AnySkin: Plug-and-play Skin Sensing for Robotic Touch

Raunaq Bhirangi[1,2,†], Venkatesh Pattabiraman[1], Enes Erciyes[1], Yifeng Cao[3], Tess Hellebrekers[4], Lerrel Pinto[1]

[1]New York University, [2]Carnegie Mellon University, [3]Columbia University, [4]Meta AI Research

[†] Correspondence to: raunaqbhirangi@nyu.edu

Abstract: While tactile sensing is widely accepted as an important and useful sensing modality, its use pales in comparison to other sensory modalities like vision and proprioception. AnySkin addresses the critical challenges that impede the use of tactile sensing – versatility, replaceability, and data reusability. Building on the simple design of ReSkin, and decoupling the sensing electronics from the sensing interface, AnySkin makes integration as straightforward as putting on a phone case and connecting a charger. Furthermore, AnySkin is the first uncalibrated tactile-sensor to report cross-instance generalizability of learned manipulation policies. To summarize, this work makes three key contributions: first, we introduce a streamlined fabrication process and a design tool for creating an adhesive-free, durable and replaceable magnetic tactile sensor; second, we characterize slip detection and policy learning with the AnySkin sensor; third, we demonstrate zero-shot generalization of models trained on one instance of AnySkin to new instances, and compare it with popular existing tactile solutions like DIGIT and ReSkin. Videos and details can be found on https://any-skin.github.io/.

Keywords: Tactile Sensing, Soft Robotics

1 Introduction

Touch sensing is widely recognized as a crucial modality for biological movement and control [1, 2]. Unlike vision, sound, or proprioception, touch provides sensing at the point of contact, allowing agents to perceive and reason about forces and pressure. However, a closer examination of robotics literature reveals a different narrative. Prominent works and current state-of-the-art in robot learning primarily utilize vision sensing in conjunction with proprioception to train manipulation skills [3, 4, 5, 6], often ignoring touch. If touch is indeed vital from a biological perspective, why does it remain a second-class citizen in sensorimotor control?

To address this question, let's examine what made cameras ubiquitous in robotics. Three key factors are at play: cost, convenience, and consistency. Cameras are relatively inexpensive (under $20), easy to integrate on a wide variety of robot platforms (e.g. multi-view, depth, ego-centric), and allow for models trained on them (e.g. object detection, segmentation) to easily transfer to images captured with new cameras. In contrast, touch sensors are often costly due to expensive fabrication processes [7] (e.g., pressure-based sensors) or the need for high-end components [8] (e.g., Biotac). They are inconvenient to use on different robot platforms, being custom-built for specific robot end-effectors and constrained form factors requiring extensive adaptation for different shapes [9, 10]. Finally, touch sensors are inconsistent. Due to boutique fabrication, sensor profiles can vary significantly even when produced through the same process [11, 12]. This inconsistency poses a challenge when transferring tactile-based models across different instances of the same sensor. This transfer is particularly critical for touch sensors due to their persistent need for replacement. Soft sensing interfaces, which are important for touch sensors to maintain a stable grip, wear out more quickly than hard interfaces, resulting in more frequent replacements.

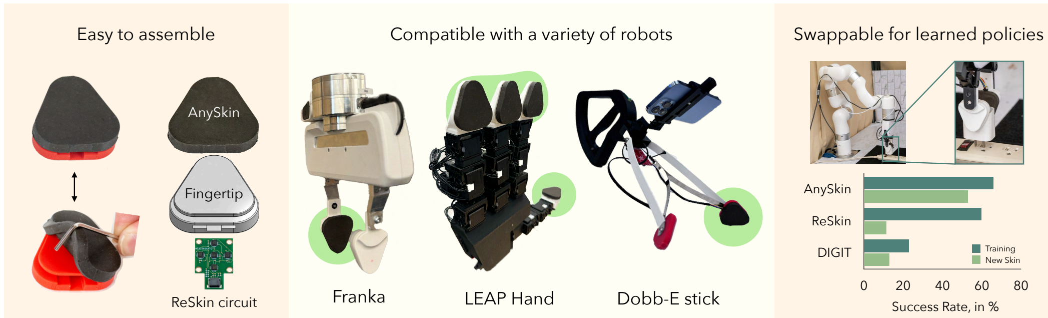


Figure 1: We present AnySkin, a skin sensor made for robotic touch that is easy to assemble, compatible with different robot end-effectors and generalizes to new skin instances. AnySkin senses contact through distortions in magnetic field generated by the magnetized skin. The flexible surface is physically separated from its electronics, which allows for easy replacability when damaged.

In this work we present AnySkin, a new touch sensor that is cheap, convenient to use and has consistent response across different sensor instances. AnySkin builds on ReSkin [11], a magnetic-field based touch sensor, by improving its fabrication, separating the sensing mechanism from the interaction surface, and developing a new self-adhering, self-aligning attachment mechanism. This allows AnySkin to (a) have stronger magnetic fields, which significantly improves its sensor response, (b) be easy to fabricate for arbitrary surface shapes, which allows easy use on different end-effectors, (c) be easy to replace the sensor without adversely affecting the data collection process or the efficacy of models trained on previous sensors (Fig. 1).

We run a suite of experiments to understand the efficacy of AnySkin vis-a-viz other prominent touch sensors. Our main findings can be summarized below:

1. AnySkin can readily be used on a variety of robots including xArm, Franka, and the four-fingered Leap hand (See fabrication details in Section 4).
2. AnySkin is compatible with ML techniques for slip detection and visuo-tactile policy learning for precise tasks such as inserting USBs (See learning details in Section 5).
3. AnySkin takes an average of 12 seconds to replace and can be reused after replacement (See replacement study in Section 5.3).
4. Models trained on one AnySkin transfer zero-shot to a different AnySkin with only a 13% reduction in performance on a plug insertion task compared to the 43% drop in performance with ReSkin [11] sensors.

AnySkin is fully open-sourced. Videos of fabrication, attachment, and robot policies are best viewed on our project website: `https://any-skin.github.io/`.

2 Related Work

2.1 Tactile sensing

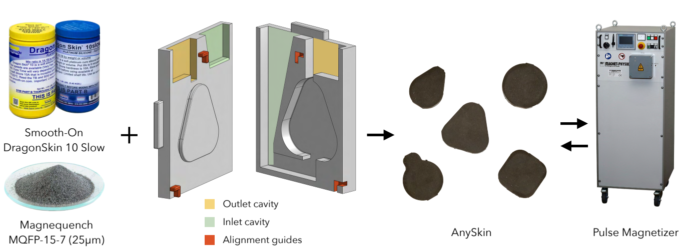
Existing literature on tactile sensing explores a wide range of modalities, each with their own set of advantages and limitations. Capacitative sensors [13, 14, 15, 16] sense contact through changes in capacitance, offering high sensitivity. However, they struggle to model shear force and are prone to breakage due to direct electrical connections between the circuitry and elastomer. Resistive sensors [7, 17, 18] are simple and durable, but tend to provide spatially discrete sensing with low spatial resolution. MEMS-based sensors [8] offer significant versatility by combining multiple sensors like audio and IMU sensors for multimodal feedback in a mm-scale form factor, but tend to use high-end components and are expensive to fabricate. Optical sensors [10, 19, 20, 21] capture high-resolution contact information using cameras to track the deformation of an elastomer, but often pose hard, stringent limits on the sensor form factor, due to the physical constraints on the camera field of view. This complicates integration for a wide range of applications and significantly increases the effort required to sensorize surfaces of different shapes and sizes.

Magnetic tactile sensors [22, 23, 24] largely overcome these limitations due to three salient advantages: (a) separation of the sensing electronics from the sensing interface to improve robustness (b) compatibility with different form factors, and (c) an ability to capture shear forces [11]. Two prominent classes of magnetic sensors in robotics right now - ReSkin [11] and uSkin [23] - use elastomeric sensing interfaces with magnetic microparticles and macro-sized magnets respectively. In this work, we build on ReSkin sensors due to their lower cost and ease of fabrication.

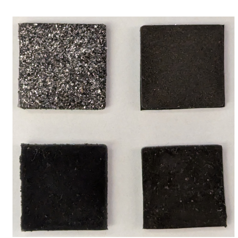
2.2 Replaceability for Tactile Sensors

Recent developments in rapid prototyping and elastomer technology have spurred a substantial rise in the number of robotic tactile sensors. Most tactile solutions rely on soft sensing interfaces to enable stable conformal contact with objects in the environment. Soft interfaces are prone to frequent wear and tear from contact-rich interactions, but discussions on replaceability for tactile sensors remain few and far between. There are two main factors to consider when evaluating replaceability: (a) the physical ease of replacing the sensory interface, and (b) signal consistency when replacing a worn out instance with a new instance. The former is far more frequently discussed [8, 11, 20, 25] and resolved by simply separating the sensing interface – generally the damage-prone soft elastomer – from the sensing electronics which last much longer. The latter, however, is much less discussed. Prior work in tactile sensing has found significant variation across different instances of the same tactile sensor [11, 12]. Higher susceptibility to wear combined with lower signal consistency across sensor instances severely restricts the scale of data explored in most existing work on tactile learning [26, 27].

This effect is even more pronounced for policy learning where imitation learning as well as reinforcement learning approaches have been used to show impressive results on real-world robots [4, 28, 29]. However, both approaches rely on significant quantities of data, be it demonstration or online interaction. The necessity of using a single sensor instance across training and testing has severely limited the extent of capabilities demonstrated with visuotactile learning [30, 31]. Recent research has either relied on policy learning in simulation using simplified models of the tactile sensor [32, 33, 34], or used analytical models for manual feature extraction and dimensionality reduction [35, 36]. The former results in a significant dilution in the amount of tactile information and is therefore restricted to less precise tasks with simpler contact reasoning. The latter techniques are often specific to the task they solve, difficult to scale and show limited generalizability beyond the restrictive settings they operate in. In this light, signal consistency across instances is central to building scalable and generalizable tactile models, and making tactile sensing a ubiquitous presence in robot learning. In Section 5, we quantitatively demonstrate the improved consistency of AnySkin signal over ReSkin, and present a direct replaceability comparison with DIGIT [20] and ReSkin through policy learning.



(a) Fabrication of AnySkin



(b) Top and bottom surfaces of skins

Figure 2: (a) AnySkin is made by mixing Smooth-On DragonSkin 10 Slow and MQFP-15-7(25μm) magnetic particles in a 1:1:2 ratio, and curing it in the two-part molds shown above. Cured skins are magnetized using a pulse magnetizer. (b) Skins made with MQP-15-7(-80 mesh) and MQFP-15-7(25μm) particles. Note the concentration of particles at the surface of the former due to the larger particle size.

3 AnySkin: Components

AnySkin builds on ReSkin [11], a tactile skin composed of a soft magnetized skin coupled with magnetometer-based sensing circuitry. By detecting distortions in magnetic fields, ReSkin measures skin deformations caused by normal and shear forces [22, 11]. Its adaptability enables integration across various applications, from robotic hands [26] to arm sleeves and even dog shoes. AnySkin uses the same 5-magnetometer circuitry as ReSkin, while introducing key design and fabrication changes to the skin to improve durability, repeatability, and replaceability, including:

- Magnetizing skins post-curing using a pulse magnetizer.
- Introducing physical separation between magnetic elastomer and magnetometer circuit.
- Utilizing finer magnetic particles to achieve a more uniform particle distribution.
- Implementing self-aligning design for reduced variability in positioning of elastomers and circuitry.

While some of these additions have been used in isolation in prior work [26, 37, 38], there has been little discussion on their effect on sensor response. In this section, we elaborate on the rationale for each choice, followed by a quantitative comparison of the sensor response in Section 5.1.

ReSkin uses a grid of cube magnets during curing to magnetize the elastomeric skins. While effective, this approach has several drawbacks, such as producing skins with relatively weak magnetic fields. As a result, although the design of ReSkin separates the sensing electronics and the sensing interface, adding physical distance between the skin and the sensors significantly weakens the signal, making it infeasible in reality. In contact-rich tasks, where the sensing skin undergoes considerable strain, the absence of physical separation leads to stress being transmitted directly to the electronics, ultimately compromising their durability.

Additionally, applying a magnetic field during elastomer curing increases variability in the signal response. Before curing, magnetic particles are free to move through the liquid elastomer under the effect of the magnetic field. As a result, the distribution of particles is influenced by the temporal evolution of the applied magnetic field, i.e. how you move the magnets into place, which can be difficult to control when fabricating. To circumvent these disadvantages, we propose using a pulse magnetizer to magnetize the skins post-curing in line with [26, 38], as shown in Fig. 2a. The pulse magnetizer can apply a large enough magnetic field to magnetize the dipoles in the magnetic elastomer. Curing outside the influence of magnetic fields allows for a more uniform distribution of magnetic particles through the bulk of the sensor, thereby improving magnetic field consistency.

However, simply changing the magnetization procedure results in other problems. Curing outside the influence of a magnetic field causes particles to settle to the bottom of the sensing skin due to gravity (See Fig. 2b). This results in lower durability as the skin sheds magnetic particles, particularly during contact-rich interactions. To get around this problem, we replace the magnetic particles with much finer particles (details in Section 4.2). The smaller particles operate in a sufficiently low Reynolds number regime to allow the elastomer to cure before they can settle on one surface of the elastomer.

Finally, since ReSkin relies on magnetic field distortions to measure contact characteristics, sensor response is strongly influenced by the relative positioning of the magnetic skin and circuitry (see Section 5.1). This means that loss of adhesion, peeling, or other relative motions between the skin and circuitry over the life of the skin can significantly affect the consistency of the signal. Ideally, we would like the skin to stay adhered until it needs to be replaced due to wear and tear. Using screws to adhere the skin as suggested in [11] results in poor durability due to a stress concentration at the screw-skin interface, especially in tasks involving shear forces. Using an adhesive like Silpoxy [20, 26] can be used to create sticker-like skins that last relatively longer but still tend to peel under repetitive cyclic loading. With AnySkin, we eliminate the need for adhesives or fasteners by modifying the design of the skins to be self-adhering. Additionally, we also eliminate the need to manually align skin and circuitry, significantly improving signal consistency as demonstrated in Section 5.1.

4 AnySkin: Fabrication

The overall fabrication procedure follows the general outline of ReSkin: Magnetic particles and elastomer are mixed in specified ratios; the resulting mixture is poured into the molds; cured skins are magnetized. The shape of the fingertip-skin assembly is designed to be triangular as shown in Fig. 1 to improve reachability. In this section, we elaborate on the details of the fabrication procedure for AnySkin, and key changes to the ReSkin fabrication procedure that result in a new, upgraded sensor.

4.1 Mold design

The shape of the magnetic skin is dictated by the molds used for curing. To create self-adhering skins as outlined in Section 3, we present a two-part mold design as shown in Fig. 2a. We choose a skin thickness of 2 mm following [11] with a triangular shape for its advantageous form factor for precise manipulation. All the experiments presented in this paper use this triangular skin. We also open-source a mold design CAD tool that generates designs for the fingertip as well as 2-part molds from just a 2D drawing. Unlike tactile sensors that require significant engineering for changes in form factor [20, 10], AnySkin makes it effortless to diversify your tactile sensor.

4.2 Elastomer composition

For AnySkin, we mix magnetic microparticles and two-part polymer (Dragonskin 10 Slow; Smooth-On) in the same 2:1:1 ratio as ReSkin, while using finer Magnequench MQFP-15-7(25μm). These particles are about 100x smaller than the MQP-15-7(-80 mesh) used with ReSkin, and do not settle before curing, due to their lower Reynolds number [39]. This ensures that magnetic particles are more evenly distributed through the volume of the skin, thereby improving consistency of the signal.

4.3 Magnetization

ReSkin is magnetized by sandwiching the magnetic elastomer mix between a grid of magnets while it is curing. This results in higher variance in distribution of magnetic particles within the core of the skin based on the exact timing of sandwiching the skins. Drawing from D'Manus [26], we use a pulse magnetizer for magnetizing the skins after curing is complete. Separating the magnetizing process from the curing process allows the skins to cure undisturbed and maintain a more uniform distribution of magnetic particles. Furthermore, the magnetic field applied by the pulse magnetizer is far stronger than the sandwich of grid magnets. This results in skins with stronger magnetic fields, which in turn enables larger separations between magnetic skin and magnetometer circuitry.

4.4 Magnetic elastomer fabrication

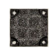
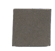
The final fabrication process follows similar steps as the ReSkin fabrication process. The molds are first aligned using the built-in alignment guides and clicked together. We use plastic clamps to hold the parts together. The two-part elastomer compound is then mixed and degassed. This is followed by the addition of magnetic micro particles and another round of mixing and degassing. Once degassing is complete, the magnetic elastomer mix is poured through the mold inlet as shown in Fig. 2 until it emerges at the outlet, pausing as necessary to allow the mixture to flow through and fill the entire mold. The filled mold is then placed in a vacuum chamber and a pressure of 29mm of Hg/in is applied, again pausing as necessary to prevent overflow as the liquid bubbles. This pressure is held for 10 minutes before releasing the vacuum. The molds are allowed to rest for 16 hours, before prying them open and trimming excess material to reveal the fully cured AnySkin.

5 Experimental Results

5.1 Comparison between ReSkin and AnySkin signal

We built upon ReSkin's fabrication methods to develop AnySkin, introducing modifications that improve durability, repeatability (5.4.1), and replaceability (5.3). To quantitatively demonstrate the effect of each of the fabrication changes (See Section 4) towards improving the consistency of AnySkin, we present a set of experiments analyzing the raw signal from the four different skins shown in Table 1, tracking the progression from ReSkin to AnySkin. All statistics presented are computed over five instances of each type.

Table 1: AnySkin shows lower variability across instances. Statistics computed over 5 samples of each type (PM: Pulse magnetizer, FP: finer particles, SA: self-aligning – AnySkin).

Experiment	 ReSkin		 +PM		 +FP		 +SA	
	B_{xy}	B_z	B_{xy}	B_z	B_{xy}	B_z	B_{xy}	B_z
Average strength, in μT	1062	302	**1818**	**5212**	**1602**	**5784**	283	1265
Normalized σ (instances)	0.54	0.87	0.34	0.12	0.21	0.15	**0.12**	**0.10**
Normalized σ (1mm displaced)	1.38	1.43	0.25	0.11	0.18	0.07	**Self-aligning**	

We first see a significant increase in the raw magnetic field by using the pulse magnetizer for magnetization. Next, finer particles results in both a lower variability across instances as well as durability due to reduced leakage of particles resulting from stress concentration on the surface as shown in Fig. 2b. Finally, the self-aligning design of AnySkin removes the possibility of misalignment between skin and circuitry which introduced significant variability in the signal. This design also adds a physical separation between the electronics and the sensing interface which improves durability while maintaining the signal strength of ReSkin.

5.2 Slip Detection

We quantify AnySkin's ability to detect slip through a controlled experiment, using an LSTM [40] to train our slip prediction models purely from tactile data. Our model is able to detect slip on unseen objects with 92% accuracy (See Section A for setup details).

5.3 Ease of replaceability

We compare the replacement time of AnySkin against other skins like DIGIT and ReSkin through a user study with 10 non-expert users. We find that ReSkin takes 236 ± 64 seconds, DIGIT takes 58 ± 22 seconds, while AnySkin takes a mere 12 ± 5 seconds to replace.

However, the most important consequence of the signal consistency and replaceability of AnySkin is its ability to enable policy generalization across different instances of the skin. We demonstrate cross-instance generalizability of AnySkin across three precise manipulation tasks shown in Fig. 3 and compare the cross-instance generalizability of policies trained on DIGIT, ReSkin and AnySkin on the plug insertion task.

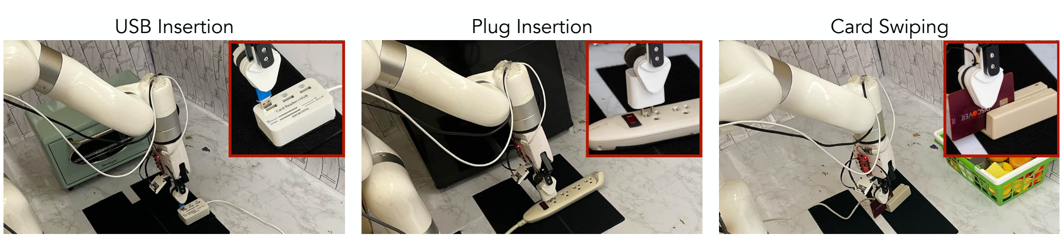


Figure 3: We evaluate the replaceability of AnySkin on a set of contact-rich, precision tasks.

The BAKU [41] architecture is used as the policy architecture. BAKU tokenizes each input using a modality-specific encoder: image inputs from cameras and DIGITs use ResNet-18 [42] encoders, while AnySkin and ReSkin inputs use MLP encoders. The policy uses a deterministic action head and action chunking [43] with exponential temporal averaging.

5.4 Replaceability in Policy Learning

The most important consequence of the signal consistency and replaceability of AnySkin outlined so far, is its ability to enable policy generalization across different instances of the skin. In this section, we demonstrate the cross-instance generalizability of AnySkin across three precise manipulation tasks. We follow this up with a comparison of the cross-instance generalizability of policies trained on DIGIT, ReSkin and AnySkin on the plug insertion task. See Section B for more details on the policy learning experimental setup (B.1) and our model architecture and training details (B.2).

5.4.1 Evaluating cross-instance generalizability

To investigate the replaceability of AnySkin in the context of policy learning, we evaluate behavior cloning policies trained using a single instance of AnySkin on a new instance. Note that we use a different training and test skin for *each* of the presented tasks to avoid over-indexing on specific skin instances. Table 2 presents a comparison between policy performance with the original and swapped skins for each of the precise, contact-rich tasks outlined above. Additionally, we train and evaluate a policy that only takes camera images as input to serve as a control experiment. We find that across tasks, performance drops by an average of just 15.6% and visuotactile policies with swapped skins continue to do significantly better than purely visual policies. This result demonstrates the strength and uniqueness of AnySkin as a tactile sensor for contact-rich manipulation.

5.4.2 Comparison across sensors

To better contextualize the significance of this result, we present a replaceability comparison with DIGIT [20] and ReSkin [11] sensors. We collect two additional datasets of 96 demonstration trajectories each for the plug insertion task with these sensors similar to AnySkin. Replaceability is evaluated by swapping the training skin for a new skin during evaluation as outlined in the previous section. Success rates from 10 evaluations across three seeds for each setting are reported in Table 2.

Table 2: Success rates (out of 10) for policies when swapping out tactile skins. All statistics computed over 3 training seeds

Task	Cameras only	Cameras + Skin	
		Original skin	Swapped skin
Cross-instance generalization			
Plug Insertion	1.7 ± 0.6	6.7 ± 1.5	5.3 ± 2.5
Card Swiping	2.0 ± 1.0	7.0 ± 1.7	6.3 ± 0.6
USB Insertion	1.7 ± 1.2	5.7 ± 1.5	3.0 ± 1.0
Comparison across sensors – Plug Insertion			
AnySkin	1.7 ± 0.6	$\mathbf{6.7 \pm 1.5}$	$\mathbf{5.3 \pm 2.5}$
ReSkin	1.7 ± 1.2	6.0 ± 1.7	1.7 ± 1.2
DIGIT	1.7 ± 1.5	2.3 ± 0.6	1.3 ± 0.6

Based on these results, we find that visuotactile policies trained with ReSkin and AnySkin have similar performance on solving the plug insertion task. However, when the skin is replaced, the performance of the ReSkin policy falls 43% to the same level as the camera-only policy, while the performance of AnySkin policies only drops by 13%. This transferability is evidence of AnySkin's superior signal consistency, and is a significant boost to scaling efforts like training large tactile models as well as real world deployment of models trained in the laboratory.

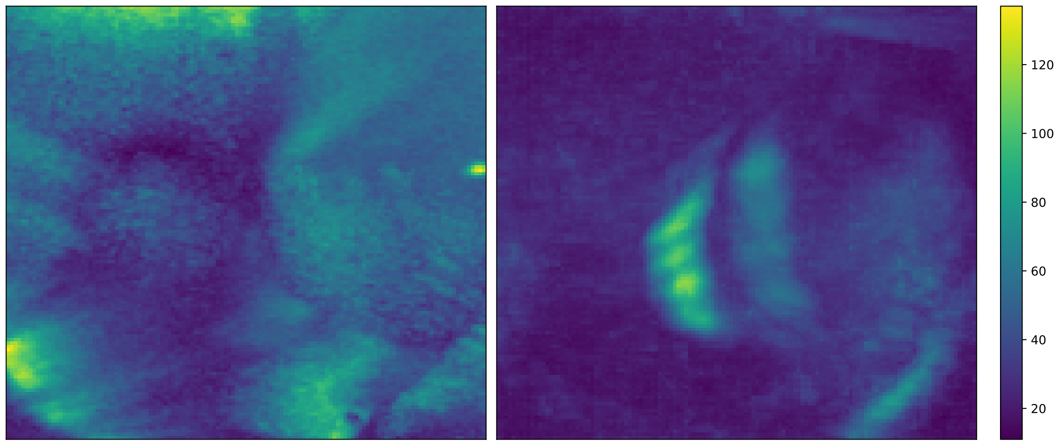


Figure 4: Pixel-wise difference between two different DIGIT sensor instances (left) and the maximum difference in response of one DIGIT sensor on the task of plug insertion across 96 demos (right).

An unexpected result from these experiments was the poor success rate of policies trained using the DIGIT sensor which has been shown to be successful in other visuotactile tasks, perhaps on less precise [32] or less reactive [12] ones, in prior work. Consequently, while there is still a gap in performance, we don't see a significant gap between the poor performance of the visuotactile policy on the original and swapped DIGIT skins. However, the high variability across instances of the DIGIT sensors is previously documented [12], and we find that a closer look at the DIGIT signal from our plug insertion dataset indicates that even if it were possible to train more performant policies, they are unlikely to generalize across instances. Across the 96 demonstration trajectories from the plug insertion, we compute the *maximum* change in pixel values across channels and across trajectories, and compare it against the pixel-wise differences between the original and a swapped instance of the DIGIT sensor in Fig. 4. Since the *maximum* change in sensory measurement through the course of the interaction is comparable to the difference in signal between two instances, it is unlikely that policies trained on one sensor will generalize to new instances.

6 Conclusion and Limitations

In this paper, we present AnySkin, a new magnetic tactile sensor. AnySkin is versatile, self-adhering and improves on signal consistency across different instances of the skin. Furthermore, to the best of our knowledge, AnySkin is the first sensor to demonstrate zero-shot generalization of visuotactile policies to new instances of the tactile skin. It could be incorporated into a large-scale data collection tool such as UMI [44] or the Stick [45, 29], and effectively used for training useful, generalizable models. Future work could explore simple calibration schemes and improve fabrication to close the gap between training and swapped skin instances, and enhance signal consistency.

Despite its potential, AnySkin still inherits some of the drawbacks of the ReSkin sensor, primary among them being interference from magnetic and ferromagnetic objects in the environment. Using machine learning approaches for noise reduction in magnetic sensors [38] or improving the skin design to incorporate a Faraday cage could help resolve these difficulties, and take tactile sensing one step closer to being a first-class citizen in robotics.

Our experiments were performed using DIGIT sensors with the standard, commercially available fingertip gel, but prior work has found some success learning visuotactile policies using optical sensors with dotted skins [46, 47]. An interesting direction for future work could be comparing the performance of behavior cloning policies using different gel tips for optical sensors. However, while this may improve learning performance with a single skin, cross-instance generalizability might still require significant changes in the fabrication of optical sensors.

Acknowledgments

Special thanks for Krishna Bodduluri, Mike Lambeta, and team from Meta AI Research for providing the DIGIT sensors for comparison. NYU authors are supported by grants from Honda, Hyundai, NSF award 2339096 and ONR awards N00014-21-1-2758 and N00014-22-1-2773. LP is supported by the Packard Fellowship.

References

[1] J. Jenner and J. Stephens. Cutaneous reflex responses and their central nervous pathways studied in man. *The Journal of physiology*, 333(1):405–419, 1982.

[2] R. S. Johansson. Sensory control of dexterous manipulation in humans. In *Hand and brain*, pages 381–414. Elsevier, 1996.

[3] Z. Fu, A. Kumar, A. Agarwal, H. Qi, J. Malik, and D. Pathak. Coupling vision and proprioception for navigation of legged robots. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 17273–17283, June 2022.

[4] A. Padalkar, A. Pooley, A. Jain, A. Bewley, A. Herzog, A. Irpan, A. Khazatsky, A. Rai, A. Singh, A. Brohan, et al. Open x-embodiment: Robotic learning datasets and rt-x models. *arXiv preprint arXiv:2310.08864*, 2023.

[5] C. Chi, S. Feng, Y. Du, Z. Xu, E. Cousineau, B. Burchfiel, and S. Song. Diffusion policy: Visuomotor policy learning via action diffusion. *arXiv preprint arXiv:2303.04137*, 2023.

[6] H. Bharadhwaj, J. Vakil, M. Sharma, A. Gupta, S. Tulsiani, and V. Kumar. Roboagent: Generalization and efficiency in robot manipulation via semantic augmentations and action chunking, 2023. URL https://arxiv.org/abs/2309.01918.

[7] S. Sundaram, P. Kellnhofer, Y. Li, J.-Y. Zhu, A. Torralba, and W. Matusik. Learning the signatures of the human grasp using a scalable tactile glove. *Nature*, 569(7758):698–702, 2019.

[8] N. Wettels, V. J. Santos, R. S. Johansson, and G. E. Loeb. Biomimetic tactile sensor array. *Advanced robotics*, 22(8):829–849, 2008.

[9] 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)*, pages 6468–6475. IEEE, 2021.

[10] 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)*, pages 10781–10787. IEEE, 2022.

[11] R. Bhirangi, T. Hellebrekers, C. Majidi, and A. Gupta. Reskin: versatile, replaceable, lasting tactile skins. In *5th Annual Conference on Robot Learning*, 2021.

[12] S. Suresh, H. Qi, T. Wu, T. Fan, L. Pineda, M. Lambeta, J. Malik, M. Kalakrishnan, R. Calandra, M. Kaess, et al. Neural feels with neural fields: Visuo-tactile perception for in-hand manipulation. *arXiv preprint arXiv:2312.13469*, 2023.

[13] O. Glauser, D. Panozzo, O. Hilliges, and O. Sorkine-Hornung. Deformation capture via soft and stretchable sensor arrays. *ACM Transactions on Graphics (TOG)*, 38(2):1–16, 2019.

[14] T.-Y. Wu, L. Tan, Y. Zhang, T. Seyed, and X.-D. Yang. Capacitivo: Contact-based object recognition on interactive fabrics using capacitive sensing. In *Proceedings of the 33rd annual acm symposium on user interface software and technology*, pages 649–661, 2020.

[15] D. Xu, A. Tairych, and I. A. Anderson. Stretch not flex: programmable rubber keyboard. *Smart Materials and Structures*, 25, 2015. URL https://api.semanticscholar.org/CorpusID:111206581.

[16] B. Xu, L. Zhong, G. Zhang, X. Liang, D. Virtue, R. Madan, and T. Bhattacharjee. Cushsense: Soft, stretchable, and comfortable tactile-sensing skin for physical human-robot interaction. *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5694–5701, 2024. URL https://api.semanticscholar.org/CorpusID:269605928.

[17] T. Bhattacharjee, A. Jain, S. Vaish, M. D. Killpack, and C. C. Kemp. Tactile sensing over articulated joints with stretchable sensors. In *2013 World Haptics Conference (WHC)*, pages 103–108. IEEE, 2013.

[18] S. Stassi, V. Cauda, G. Canavese, and C. F. Pirri. Flexible tactile sensing based on piezoresistive composites: A review. *Sensors*, 14(3):5296–5332, 2014.

[19] W. Yuan, S. Dong, and E. H. Adelson. Gelsight: High-resolution robot tactile sensors for estimating geometry and force. *Sensors*, 17(12):2762, 2017.

[20] 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*, 5(3):3838–3845, 2020.

[21] J. Di, Z. Dugonjic, W. Fu, T. Wu, R. Mercado, K. Sawyer, V. R. Most, G. Kammerer, S. Speidel, R. E. Fan, G. Sonn, M. R. Cutkosky, M. Lambeta, and R. Calandra. Using fiber optic bundles to miniaturize vision-based tactile sensors, 2024. URL https://arxiv.org/abs/2403.05500.

[22] T. Hellebrekers, O. Kroemer, and C. Majidi. Soft magnetic skin for continuous deformation sensing. *Advanced Intelligent Systems*, 1(4):1900025, 2019.

[23] T. P. Tomo, M. Regoli, A. Schmitz, L. Natale, H. Kristanto, S. Somlor, L. Jamone, G. Metta, and S. Sugano. A new silicone structure for uskin—a soft, distributed, digital 3-axis skin sensor and its integration on the humanoid robot icub. *IEEE Robotics and Automation Letters*, 3(3): 2584–2591, 2018.

[24] Y. Yan, Z. Hu, Z. Yang, W. Yuan, C. Song, J. Pan, and Y. Shen. Soft magnetic skin for super-resolution tactile sensing with force self-decoupling. *Science Robotics*, 6, 2021. URL https://api.semanticscholar.org/CorpusID:232040651.

[25] Gelsight mini. https://www.gelsight.com/gelsightmini/, 2023.

[26] R. Bhirangi, A. DeFranco, J. Adkins, C. Majidi, A. Gupta, T. Hellebrekers, and V. Kumar. All the feels: A dexterous hand with large-area tactile sensing. *IEEE Robotics and Automation Letters*, 2023.

[27] S. Funabashi, G. Yan, A. Geier, A. Schmitz, T. Ogata, and S. Sugano. Morphology-specific convolutional neural networks for tactile object recognition with a multi-fingered hand. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 57–63. IEEE, 2019.

[28] I. Akkaya, M. Andrychowicz, M. Chociej, M. Litwin, B. McGrew, A. Petron, A. Paino, M. Plappert, G. Powell, R. Ribas, et al. Solving rubik's cube with a robot hand. *arXiv preprint arXiv:1910.07113*, 2019.

[29] H. Etukuru, N. Naka, Z. Hu, S. Lee, J. Mehu, A. Edsinger, C. Paxton, S. Chintala, L. Pinto, and N. M. M. Shafiullah. Robot utility models: General policies for zero-shot deployment in new environments. *arXiv preprint arXiv:2409.05865*, 2024.

[30] H. Li, Y. Zhang, J. Zhu, S. Wang, M. A. Lee, H. Xu, E. Adelson, L. Fei-Fei, R. Gao, and J. Wu. See, hear, and feel: Smart sensory fusion for robotic manipulation. In *6th Annual Conference on Robot Learning*, 2023.

[31] Y. Chen, A. Sipos, M. V. der Merwe, and N. Fazeli. Visuo-tactile transformers for manipulation, 2022. URL `https://arxiv.org/abs/2210.00121`.

[32] H. Qi, B. Yi, S. Suresh, M. Lambeta, Y. Ma, R. Calandra, and J. Malik. General in-hand object rotation with vision and touch. In *Conference on Robot Learning*, pages 2549–2564. PMLR, 2023.

[33] J. Yin, H. Qi, J. Malik, J. Pikul, M. Yim, and T. Hellebrekers. Learning in-hand translation using tactile skin with shear and normal force sensing. *arXiv preprint arXiv:2407.07885*, 2024.

[34] Y. Yuan, H. Che, Y. Qin, B. Huang, Z.-H. Yin, K.-W. Lee, Y. Wu, S.-C. Lim, and X. Wang. Robot synesthesia: In-hand manipulation with visuotactile sensing. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6558–6565. IEEE, 2024.

[35] 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*, pages 3988–3993. IEEE, 2014.

[36] S. Kim and A. 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.

[37] V. H. Sundaram, R. Bhirangi, M. E. Rentschler, A. Gupta, and T. Hellebrekers. Dragonclaw: A low-cost pneumatic gripper with integrated magnetic sensing. In *2023 IEEE International Conference on Soft Robotics (RoboSoft)*, pages 1–8. IEEE, 2023.

[38] R. Bhirangi, C. Wang, V. Pattabiraman, C. Majidi, A. Gupta, T. Hellebrekers, and L. Pinto. Hierarchical state space models for continuous sequence-to-sequence modeling. In *Forty-first International Conference on Machine Learning*, 2024.

[39] G. Falkovich. *Fluid Mechanics: A Short Course for Physicists*. Cambridge University Press, 2011.

[40] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Computation*, 9(8): 1735–1780, 1997.

[41] S. Haldar, Z. Peng, and L. Pinto. Baku: An efficient transformer for multi-task policy learning. *arXiv preprint arXiv:2406.07539*, 2024.

[42] 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*, pages 770–778, 2016.

[43] T. Z. Zhao, V. Kumar, S. Levine, and C. Finn. Learning fine-grained bimanual manipulation with low-cost hardware. *arXiv preprint arXiv:2304.13705*, 2023.

[44] C. Chi, Z. Xu, C. Pan, E. Cousineau, B. Burchfiel, S. Feng, R. Tedrake, and S. Song. Universal manipulation interface: In-the-wild robot teaching without in-the-wild robots. *arXiv preprint arXiv:2402.10329*, 2024.

[45] N. M. M. Shafiullah, A. Rai, H. Etukuru, Y. Liu, I. Misra, S. Chintala, and L. Pinto. On bringing robots home. *arXiv preprint arXiv:2311.16098*, 2023.

[46] S. Dong and A. Rodriguez. Tactile-based insertion for dense box-packing. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 7953–7960. IEEE, 2019.

[47] 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)*, pages 6437–6443. IEEE, 2021.

[48] A. Iyer, Z. Peng, Y. Dai, I. Guzey, S. Haldar, S. Chintala, and L. Pinto. Open teach: A versatile teleoperation system for robotic manipulation. *arXiv preprint arXiv:2403.07870*, 2024.

A Slip Detection Setup

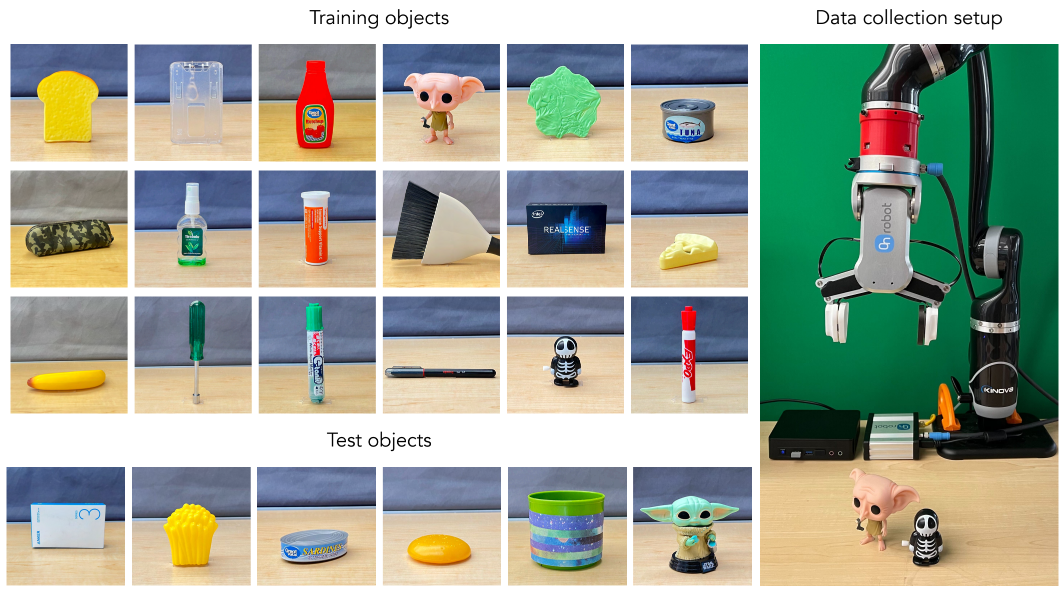


Figure 5: Experimental setup used for slip detection experiments, where we train LSTM models on data collected by a Jaco Robot equipped with the AnySkin sensor (right). We train on a set of training objects (left top) and evaluate it on a set of unseen test objects (left bottom).

For slip detection, we use a Kinova Jaco arm and an Onrobot RG-2 gripper with integrated AnySkin. An object held by a human operator is grasped and lifted up slowly for 1 second. We use a set of 40 daily objects – 30 for training and 10 evaluation – with varying shapes, weights and materials. Figure 5 shows the training objects for the slip detection. 6 objects were held out for testing and evaluating the learnt LSTM model for slip detection. We collect 6 trajectories for each object by changing the grasping force, width and location. After the data collection is complete, a human annotator labels the sequence as slip or no-slip from the corresponding videos. Objects used, along with the data collection setup can be found on our website. The data collection frequency for tactile data is 100 Hz. We subsample the signal by 15 along the temporal axis and take the first difference.

B Policy Learning Setup

Our policy learning setup consists of a UFactory xArm 7DOF robot and four cameras – three fixed to the frame and one egocentric camera attached to the robot's wrist, with AnySkin sensor on the gripper tip. A Meta Quest 3 and the accompanying joystick controller are used to teleoperate the robot using Open-Teach [48], an open-source teleoperation framework, and collect 96 demonstrations per task. For each task, learned policies are evaluated on locations of the target object unseen in training data.

B.1 Experiment Setup

We demonstrate the replaceability of AnySkin on a set of three contact-rich manipulation tasks shown in Fig. 3:

- **Plug insertion**: The robot starts with a plug grasped in the gripper. The task requires the robot to move to the first socket on the socket strip and insert the plug. The location of the socket strip is randomized in a 20 cm $\times$ 7 cm box on the table, and learned policies are tested on socket locations unseen in the training data.
- **Card swiping**: The robot starts with a credit card grasped in the gripper. It must move to the location of a credit card machine on the table and swipe the credit card. The location

of the credit card machine is randomized in a 40 cm $\times$ 15 cm box and angled in the range $(-30^\circ$ to $30^\circ)$ on the table, and learned policies are tested on card machine locations unseen in the training data.

- **USB insertion**: The robot starts with a USB cable grasped in the gripper. It must move to the location of the USB port on the table and insert the cable. The location of the USB port is randomized in a 20 cm $\times$ 15 cm box on the table, and learned policies are tested on port locations unseen in the training data.

To ensure that learned policies rely on *both* vision and tactile information, we vary the configuration of the target object, ie. the socket strip, the card machine and the USB port for plug insertion, card swiping and USB insertion respectively in the demonstration dataset. For all the evaluations presented here, we use a set of held-out configurations of the target object as shown in Fig. 6.

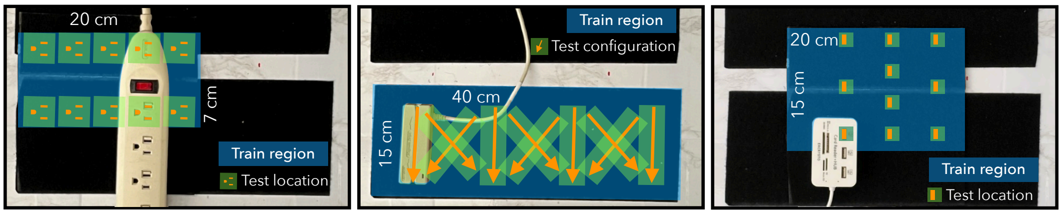


Figure 6: Training and test locations of the target objects interacted with for plug insertion, card swiping and USB insertion (left to right). The blue region represents the extent of variation in the location of the target object, while the green-orange blocks denote held-out test configurations used for evaluation.

B.2 Model Architecture and Training

Our policies are trained using behavior cloning. For each task presented in this section, we collect a set of 96 demonstration trajectories, with data from the four cameras in addition to unchanged instances of the respective tactile sensor(s) being used. The BAKU [41] architecture is used as the policy architecture. BAKU tokenizes each input using a modality-specific encoder: image inputs from cameras and DIGITs are encoded using ResNet-18 [42] encoders, while AnySkin and ReSkin inputs are encoded using an MLP encoder. An action token is appended to the set of encoded tokens before passing the sequence through a transformer encoder, and the output corresponding to the action token is used to predict actions. We use action chunking [43] and predict the next 10 actions at every timestep. For every training setting, we train three separate models corresponding to three different seeds, and present aggregated statistics on 10 policy rollouts.

B.3 Comparison across sensors

To better contextualize the significance of this result, we present a replaceability comparison with DIGIT [20] and ReSkin [11] sensors. We collect two additional datasets of 96 demonstration trajectories each for the plug insertion task with these sensors similar to AnySkin. Replaceability is evaluated by swapping the training skin for a new skin during evaluation as outlined in the previous section. Success rates from 10 evaluations across three seeds for each setting are reported in Table 2.

Based on these results, we find that visuotactile policies trained with ReSkin and AnySkin have similar performance on solving the plug insertion task. However, when the skin is replaced, the performance of the ReSkin policy falls 43% to the same level as the camera-only policy, while the performance of AnySkin policies only drops by 13%. This transferability is evidence of AnySkin's superior signal consistency, and is a significant boost to scaling efforts like training large tactile models as well as real world deployment of models trained in the laboratory.

An unexpected result from these experiments was the poor success rate of policies trained using the DIGIT sensor which has been shown to be successful in other visuotactile tasks, perhaps on less precise [32] or less reactive [12] ones, in prior work. Consequently, while there is still a gap in performance, we don't see a significant gap between the poor performance of the visuotactile policy on the original and swapped DIGIT skins. However, the high variability between instances of DIGIT sensors is previously documented [12], and we find that a closer look at the DIGIT signal from our plug insertion dataset indicates that even if it were possible to train more performant policies, they are unlikely to generalize between instances. Across the 96 demonstration trajectories from the plug insertion, we compute the *maximum* change in pixel values across channels and across trajectories, and compare it against the pixel-wise differences between the original and a swapped instance of the DIGIT sensor in Fig. 4. Since the *maximum* change in sensory measurement through the course of the interaction is comparable to the difference in signal between two instances, it is unlikely that policies trained on one sensor will generalize to new instances.