-only policy. Only the 40 mm rectangular prism was used for training; the rest are unseen objects for generalization evaluation. Each object was tested under three deviation levels, represented by $\delta(t_0)$, which is the corrective movement distance required to complete an insertion task.

Testing Objects	$\delta(t_0)$ = 1 mm	$\delta(t_0)$ = 2 mm	$\delta(t_0)$ = 3 mm	Avg. Success Rate
4 cm rectangular prism	100%	87.5%	75.0%	87.5%
4 cm cylinder	75.0%	62.5%	62.5%	66.7%
4 cm elliptical cylinder	87.5%	62.5%	50.0%	66.7%
6 cm rectangular prism	87.5%	75.0%	50.0%	70.8%
6 cm cylinder	75.0%	75.0%	50.0%	66.7%
6 cm elliptical cylinder	75.0%	62.5%	25.0%	54.2%
USB Plug	75.0%	62.5%	50.0%	62.5%
Metal Key	87.5%	87.5%	87.5%	87.5%

several key insights. As the starting deviation distance $\delta(t_0)$ increases, the success rate decreases. This is because a larger deviation $\delta(t_0)$ causes the contact point to be closer to the rotational center, resulting in a shorter moment arm and thus a smaller torque. Consequently, the dynamic torsional deformations at the edge contact are reduced, leading to a weaker gyroscope signal and making direction discrimination more difficult. As shown in Fig. 6, the narrow edges of the elliptical cylinder and the USB plug are only 10 mm and 4.5 mm wide, respectively, which can lead to incorrect direction inference after contact. Given the minimal object dimensions, a 3 mm deviation already places the contact point near the grasp center of mass, resulting in negligible rotational and translational disturbances. Consequently, larger perturbations become feasible when grasping objects of increased size. Furthermore, the visuotactile policy experiments demonstrate that vision provides effective coarse localization (typically within 1–2 mm), after which tactile feedback enables precise fine-tuning through subtle corrective adjustments. Collisions between objects of different shapes and the slot produce different contact modes, which in turn lead to varying dynamic translational and torsional deformations of the elastomeric tips, diminishing the success rate for inserting generalized objects. Experimental results with a USB plug and a metal key demonstrate that TranTac can be generalized to objects of various materials and geometries, as its sensing mechanism is based exclusively on directional deformation of the elastomeric tips.

V. Limitations

Although TranTac demonstrates its ability to leverage transient dynamic tactile information for completing contact-rich insertion tasks, several limitations remain.

Limited sensing for spatial information: Current design of TranTac prioritizes reducing sensing channels and maximizing the temporal resolution, but lacks the sensing capability of spatial patterns. One possible future direction is to increase the number of IMU integrated and to adapt super-resolution techniques. Existing research suggests that spatial information is partially encoded in the temporal structure of vibrotactile signals and thereby can be extracted via decoding methods [32].

Limited sensing for pseudo-static contact: Since TranTac is centered on dynamic tactile sensing, it lacks the ability to measure constant and pseudo-static deformation. Thus, if it touches an object that slowly deforms the gripper tip, TranTac can barely detect any static pressure. As a result, integrating other types of sensing components, such as a magnetometer and corresponding magnetized elastomeric materials, is one direction to be explored in the future.

Incomplete physical modeling: The physical model describing the relationship between the 6D pose of the grasped object and the corresponding 6-axis IMU signals remains underexplored. Further research is needed to establish an explainable model for more robust pose estimation.

Sensor size optimization: The current design of elastomeric gripper tip is four times larger than the size of the IMU chip, suggesting that the size of the elastomeric tip can be further reduced with a more optimized mechanical design. A more compact gripper tip will allow the robot to manipulate objects of smaller sizes.

Limited policy inference speed: Although we have high sampling rate continuous tactile signal input, the iterative denoising process of the diffusion policy requires approximately 60 ms per inference, which limits the robot's real-time feedback capabilities. Further research is needed to explore algorithms that balance both inference speed and the ability to model multimodal action distributions.

VI. Conclusion

We present TranTac, a cost-effective tactile sensing and control framework designed to enable delicate robotic manipulation. We utilize a compact elastomeric gripper tip integrated with IMU to capture transient and subtle tactile signals arising from dynamic translational and torsional deformations of the tip when the grasped object has environmental contact. To infer extrinsic contact direction, tactile signals are encoded and fused with visual features, then fed into an action diffusion model. This establishes a sensing-to-control mapping between the 6-DoF deformation of the elastomeric gripper tips and the corresponding 6-DoF pose adjustments of the robot end-effector during insertion tasks Our experiments demonstrate that TranTac, when used with a visuotactile policy, achieves a high average success rate of 79% in grasping and insertion tasks. For downward insertion tasks specifically, it reaches 88% success on the training object and nearly 70% on average for unseen objects of varying shapes, sizes, and materials using the tactile-only policy. The data efficiency, low cost, compact design, and robust performance of TranTac underscore its potential for advancing delicate robotic manipulation in contact-rich environments.

References

[1] Raunaq Bhirangi, Venkatesh Pattabiraman, Enes Erciyes, Yifeng Cao, Tess Hellebrekers, and Lerrel Pinto. Anyskin: Plug-and-play skin sensing for robotic touch. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pages 16563–16570. IEEE, 2025.

[2] Kevin Black, Noah Brown, James Darpinian, Karan Dhabalia, Danny Driess, Adnan Esmail, Michael Robert Equi, Chelsea Finn, Niccolo Fusai, Manuel Y Galliker, et al. $\pi_{0.5}$: a vision-language-action model with open-world generalization. In *9th Annual Conference on Robot Learning*, 2025.

[3] Anthony Brohan, Noah Brown, Justice Carbajal, Yevgen Chebotar, Joseph Dabis, Chelsea Finn, Keerthana Gopalakrishnan, Karol Hausman, Alex Herzog, Jasmine Hsu, et al. Rt-1: Robotics transformer for real-world control at scale. *arXiv preprint arXiv:2212.06817*, 2022.

[4] Max Yiye Cao, Stephen Laws, and Ferdinando Rodriguez y Baena. Six-axis force/torque sensors for robotics applications: A review. *IEEE Sensors Journal*, 21(24):27238–27251, 2021.

[5] Cheng Chi, Zhenjia Xu, Siyuan Feng, Eric Cousineau, Yilun Du, Benjamin Burchfiel, Russ Tedrake, and Shuran Song. Diffusion policy: Visuomotor policy learning via action diffusion. *The International Journal of Robotics Research*, 44(10-11):1684–1704, 2025.

[6] Ravinder S. Dahiya, Giorgio Metta, Maurizio Valle, and Giulio Sandini. Tactile sensing—from humans to humanoids. *IEEE Transactions on Robotics*, 26(1):1–20, 2010.

[7] Siyuan Dong, Devesh K Jha, Diego Romeres, Sangwoon Kim, Daniel Nikovski, and Alberto Rodriguez. Tactile-rl for insertion: Generalization to objects of unknown geometry. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6437–6443. IEEE, 2021.

[8] Maximilian Du, Olivia Y Lee, Suraj Nair, and Chelsea Finn. Play it by ear: Learning skills amidst occlusion through audio-visual imitation learning. *arXiv preprint arXiv:2205.14850*, 2022.

[9] Jana Egli, Benedek Forrai, Thomas Buchner, Jiangtao Su, Xiaodong Chen, and Robert K Katzschmann. Sensorized soft skin for dexterous robotic hands. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 18127–18133. IEEE, 2024.

[10] Niklas Funk, Erik Helmut, Georgia Chalvatzaki, Roberto Calandra, and Jan Peters. Evetac: An event-based optical tactile sensor for robotic manipulation. *IEEE Transactions on Robotics*, 40:3812–3832, 2024.

[11] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[12] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

[13] Binghao Huang, Yixuan Wang, Xinyi Yang, Yiyue Luo, and Yunzhu Li. 3d-vitac: Learning fine-grained manipulation with visuo-tactile sensing. In *8th Annual Conference on Robot Learning*, 2024.

[14] Roland S Johansson and J Randall Flanagan. Coding and use of tactile signals from the fingertips in object manipulation tasks. *Nature Reviews Neuroscience*, 10(5):345–359, 2009.

[15] Zhanat Kappassov, Juan-Antonio Corrales, and Véronique Perdereau. Tactile sensing in dexterous robot hands. *Robotics and Autonomous Systems*, 74:195–220, 2015.

[16] Ulf Kasolowsky and Berthold Bäuml. Fine manipulation using a tactile skin: Learning in simulation and sim-to-real transfer. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 13120–13127. IEEE, 2024.

[17] Moo Jin Kim, Karl Pertsch, Siddharth Karamcheti, Ted Xiao, Ashwin Balakrishna, Suraj Nair, Rafael Rafailov, Ethan Foster, Grace Lam, Pannag Sanketi, et al. Openvla: An open-source vision-language-action model. In *9th Annual Conference on Robot Learning*, 2025.

[18] Sangwoon Kim and Alberto Rodriguez. Active extrinsic contact sensing: Application to general peg-in-hole insertion. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 10241–10247. IEEE, 2022.

[19] Kenta Kumagai and Kazuhiro Shimonomura. Event-based tactile image sensor for detecting spatio-temporal fast phenomena in contacts. In *2019 IEEE World Haptics Conference (WHC)*, pages 343–348. IEEE, 2019.

[20] Naveen Kuppuswamy, Alex Alspach, Avinash Uttamchandani, Sam Creasey, Takuya Ikeda, and Russ Tedrake. Soft-bubble grippers for robust and perceptive manipulation. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 9917–9924. IEEE, 2020.

[21] Mike Lambeta, Po-Wei Chou, Stephen Tian, Brian Yang, Benjamin Maloon, Victoria Rose Most, Dave Stroud, Raymond Santos, Ahmad Byagowi, Gregg 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*, 5(3):3838–3845, 2020.

[22] Hao Li, Yizhi Zhang, Junzhe Zhu, Shaoxiong Wang, Michelle A Lee, Huazhe Xu, Edward Adelson, Li Fei-Fei, Ruohan Gao, and Jiajun Wu. See, hear, and feel: Smart sensory fusion for robotic manipulation. In *7th Annual Conference on Robot Learning*, pages 1368–1378, 2023.

[23] Qiang Li, Oliver Kroemer, Zhe Su, Filipe Fernandes Veiga, Mohsen Kaboli, and Helge Joachim Ritter. A review of tactile information: Perception and action through touch. *IEEE Transactions on Robotics*, 36(6):1619–1634, 2020.

[24] Changyi Lin, Han Zhang, Jikai Xu, Lei Wu, and Huazhe Xu. 9dtact: A compact vision-based tactile sensor for accurate 3d shape reconstruction and generalizable 6d force estimation. *IEEE Robotics and Automation Letters*, 9(2):923–930, 2023.

[25] Zeyi Liu, Cheng Chi, Eric Cousineau, Naveen Kuppuswamy, Benjamin Burchfiel, and Shuran Song. Maniwav: Learning robot manipulation from in-the-wild audio-visual data. In *8th Annual Conference on Robot Learning*, 2024.

[26] Jianlan Luo, Charles Xu, Fangchen Liu, Liam Tan, Zipeng Lin, Jeffrey Wu, Pieter Abbeel, and Sergey Levine. Fmb: a functional manipulation benchmark for generalizable robotic learning. *The International Journal of Robotics Research*, 44(4):592–606, 2025.

[27] Yiyue Luo, Yunzhu Li, Pratyusha Sharma, Wan Shou, Kui Wu, Michael Foshey, Beichen Li, Tomás Palacios, Antonio Torralba, and Wojciech Matusik. Learning human–environment interactions using conformal tactile textiles. *Nature Electronics*, 4(3):193–201, 2021.

[28] Daolin Ma, Siyuan Dong, and Alberto Rodriguez. Extrinsic contact sensing with relative-motion tracking from distributed tactile measurements. In *2021 IEEE international conference on robotics and automation (ICRA)*, pages 11262–11268. IEEE, 2021.

[29] Jared Mejia, Victoria Dean, Tess Hellebrekers, and Abhinav Gupta. Hearing touch: Audio-visual pretraining for contact-rich manipulation. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6912–6919. IEEE, 2024.

[30] Ethan Perez, Florian Strub, Harm De Vries, Vincent Dumoulin, and Aaron Courville. Film: Visual reasoning with a general conditioning layer. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

[31] Amin Rigi, Fariborz Baghaei Naeini, Dimitrios Makris, and Yahya Zweiri. A novel event-based incipient slip detection using dynamic active-pixel vision sensor (davis). *Sensors*, 18(2):333, 2018.

[32] Yitian Shao, Vincent Hayward, and Yon Visell. Compression of dynamic tactile information in the human hand. *Science advances*, 6(16):eaaz1158, 2020.

[33] Kazuhiro Shimonomura. Tactile image sensors employing camera: A review. *Sensors*, 19(18):3933, 2019.

[34] Yuki Shirai, Devesh K Jha, Arvind U Raghunathan, and Dennis Hong. Tactile tool manipulation. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 12597–12603. IEEE, 2023.

[35] Ian H Taylor, Siyuan Dong, and Alberto 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)*, pages 10781–10787. IEEE, 2022.

[36] Octo Model Team, Dibya Ghosh, Homer Walke, Karl Pertsch, Kevin Black, Oier Mees, Sudeep Dasari, Joey Hejna, Tobias Kreiman, Charles Xu, et al. Octo: An open-source generalist robot policy. *arXiv preprint arXiv:2405.12213*, 2024.

[37] Abitha Thankaraj and Lerrel Pinto. That sounds right: Auditory self-supervision for dynamic robot manipulation. In *7th Annual Conference on Robot Learning*, 2023.

[38] Neeli Tummala, Yitian Shao, and Yon Visell. Spatiotemporal organization of touch information in tactile neuron population responses. In *2023 IEEE World Haptics Conference (WHC)*, pages 183–189. IEEE, 2023.

[39] Benjamin Ward-Cherrier, Nicholas Pestell, and Nathan F Lepora. Neurotac: A neuromorphic optical tactile sensor applied to texture recognition. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2654–2660. IEEE, 2020.

[40] Yansong Wu, Zongxie Chen, Fan Wu, Lingyun Chen, Liding Zhang, Zhenshan Bing, Abdalla Swikir, Sami Haddadin, and Alois Knoll. Tacdiffusion: Force-domain diffusion policy for precise tactile manipulation. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pages 11831–11837. IEEE, 2025.

[41] Han Xue, Jieji Ren, Wendi Chen, Gu Zhang, Yuan Fang, Guoying Gu, Huazhe Xu, and Cewu Lu. Reactive diffusion policy: Slow-fast visual-tactile policy learning for contact-rich manipulation. In *Proceedings of Robotics: Science and Systems (RSS)*, 2025.

[42] Chaojie Yan, Jun Wu, and Qiuguo Zhu. Learning-based contact status recognition for peg-in-hole assembly. In *2021 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pages 6003–6009. IEEE, 2021.

[43] Wenzhen Yuan, Siyuan Dong, and Edward H Adelson. Gelsight: High-resolution robot tactile sensors for estimating geometry and force. *Sensors*, 17(12):2762, 2017.

[44] Tony Z Zhao, Vikash Kumar, Sergey Levine, and Chelsea Finn. Learning fine-grained bimanual manipulation with low-cost hardware. *arXiv preprint arXiv:2304.13705*, 2023.

[45] Brianna Zitkovich, Tianhe Yu, Sichun Xu, Peng Xu, Ted Xiao, Fei Xia, Jialin Wu, Paul Wohlhart, Stefan Welker, Ayzaan Wahid, et al. Rt-2: Vision-language-action models transfer web knowledge to robotic control. In *7th Annual Conference on Robot Learning*, 2023.