Learning Heterogeneous Tactile Representations with Graph Neural Networks for Dexterous Manipulation

Tai Yamada, Satoshi Funabashi, Steven Oh, Pranav Ponnivalavan, Tetsuya Ogata, and Shigeki Sugano

***Abstract*— Tactile feedback is essential for dexterous manipulation, but most learned tactile representations rely on a single sensing modality. We propose a heterogeneous tactile graph representation that unifies multimodal tactile observations across a robotic hand. Our system combines vision-based tactile sensors on the fingertips with uSkin magnetic tactile sensors on the palm, capturing both fine contact structure and distributed force information. These signals are structured as a graph aligned with the hand and processed using a graph convolutional network and masked autoencoder to produce a compact tactile embedding. Pre-trained on play data from 30 objects, the representation achieves near-perfect object classification across eight cup variants and shows structured latent transitions during teleoperated manipulation.**

I. Introduction

Humans rely heavily on tactile feedback during dexterous manipulation. The fingertips contain much higher densities of mechanoreceptors than the palm, enabling fine tactile discrimination, while the palm provides broader contact information during grasping [1]. Despite advances in perception and control, dexterous manipulation remains difficult for multi-fingered robot hands in unstructured environments. Although vision improves object recognition and pose estimation, it cannot directly capture contact-level signals such as force distribution, friction, or slip. Tactile sensing is therefore critical for robust manipulation.

Many tactile sensing modalities have been developed to capture different aspects of contact. Vision-based tactile sensors provide high spatial resolution and capture fine surface deformation and texture [2]–[9], while sparse taxel-based sensors—including piezoelectric [10], [11], capacitive [12], [13], and magnetic tactile sensors [14], [15]—are well suited for measuring force-related signals such as contact pressure and distributed forces. Learning effective tactile representations is therefore essential for contact-aware manipulation. Prior work, however, mainly focuses on a single sensing modality. Vision-based tactile sensing often uses pretrained encoders or self-supervised representations [16], and has also been fused with vision through contrastive or cross-attention methods [17]. In contrast, sparse taxel-based sensing commonly relies on canonical representations [18], 3D spatial encodings [19], or pretraining strategies for low-dimensional tactile signals [20], [21]. Methods that jointly model heterogeneous tactile modalities remain largely unexplored.

This research was supported by the Grant-in-Aid for Early-Career Scientists (KAKENHI) from the Japan Society for the Promotion of Science (JSPS), Grant No. 1167393.

All authors are with Waseda University, Okubo 3-4-1, Shinjuku, Tokyo 169-8555, Japan

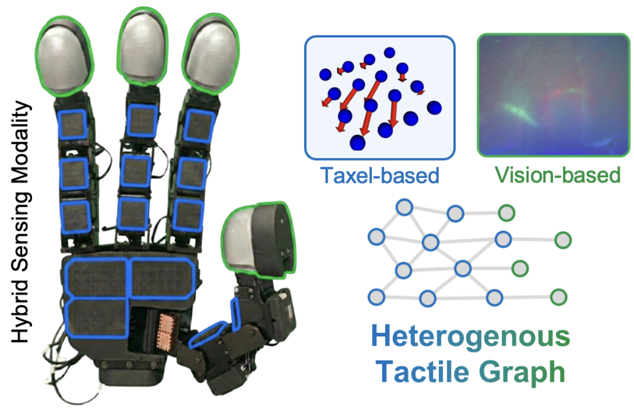

Fig. 1: Heterogeneous tactile graph representation combining taxel-based and vision-based tactile sensing.

Inspired by the heterogeneous sensing of the human hand, we propose a heterogeneous tactile graph representation that unifies tactile observations from multiple sensing modalities. Our system combines uSkin magnetic tactile sensors [14] on the palm with custom vision-based tactile sensors on the fingertips (See Appendix I). These multimodal observations are structured as a graph aligned with the physical layout of the robotic hand and processed with a graph convolution network to learn a unified tactile embedding. We evaluate the representation on object classification across multiple cup variants and analyze its behavior during teleoperated manipulation.

II. Heterogeneous Graph Representation

We model tactile sensing across the robotic hand as a heterogeneous graph that reflects both the spatial layout of the hand and the sensing modalities of the sensors. The system integrates two types of tactile sensing: high-resolution vision-based tactile sensors mounted on the fingertips and magnetic tactile sensors distributed across the palm. Each vision-based tactile sensor produces images with a resolution of 680×480. To reduce dimensionality while preserving spatial contact structure, each image is encoded by a convolutional neural network (CNN) into a 16×12 feature map. Each spatial location in this feature map is treated as a graph node, resulting in 192 nodes per fingertip sensor.

The Allegro hand is equipped with four fingertip sensors, producing $4 \times 192 = 768$ fingertip nodes. In addition, the palm contains 248 uSkin magnetic taxels, each measuring tri-axial force (f_x, f_y, f_z), and each taxel is treated as an additional graph node. Combining both modalities yields a heterogeneous tactile graph with $768 + 248 = 1016$

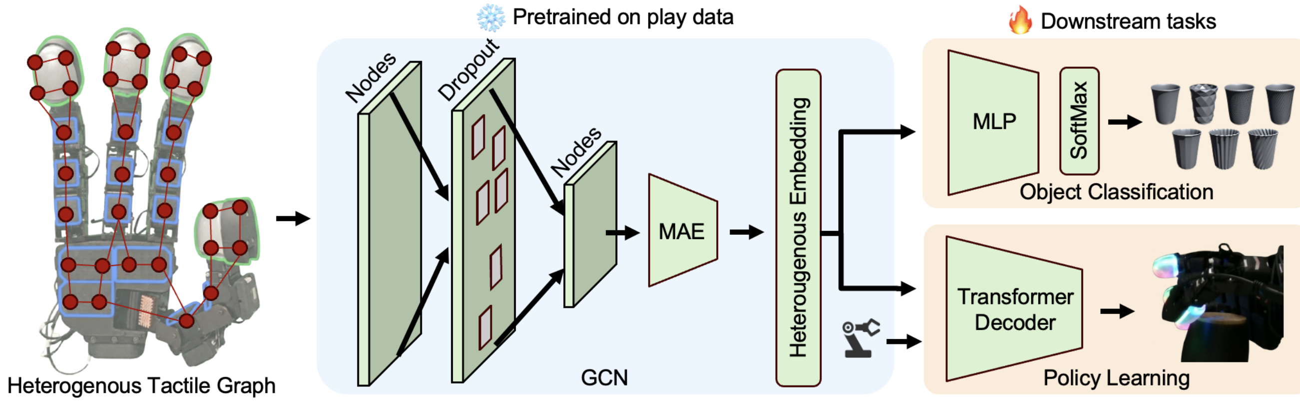

Fig. 2: Overview of the proposed heterogeneous tactile graph representation.

TABLE I: Object classification performance.

Class	Precision	Recall	F1-score
Original Cup	1.000	1.000	1.000
Standard Cup	0.988	0.995	0.992
Low-Poly Cup	0.999	1.000	0.999
Cup with Double Strip Deformation	1.000	1.000	1.000
Cup with Single Strip Deformation	1.000	0.952	0.975
Polygonal Cup	0.954	1.000	0.976
Wavy Cup	1.000	1.000	1.000
Wavy Twisted Cup	1.000	0.982	0.991

nodes. Edges are defined according to spatial proximity and the kinematic structure of the hand, allowing neighboring tactile elements to exchange information across both local contact regions and larger hand-level structure. The graph is processed by a graph convolutional network (GCN), and the resulting latent representation is further compressed using a masked autoencoder to form the final heterogeneous tactile embedding. This embedding is designed to capture both fine local contact structure from the fingertips and distributed force information from the palm. The encoder is pre-trained using play data collected from interactions with 30 objects, and during downstream tasks the encoder is frozen while task-specific heads are trained for applications such as object classification.

III. Experiments

A. Object Classification

We evaluate the learned representation on an object classification task using eight cup variants with different geometric deformations. During data collection, the robotic hand interacts with each object through grasping and manipulation motions, allowing both fingertip and palm sensors to capture contact signals. The frozen GCN encoder produces a tactile embedding that is used as input to a lightweight classification head. Table I reports the per-class precision, recall, and F1-score. The results show near-perfect performance across most object variants, suggesting that the heterogeneous tactile representation effectively captures geometric and contact-related differences between them. Future work will extend this evaluation to broader material understanding tasks involving texture and hardness.

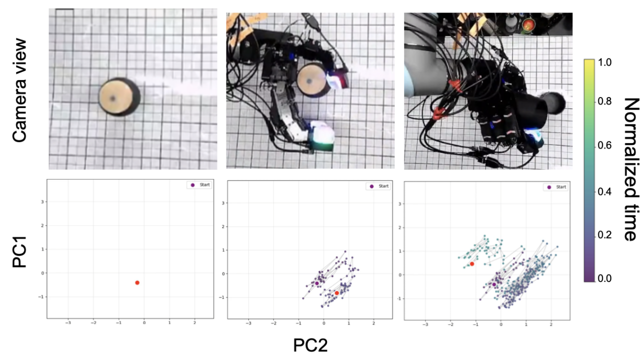

Fig. 3: Frames from a teleoperated manipulation sequence and the corresponding tactile latent space.

B. Latent Space Analysis During Manipulation

To analyze how the learned embedding evolves during manipulation, we examine teleoperated demonstrations collected using a motion capture glove for the Allegro hand and a downsized leader device for arm control [22] and real world robot setup described in Appendix II. The heterogeneous graph encoder is pre-trained on play data from 30 objects and frozen during evaluation. Figure 3 shows a PCA visualization of the tactile latent space during a manipulation sequence in which the hand grasps a cup, opens the lid using the thumb and index finger, and pours its contents. Before contact, the latent representation remains compact. When the palm first contacts the cup, the embedding begins to separate into more structured regions. As the fingertips engage in the lid-opening and pouring actions, the latent trajectory spreads further in the embedding space. These results suggest that the learned heterogeneous tactile representation captures meaningful contact transitions during manipulation by integrating distributed palm forces with fine fingertip contact structure.

IV. Conclusion

We propose a heterogeneous tactile graph representation integrating vision-based fingertip and magnetic palm sensing. Experiments on object classification and manipulation demonstrate its effectiveness. Future work will extend to texture and material classification and analyze its impact on policy learning.

References

[1] R. S. Johansson and A. B. Vallbo, "Tactile sensibility in the human hand: relative and absolute densities of four types of mechanoreceptive units in glabrous skin." *The Journal of Physiology*, vol. 286, no. 1, pp. 283–300, 1979.

[2] Q. Cong, S. Oh, W. Fan, S. Luo, K. Althoefer, and D. Zhang, "Taceva: A performance evaluation framework for vision-based tactile sensors," *Advanced Intelligent Systems*, p. e202501179, 2025.

[3] K. Vlack, T. Mizota, N. Kawakami, K. Kamiyama, H. Kajimoto, and S. Tachi, "Gelforce: a vision-based traction field computer interface," in *CHI'05 extended abstracts on Human factors in computing systems*, 2005, pp. 1154–1155.

[4] A. Yamaguchi, "Fingervision with whiskers: Light touch detection with vision-based tactile sensors," in *2021 Fifth IEEE International Conference on Robotic Computing (IRC)*. IEEE, 2021, pp. 56–64.

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

[6] A. Padmanabha, F. Ebert, S. Tian, R. Calandra, C. Finn, and S. Levine, "Omnitact: A multi-directional high-resolution touch sensor," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 618–624.

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

[8] W. Fan, H. Li, W. Si, S. Luo, N. Lepora, and D. Zhang, "Vitactip: Design and verification of a novel biomimetic physical vision-tactile fusion sensor," *arXiv preprint arXiv:2402.00199*, 2024.

[9] J. Xu, W. Chen, H. Qian, D. Wu, and R. Chen, "Thintact: Thin vision-based tactile sensor by lensless imaging," *IEEE Transactions on Robotics*, vol. 41, pp. 1139–1154, 2025.

[10] H. Pei, J. Jing, Y. Chen, J. Guo, and N. Chen, "3d printing of pvdf-based piezoelectric nanogenerator from programmable metamaterial design: Promising strategy for flexible electronic skin," *Nano Energy*, vol. 109, p. 108303, 2023.

[11] P. Yu, W. Liu, C. Gu, X. Cheng, and X. Fu, "Flexible piezoelectric tactile sensor array for dynamic three-axis force measurement," *Sensors*, vol. 16, no. 6, p. 819, 2016.

[12] A. Schmitz, M. Maggiali, L. Natale, B. Bonino, and G. Metta, "A tactile sensor for the fingertips of the humanoid robot icub," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2010, pp. 2212–2217.

[13] S. Somlor, R. Hartanto, A. Schmitz, and S. Sugano, "A novel tri-axial capacitive-type skin sensor," *Advanced Robotics*, vol. 29, no. 21, pp. 1375–1391, Nov. 2015, publisher Copyright: © 2015 © 2015 Taylor & Francis and The Robotics Society of Japan.

[14] 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*, vol. 3, no. 3, pp. 2584–2591, 2018.

[15] A. Alfadhel, M. A. Khan, S. Cardoso de Freitas, and J. Kosel, "Magnetic tactile sensor for braille reading," *IEEE Sensors Journal*, vol. 16, no. 24, pp. 8700–8705, 2016.

[16] C. Higuera, A. Sharma, C. K. Bodduluri, T. Fan, P. Lancaster, M. Kalakrishnan, M. Kaess, B. Boots, M. Lambeta, T. Wu, and M. Mukadam, "Sparsh: Self-supervised touch representations for vision-based tactile sensing," 2024.

[17] J. Li, T. Wu, J. Zhang, Z. Chen, H. Jin, M. Wu, Y. Shen, Y. Yang, and H. Dong, "Adaptive visuo-tactile fusion with predictive force attention for dexterous manipulation," in *2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2025, pp. 3232–3239.

[18] T. Wu, J. Li, J. Zhang, M. Wu, and H. Dong, "Canonical representation and force-based pretraining of 3d tactile for dexterous visuo-tactile policy learning," in *2025 IEEE International Conference on Robotics and Automation (ICRA)*, 2025, pp. 6786–6792.

[19] B. Huang, Y. Wang, X. Yang, Y. Luo, and Y. Li, "3d-vitac: Learning fine-grained manipulation with visuo-tactile sensing," in *8th Annual Conference on Robot Learning*.

[20] I. Guzey, B. Evans, S. Chintala, and L. Pinto, "Dexterity from touch: Self-supervised pre-training of tactile representations with robotic play," *arXiv preprint arXiv:2303.12076*, 2023.

[21] A. Sharma, C. Higuera, C. K. Bodduluri, Z. Liu, T. Fan, T. Hellebrekers, M. Lambeta, B. Boots, M. Kaess, T. Wu, F. R. Hogan, and M. Mukadam, "Self-supervised perception for tactile skin covered dexterous hands," in *9th Annual Conference on Robot Learning*, 2025.

[22] K. Shaw, Y. Li, J. Yang, M. K. Srirama, R. Liu, H. Xiong, R. Mendonca, and D. Pathak, "Bimanual dexterity for complex tasks," in *8th Annual Conference on Robot Learning*, 2024.

APPENDIX I. VISION-BASED TACTILE SENSOR FABRICATION

The vision-based tactile sensor mounted on the fingertip was fabricated using readily available materials. The sensor consists of a soft contact surface, an internal marker layer, and an embedded camera for capturing deformation patterns during contact. This design enables high-resolution tactile observation at the fingertip.

Details of the sensor structure are shown in Fig. 4.

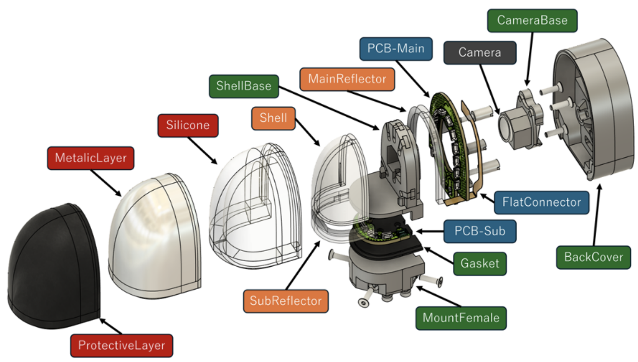

Fig. 4: Fabrication breakdown of customized vision-based tactile sensor.

APPENDIX II. SYSTEM SETUP

The overall experimental setup is shown in Fig. 5. The system consists of UR5e robot arm and Allegro hand equipped with vision-based tactile sensors on the finger tip, and taxel-based uSkin tactile sensor on the palm.

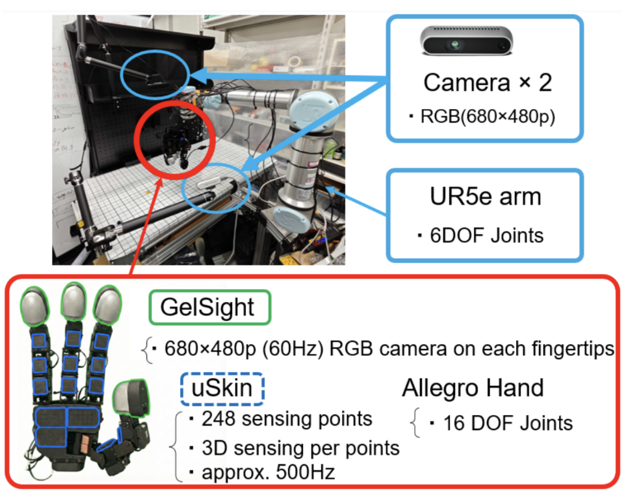

Fig. 5: Experimental system setup.