TACTX: Learning Shared Tactile Representations Across Diverse Sensors

Anonymous Author(s)
Affiliation
Address
email

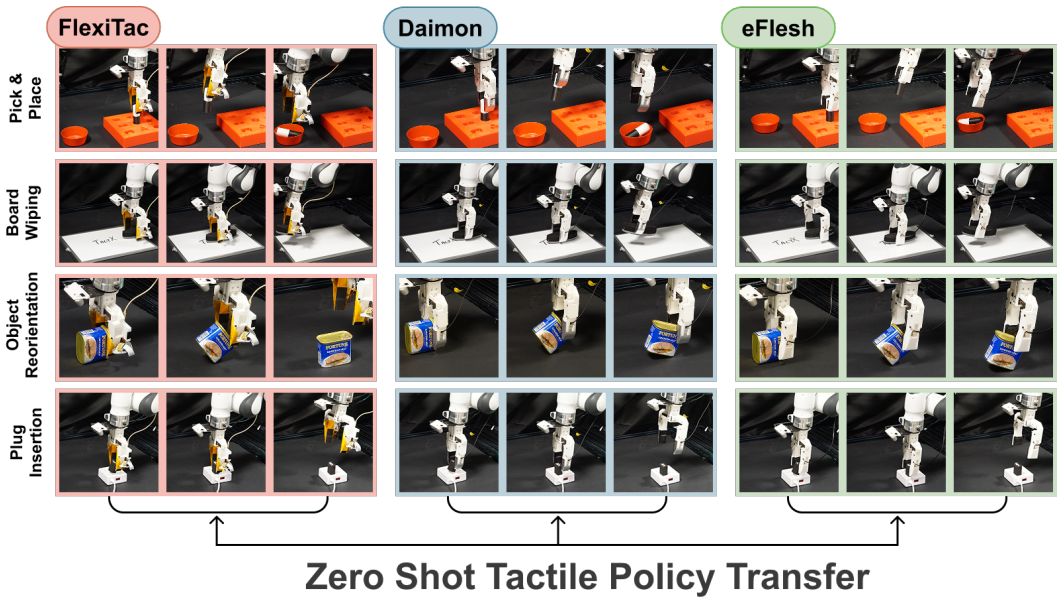


Figure 1: TACTX learns a shared latent representation that aligns heterogeneous tactile sensors and enables zero-shot transfer of tactile-conditioned policies.

Abstract: Tactile sensors provide critical information for contact-rich manipulation, yet tactile representations and policies remain tightly coupled to each specific sensor, limiting transferability across robots and hardware platforms. We propose TACTX, a framework for learning a transferable tactile representation across sensors spanning three fundamentally different transduction modalities: resistive, magnetic, and vision-based. TACTX maps heterogeneous tactile observations into a shared latent space through modality-specific encoders trained on paired contact data. Such paired interactions provide a natural alignment signal across modalities, and the encoders are jointly trained across all sensor pairs, inducing a consistent latent space for all sensor types. Our experiments show that TACTX aligns tactile representations across sensors while preserving object-level contact information, as evidenced by sensor-identity prediction and object classification in the learned latent space. We evaluate TACTX on four contact-rich manipulation tasks—pick-and-place, plug insertion, board wiping, and object reorientation—and show that policies trained with one sensor transfer zero-shot to physically distinct sensors through the shared latent. This improves the average success rate from 27.5% for vision-only policy to 45.9%, providing a step toward sensor-agnostic tactile manipulation.

Keywords: Tactile Sensing, Cross-Sensor Transfer, Robot Manipulation

Homepage: https://tactx-corl.github.io/

Submitted to the 10th Conference on Robot Learning (CoRL 2026). Do not distribute.

1 Introduction

Contact-rich manipulation depends on information that vision alone cannot provide. Tactile sensing captures forces, slips, and contact geometries at the robot–world interface, and prior work has shown that it is essential for robust manipulation in cluttered, occluded, or precision-critical settings [1, 2, 3, 4]. However, sensors differ not only in form factor, but also in how they physically measure contact, including through optical deformation, magnetic-field changes, and resistive pressure responses. This diversity has enabled specialized capabilities, but it also makes tactile representations and policies highly hardware-dependent. A policy trained with one tactile sensor is usually tied to that sensor's observation space, and replacing the sensor often requires collecting new demonstrations and retraining the downstream policy.

Cross-sensor tactile learning aims to reduce this dependence, but existing work has largely focused on transferring between vision-based tactile sensors [5, 6, 7, 8, 9]. Calibration or direct transfer across sensors is often insufficient because the same contact can produce substantially different signal distributions on different hardware. The more general question is how tactile representations can be shared across sensors that measure contact through different physical modalities, such as vision-based tactile images [10, 11, 12, 13], magnetic fields [14, 15, 16], and resistive pressure maps [17, 18].

To address this challenge, we propose TACTX, a framework for learning sensor-agnostic tactile representations across heterogeneous tactile sensors. TACTX learns from paired contact data collected using a gripper with a different tactile sensor mounted on each finger. Each grasp produces paired observations, where the two sensors measure the same contact point. Since multiple sensors cannot be mounted simultaneously, we collect data for each sensor pair and train the encoders jointly across pairs, inducing a globally consistent latent space from pairwise supervision. This pairwise formulation also provides a natural extension for incorporating additional sensors, since new modalities can be connected to the shared space through paired contact data.

TACTX combines contrastive alignment with self- and cross-reconstruction: contrastive learning pulls paired contacts from different sensors together in latent space [5, 8, 19], while reconstruction encourages the latent to preserve object- and contact-level structure. We evaluate this shared representation through both representation-level analyses and zero-shot policy-transfer experiments. Our results show that TACTX aligns three heterogeneous tactile sensors into a common latent space while supporting tactile-conditioned robot policies [20] trained with one sensor and deployed zero-shot with another across four contact-rich manipulation tasks: pick-and-place, plug insertion, board wiping, and object reorientation.

Our contributions are as follows. First, we present TACTX, a framework for learning shared tactile representations across fundamentally different sensing modalities—vision-based, magnetic, and resistive—going beyond prior cross-sensor settings that focus on vision-based tactile sensors. Second, we introduce a pairwise training strategy that uses paired contact data to align all sensor pairs into a globally consistent latent space. Third, we demonstrate the practical utility of this shared latent space for robotic manipulation, showing that tactile-conditioned policies trained with one sensor can be deployed on physically distinct sensors without retraining, improving over vision-only transfer baselines by approximately 20%.

2 Related Work

Contact-Rich Manipulation. Tactile feedback has been shown to improve robotic manipulation across a wide range of contact-rich tasks: peg-in-hole and electronics insertion [3], in-hand reorientation [21] and dexterous manipulation [2, 22, 23, 24], sliding, wiping, and pivoting [4, 25], cable routing [26], and grasping under clutter or occlusion [1, 27]. These results collectively establish that tactile sensing provides information that vision alone cannot, particularly for sub-millimeter precision and contact-state estimation under occlusion.

Tactile Sensors and Representations. Modern tactile sensors span a wide range of transduction principles, including vision-based [10, 11, 13, 28, 29, 30, 31], magnetic [14, 15, 16, 32], resistive and capacitive [17, 18], and piezoelectric or acoustic [33, 34, 35] designs. Recent work has produced strong per-sensor representations, ranging from sensor-specific encoders trained end-to-end [1, 36] to large-scale self-supervised pretraining [37, 38, 39]. Comparative benchmarks [8, 40] have shown that the usefulness of tactile information depends strongly on sensor modality, material properties, and the task, motivating representations that generalize across sensors rather than being tied to any single device.

Cross-Embodiment and Cross-Sensor Representations. A growing line of work seeks unified representations across heterogeneous hardware. In cross-embodiment robot learning, shared *latent action spaces* enable a single policy to drive multiple dexterous hands or arms [41, 42, 43, 44, 45], often via per-embodiment encoder–decoder pairs that map into a common space. In cross-sensor tactile learning, prior work has primarily focused on aligning sensors within a shared sensing substrate—typically the vision-based family, where signals share a common image-like representation [5, 6, 7, 8, 46, 47, 48]. A smaller body of work addresses sensors with no shared substrate, either by mapping signals to a unified input format [49, 50, 51, 52] or by aligning distributions without paired data [53]. TACTX extends this direction to three transduction modalities simultaneously—resistive, magnetic, and vision-based—and aligns them in latent space without any per-sensor input transformation.

3 Methodology

Our goal is to align tactile observations from fundamentally different sensing modalities into a shared representation. This is challenging because **vision-based**, **magnetic**, and **resistive** sensors produce observations with different structures, dimensionalities, and sampling rates. TACTX addresses this by using modality-specific encoders to map each observation into a shared latent space $\mathcal{Z}$, contrastive learning to align paired contacts across sensors, and reconstruction to preserve contact-relevant information. An overview of the data collection, encoder–decoder architecture, and training losses is shown in Figure 2.

Data collection. To construct positive pairs across sensors, TACTX collects paired contact observations by mounting different tactile sensors on the same gripper. For each sensor pair $S_i, S_j \in \mathcal{S}$, we record quasi-static grasps of rigid symmetric objects, where each grasp provides one measurement from each sensor for the same physical contact. The resulting temporally aligned pairs $(x_i^{(t)}, x_j^{(t)})$ over diverse contact points and 10 indentors form the pair dataset $\mathcal{D}_{ij} = (x_i^{(t)}, x_j^{(t)})$. We collect such paired datasets across all sensor pairs, which jointly supervise the sensor-specific encoders. Details on sensor instantiations, hardware, objects, and protocol are provided in Appendix A.

Architecture. For each sensor $i \in \mathcal{S}$, an encoder f_i maps its native signal $x_i \in \mathcal{X}_i$ through a signal-specific backbone and projection head, which outputs the parameters of a posterior $q_i(z \mid x_i) = \mathcal{N}(\mu_i(x_i), \mathrm{diag}(\sigma_i^2(x_i)))$ over the shared latent $z \in \mathcal{Z} \subset R^{16}$. The low-dimensional latent encourages the encoders to learn shared contact features while compressing sensor-specific detail. Each sensor has its own decoder $g_i : \mathcal{Z} \to \mathcal{X}_i$, used for both self- and cross-reconstruction. All encoders are trained from scratch such that modalities start from equal footing. Per-sensor backbones and decoder architectures are given in Appendix B.

Forward pass. A single training example is one left–right pair (x_i, x_j) from $\mathcal{D}_{ij}$: the signal x_i from the sensor on one finger and x_j from the sensor on the other, both observing the same contact. The two encoders produce posteriors $q_i(z_i \mid x_i)$ and $q_j(z_j \mid x_j)$; we align the posterior means μ_i, μ_j in the shared latent (they should coincide, since both describe the same contact), and sample $z_i \sim q_i$, $z_j \sim q_j$ via the reparameterization trick for reconstruction. Each sampled latent is decoded both by its *own* sensor's decoder (*self*-reconstruction, e.g. $g_i(z_i) \to x_i$) and by the *paired* sensor's decoder

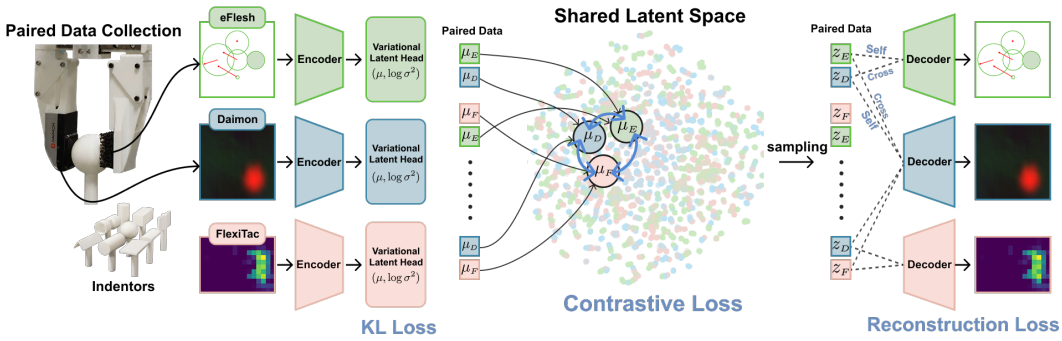


Figure 2: **TACTX** trains on paired contacts from two sensors at a time. Paired observations are encoded into a shared latent space, aligned with InfoNCE, and decoded through self- and cross-reconstruction. Other pairs are trained analogously, yielding a single latent space shared by all three sensors.

(*cross*-reconstruction, e.g. $g_j(z_i) \rightarrow x_j$): the latent from one finger must reconstruct the other finger's ground-truth signal. At inference, we use the posterior mean $z = \mu_i(x_i)$ as a deterministic latent representation so that the downstream policy receives a stable input.

Training objective. Each step jointly optimizes three terms over every sensor pair (i, j):

$$\mathcal{L}_{\text{TACTX}} = \sum_{(i,j)} \Big[\lambda_{\text{recon}} \mathcal{L}_{\text{recon}}^{(i,j)} + \alpha(t) \mathcal{L}_{\text{align}}^{(i,j)} + \beta(t) \mathcal{L}_{\text{KL}}^{(i,j)} \Big]. \tag{1}$$

The *reconstruction* term combines the self- and cross-reconstruction flows described above,

$$\mathcal{L}_{\text{recon}}^{(i,j)} = \underbrace{\|g_i(z_i) - x_i\|_1 + \|g_j(z_j) - x_j\|_1}_{\text{self}} + \underbrace{\|g_i(z_j) - x_i\|_1 + \|g_j(z_i) - x_j\|_1}_{\text{cross}}, \tag{2}$$

where each term is mean-reduced over its target; cross-reconstruction forces shared content through the latent and ties the modalities together.

The *alignment* term is a symmetric NT-Xent loss [54] on L2-normalized posterior means $\tilde{\mu}_i = \mu_i / \|\mu_i\|_2$ with temperature τ=0.01. For a batch of N paired contacts, the two posterior means from the same contact form a positive pair, while the remaining embeddings in the batch serve as negatives.

$$\mathcal{L}_{\text{align}}^{(i,j)} = -\frac{1}{N} \sum_{n=1}^{N} \log \frac{\exp(\tilde{\mu}_i^{(n)} \cdot \tilde{\mu}_j^{(n)} / \tau)}{\sum_{m=1}^{N} \exp(\tilde{\mu}_i^{(n)} \cdot \tilde{\mu}_j^{(m)} / \tau)}. \tag{3}$$

The *KL* term regularizes each posterior toward a shared prior $\mathcal{N}(0, I)$, giving all modalities a common target region. We use λ_{recon}=1, $\alpha(t)$ optionally ramped via a reconstruction-first curriculum, and $\beta(t)$ warmed up from 0 to $\beta_{\max}$=0.1 over the first 30 epochs. Full schedules are in Appendix C.

Pairwise joint training. Each trajectory contains only two sensors, so each training step samples from each available pair dataset $\mathcal{D}_{ij}$ and optimizes the resulting paired batches jointly. This exposes every sensor encoder to paired supervision at each step, while the shared latent space and common prior tie the pairwise alignments into a globally consistent representation.

4 Experimental Evaluation

TACTX is designed to align tactile sensors across different sensing modalities while preserving the contact information needed for manipulation. While prior cross-sensor tactile methods largely evaluate transfer within sensors that share a similar sensing substrate, we evaluate TACTX across Daimon (vision-based) [29], eFlesh (magnetic) [16], and FlexiTac (resistive) [17] tactile sensors whose raw observations differ in geometry, dimensionality, and physical sensing mechanisms. We evaluate the learned latent space by testing whether it is sensor-invariant, jointly aligned across all

three sensors from pairwise supervision, and still preserves tactile content through object-level prediction and self- and cross-reconstruction. Finally, we test whether the shared latent can serve as a sensor-agnostic tactile interface for Action Chunking with Transformers (ACT) policies [20], allowing a policy trained with one tactile sensor to be deployed on another sensor without retraining. Our experiments explore cross-sensor alignment (4.1), three-way alignment from pairwise data (4.2), tactile content preservation (4.3), and zero-shot policy transfer through the shared latent (4.4).

4.1 How well do different sensors align in a shared representation space?

We first evaluate whether TACTX aligns tactile observations from different sensing modalities into a shared latent space. We compare TACTX with three objective variants: a reconstruction-only model (using Eq. (2)), a contrastive-only model (using Eq. (3)), and an L2-alignment model that replaces the contrastive objective with a direct pull between paired latents. We quantify alignment by computing the cosine similarity between positive contact pairs, and evaluate sensor invariance with a linear sensor-identity probe on the frozen latent.

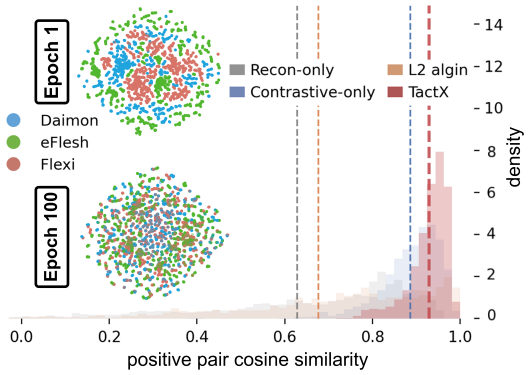


Figure 3: **Transitive cross-sensor alignment.** Cosine similarity along the Daimon→eFlesh→FlexiTac path measures global latent alignment, with dashed lines indicating the mean for each method.

The results in Figure 4 show that TACTX consistently aligns paired contacts across all sensor pairs. As expected, contrastive-only training gives high positive-pair similarity, since it is optimized solely to pull paired contacts together. TACTX achieves comparable alignment while also retaining the reconstruction objective, which is important for preserving tactile content as evaluated in Section 4.3.

The t-SNE visualization shows the same trend qualitatively. Before training, samples from different sensors occupy separate regions, reflecting the gap between their raw sensing modalities. After training, the three sensor domains become substantially more mixed. The sensor-prediction probe further supports this observation. We train this probe as a linear classifier on the learned latent space with the encoders frozen, so only the classifier is trained to predict sensor identity. Because lower sensor-prediction accuracy indicates stronger sensor invariance, the ideal representation should approach the 33.3% chance level. TACTX reduces sensor prediction accuracy from 67.5% for reconstruction-only training to 47.5%, the closest to chance among the reconstruction-based variants. These results indicate that TACTX learns a shared latent space in which heterogeneous tactile sensors are well aligned.

4.2 Can pairwise data align multiple sensors?

We next ask whether pairwise supervision is sufficient to produce a globally consistent tactile space of multiple sensors. Our data are collected from two sensors at a time, so the model never observes optical, magnetic, and resistive tactile readings from the same contact simultaneously. This makes three-way alignment nontrivial: the model must align each observed pair while placing all three modalities into a single shared coordinate system.

We evaluate this property using transitive alignment between Daimon and FlexiTac through eFlesh. We denote Daimon, eFlesh, and FlexiTac as D, E, and F, respectively. Each positive pair contains two observations of the same contact, one from each sensor. The model is trained on D–E and E–F positive pairs, but does not observe D–F contacts jointly in this test; therefore, successful D–F alignment requires consistency through the shared E bridge. For each E–F pair, we find the nearest E latent in the D–E set, take its paired D latent, and measure the cosine similarity between this matched D latent and the F latent. Figure 3 shows that TACTX achieves the strongest transi-

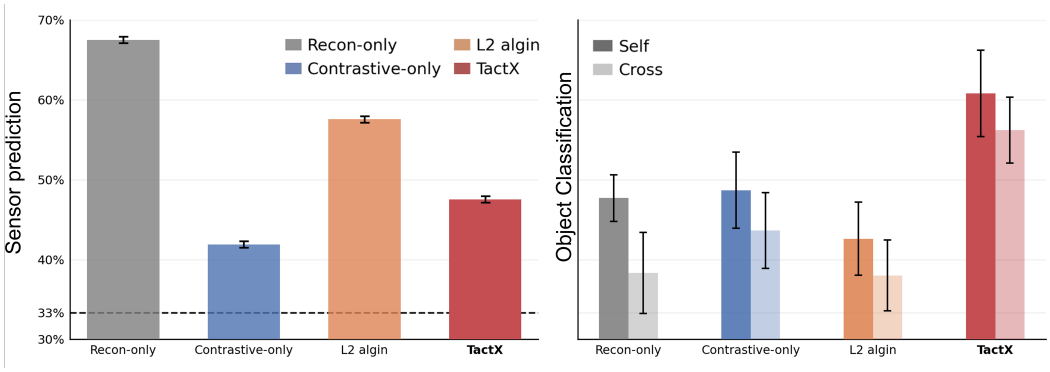


Figure 4: **Sensor invariance and semantic preservation in the shared latent space.** Sensor-prediction accuracy measures whether sensor identity remains recoverable from frozen latents, where lower values closer to the 33.3% chance level indicate stronger sensor invariance. Object-classification accuracy evaluates whether object-level information is preserved, where "Self" denotes training and testing on the same sensor and "Cross" denotes training on one sensor and testing on aligned latents from the other sensors.

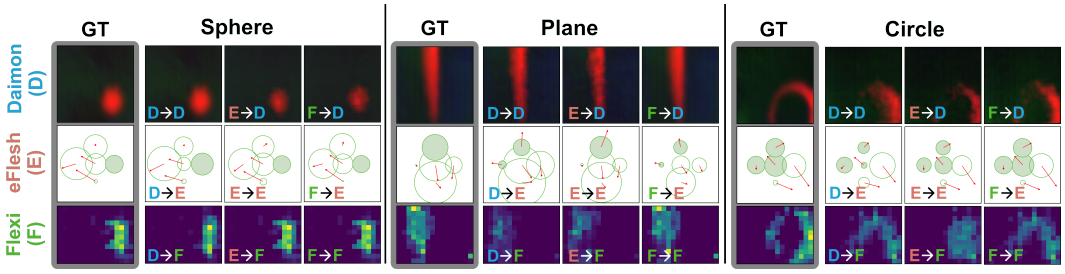


Figure 5: **Self- and cross-reconstruction from the shared latent.** We visualize representative validation contacts from sphere, plane, and circle indentors. For each sensor, the first column is the ground-truth observation, the diagonal entries are self-reconstructions, and the off-diagonal entries are cross-reconstructions decoded from the nearest latent representations of the other sensors in the validation set.

tive alignment, increasing the D–F cosine from 0.626 with reconstruction-only training and 0.679 with L2-alignment to 0.928. This indicates that the model does not simply learn independent pairwise mappings, but instead induces a globally consistent shared latent space across multiple tactile modalities from pairwise supervision.

4.3 Does the shared latent preserve critical tactile information?

We next evaluate whether the shared latent space preserves tactile content, rather than only removing sensor identity. This is important because a fully sensor-invariant representation is useful for manipulation only if it still retains information about the underlying contact geometry. We examine this property through reconstruction from the latent space. For each target sample in the validation set, we decode the sensor's own latent for self-reconstruction. For cross-reconstruction, we do not use the paired observation directly; instead, we retrieve the nearest latent from the other sensor in the validation set and decode it into the target modality.

The results in Figure 5 show that TACTX preserves the dominant contact patterns across all three sensing modalities. Self-reconstructions recover the original contact structure, and cross-reconstructions from nearest aligned latents remain visually consistent with the corresponding ground truth. This indicates that the shared latent does not merely discard sensor-specific information, but retains object- and contact-level structure that can be decoded across modalities.

We further evaluate tactile content preservation by training object classifiers on frozen latent features and testing them on unseen contact points from 10 object classes. The results in Figure 4 show that TACTX achieves the highest accuracy in both settings, reaching 60.8% for self-sensor evaluation

Table 2: **Cross-sensor policy transfer across all tasks.** Each entry reports the number of successes out of 10 trials, shown as mean ± std over 3 runs. **Bold** indicates the best performance for each source-deploy combination and task.

Method	Source	Deploy	P&P	P&P (OOD)	Insertion	Wiping	Reorient
Vision Transfer	Daimon	eFlesh	5.3 ± 0.9	**1.7 ± 0.9**	2.7 ± 0.5	3.0 ± 0.0	0.3 ± 0.5
		FlexiTac	5.3 ± 0.5	1.3 ± 0.5	1.3 ± 0.9	1.3 ± 0.9	0.7 ± 0.9
	eFlesh	Daimon	7.7 ± 0.5	0.0 ± 0.0	3.7 ± 0.5	0.0 ± 0.0	**7.0 ± 0.0**
		FlexiTac	1.3 ± 0.5	0.7 ± 0.5	0.3 ± 0.5	0.3 ± 0.5	1.3 ± 0.5
	FlexiTac	Daimon	7.0 ± 0.8	6.0 ± 0.0	3.3 ± 0.5	0.7 ± 0.5	7.3 ± 0.9
		eFlesh	6.7 ± 0.5	1.0 ± 0.0	5.0 ± 0.0	0.3 ± 0.5	0.0 ± 0.0
Binary Contact Transfer	Daimon	eFlesh	4.0 ± 0.0	1.3 ± 0.9	2.7 ± 2.1	2.0 ± 2.8	1.7 ± 1.7
		FlexiTac	3.3 ± 1.2	**2.7 ± 0.5**	**2.0 ± 0.8**	**2.0 ± 1.4**	4.0 ± 1.6
	eFlesh	Daimon	3.3 ± 0.9	0.0 ± 0.0	0.3 ± 0.5	1.0 ± 1.4	2.3 ± 2.1
		FlexiTac	**4.0 ± 1.6**	**1.3 ± 1.2**	0.7 ± 0.9	0.3 ± 0.5	**6.0 ± 2.2**
	FlexiTac	Daimon	3.3 ± 1.7	2.7 ± 1.7	6.7 ± 0.5	1.3 ± 1.9	7.3 ± 2.4
		eFlesh	1.0 ± 0.8	2.7 ± 1.9	0.3 ± 0.5	1.0 ± 0.8	2.7 ± 0.9
TactX Transfer (Ours)	Daimon	eFlesh	**8.3 ± 0.5**	1.0 ± 0.0	**4.0 ± 0.8**	**4.0 ± 0.0**	**3.7 ± 0.9**
		FlexiTac	**5.3 ± 1.2**	2.0 ± 0.8	1.3 ± 1.9	0.3 ± 0.5	**6.7 ± 0.9**
	eFlesh	Daimon	**9.0 ± 1.4**	**0.7 ± 0.5**	**6.0 ± 1.6**	**6.0 ± 0.8**	5.0 ± 1.4
		FlexiTac	**1.3 ± 0.9**	0.0 ± 0.0	**3.7 ± 1.2**	**1.3 ± 0.5**	5.0 ± 0.8
	FlexiTac	Daimon	**8.0 ± 1.4**	**8.3 ± 1.2**	**8.3 ± 0.9**	**6.3 ± 0.9**	**7.7 ± 1.2**
		eFlesh	**6.7 ± 0.9**	**3.0 ± 0.8**	**4.7 ± 0.5**	**5.7 ± 0.5**	**4.3 ± 0.5**

and 56.2% for cross-sensor evaluation. This indicates that the shared latent preserves object-level tactile structure even when evaluated across sensors.

4.4 Can robot policies transfer across sensors through the shared latent?

We finally evaluate whether the shared latent space can support tactile-conditioned policy learning across physically different sensors. We consider four contact-rich manipulation tasks: pick-and-place, plug insertion, board wiping, and object reorientation. For pick-and-place, we additionally evaluate an out-of-distribution color setting, where policies trained on black objects are tested on white objects with the same geometry. For each task, we train an ACT policy [20] using demonstrations from one tactile sensor and evaluate it under a different tactile sensor.

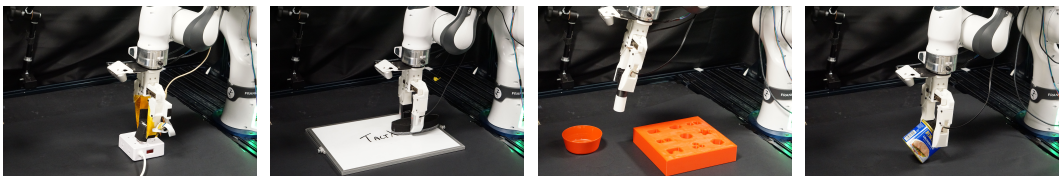
Table 1 summarizes the policy results. In the in-domain setting, tactile input improves over vision-only policies, confirming that contact information is useful for these tasks. Raw tactile observations provide a strong same-sensor performance, and we treat this as an oracle upper bound: the goal of cross-sensor transfer is to *approach* this value. TACTX retains most of this benefit using a shared latent representation in place of sensor-specific raw inputs.

Table 1: **Average same-sensor policy performance.** Results are averaged over three same-sensor train-test evaluations.

Task	Vision	+ Tactile GT	+ TACTX
P&P	8.33	9.33	**10.00**
P&P (OOD)	6.67	**8.00**	7.33
Insertion	4.00	**7.33**	6.00
Wiping	4.33	**8.33**	7.33
Reorientation	8.00	9.33	**9.67**

The main result is cross-sensor transfer, shown in Table 2. Vision-only policies provide a sensor-independent baseline but degrade on contact-rich or visually shifted settings. Binary contact transfer tests whether a minimal contact/no-contact signal is sufficient, but it removes spatial and geometric contact information. In contrast, TACTX achieves the best cross-sensor performance on most tasks and sensor-pairs when transferring between sensors zero-shot. Averaged over all transfer directions and tasks, TACTX improves the success rate from 27.5% to 45.9% over vision-only transfer.

Contact-rich tasks such as board wiping and object reorientation show the largest improvements over the vision-only baseline. Binary contact transfer provides only limited benefit in these settings, suggesting that the shared latent captures richer contact geometry than a simple contact/no-contact signal and recovers failures that vision alone cannot resolve. A further limitation of binary contact



(a) Plug insertion (b) Board wiping (c) Pick-and-place (d) Object reorientation

Figure 6: **Downstream manipulation tasks.** We evaluate zero-shot tactile policy transfer on four contact-rich tasks: plug insertion, board wiping, pick-and-place, and object reorientation.

transfer is its sensitivity to the contact threshold: we use three separate sensor-specific thresholds that are held fixed across all tasks (Appendix D.3), and this threshold mismatch between tasks leads to higher variance and inconsistent results across task conditions.

We also observe that transfer is not symmetric across sensors. The weakest direction is eFlesh to FlexiTac, where all methods perform poorly. This suggests that a policy trained with the lower-dimensional magnetic signal may not learn to use the finer spatial structure available from the resistive sensor at deployment. The reverse is stronger, indicating that policies trained with a richer tactile representation perform more gracefully when deployed with a lower-bandwidth sensor. Overall, these results suggest that TACTX does not simply improve latent metrics, but provides a practical shared representation for zero-shot tactile policy transfer across heterogeneous sensors.

5 Conclusion

We introduced TACTX, a framework for learning sensor-agnostic tactile representations for contact-rich manipulation. TACTX aligns heterogeneous tactile observations from vision-based, magnetic, and resistive sensors into a shared latent space using paired contact data, contrastive alignment, and reconstruction. By addressing sensors with fundamentally different tactile transduction modalities, TACTX goes beyond transfer among visually similar tactile sensors and enables a common representation for structurally diverse touch signals. This shared representation enables zero-shot cross-sensor policy transfer, allowing policies trained with one tactile sensor to be deployed with a physically distinct sensor without retraining. Across multiple manipulation tasks, TACTX improves zero-shot cross-sensor policy transfer over vision-only baselines, demonstrating that aligned tactile representations can help policies generalize across heterogeneous tactile hardware. These results suggest a scalable path toward shared tactile representations that reduce the need to collect new demonstrations and retrain policies for each new sensor.

6 Limitations

While our results demonstrate the feasibility of zero-shot cross-sensor tactile policy transfer, TACTX still has several limitations. First, our method relies on paired contact data to align heterogeneous tactile sensors, which requires different sensors to observe corresponding contact events under comparable object poses, contact locations, and interaction conditions. This assumption may be more difficult to satisfy for asymmetric or geometrically complex objects, where two sensors can produce different tactile readings due to differences in placement or local contact geometry rather than sensing modality alone. Second, our current data is collected primarily from quasi-static gripping interactions, which provide clean alignment supervision but do not fully capture the dynamic contact variations that arise during manipulation. For example, although TACTX transfers effectively on board wiping overall, failures can occur under large shear changes or sustained sliding contact. Future work will address these limitations by exploring weaker forms of supervision, such as weakly paired, unpaired, or self-supervised alignment, and by extending data collection to dynamic tactile interactions such as sliding, pushing, and other contact-rich motions.

References

[1] R. Calandra, A. Owens, D. Jayaraman, J. Lin, W. Yuan, J. Malik, E. H. Adelson, and S. Levine. More than a feeling: Learning to grasp and regrasp using vision and touch. *IEEE Robotics and Automation Letters*, 3(4):3300–3307, Oct. 2018. ISSN 2377-3774. doi:10.1109/lra.2018.2852779. URL http://dx.doi.org/10.1109/LRA.2018.2852779.

[2] Z.-H. Yin, B. Huang, Y. Qin, Q. Chen, and X. Wang. Rotating without seeing: Towards in-hand dexterity through touch, 2023. URL https://arxiv.org/abs/2303.10880.

[3] D. Palenicek, T. Gruner, T. Schneider, A. Böhm, J. Lenz, I. Pfenning, E. Krämer, and J. Peters. Learning tactile insertion in the real world, 2024. URL https://arxiv.org/abs/2405.00383.

[4] M. Oller, D. Berenson, and N. Fazeli. Tactile-driven non-prehensile object manipulation via extrinsic contact mode control, 2024. URL https://arxiv.org/abs/2405.18214.

[5] F. Yang, C. Feng, Z. Chen, H. Park, D. Wang, Y. Dou, Z. Zeng, X. Chen, R. Gangopadhyay, A. Owens, and A. Wong. Binding touch to everything: Learning unified multimodal tactile representations, 2024. URL https://arxiv.org/abs/2401.18084.

[6] J. Zhao, Y. Ma, L. Wang, and E. H. Adelson. Transferable tactile transformers for representation learning across diverse sensors and tasks, 2024. URL https://arxiv.org/abs/2406.13640.

[7] R. Feng, J. Hu, W. Xia, T. Gao, A. Shen, Y. Sun, B. Fang, and D. Hu. Anytouch: Learning unified static-dynamic representation across multiple visuo-tactile sensors, 2025. URL https://arxiv.org/abs/2502.12191.

[8] 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. URL https://arxiv.org/abs/2410.24090.

[9] S. Rodriguez, Y. Dou, M. Oller, A. Owens, and N. Fazeli. Cross-sensor touch generation, 2025. URL https://arxiv.org/abs/2510.09817.

[10] W. Yuan, S. Dong, and E. H. Adelson. Gelsight: High-resolution robot tactile sensors for estimating geometry and force. *Sensors (Basel, Switzerland)*, 17, 2017. URL https://api.semanticscholar.org/CorpusID:3474913.

[11] M. Lambeta, P.-W. Chou, S. Tian, B. Yang, B. Maloon, V. R. Most, D. Stroud, R. Santos, A. Byagowi, G. Kammerer, D. Jayaraman, and R. Calandra. 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. ISSN 2377-3774. doi:10.1109/lra.2020.2977257. URL http://dx.doi.org/10.1109/LRA.2020.2977257.

[12] B. Ward-Cherrier, N. Pestell, L. Cramphorn, B. Winstone, M. Giannaccini, J. Rossiter, and N. Lepora. The tactip family: Soft optical tactile sensors with 3d-printed biomimetic morphologies. *Soft Robotics*, 5, 01 2018. doi:10.1089/soro.2017.0052.

[13] C. Lin, H. Zhang, J. Xu, L. Wu, and H. Xu. 9dtact: A compact vision-based tactile sensor for accurate 3d shape reconstruction and generalizable 6d force estimation, 2023. URL https://arxiv.org/abs/2308.14277.

[14] T. Tomo, A. Schmitz, W. Wong, H. Kristanto, S. Somlor, J. Hwang, L. Jamone, and S. Sugano. Covering a robot fingertip with uskin: A soft electronic skin with distributed 3-axis force sensitive elements for robot hands. *IEEE Robotics and Automation Letters*, PP:1–1, 08 2017. doi:10.1109/LRA.2017.2734965.

[15] R. Bhirangi, T. Hellebrekers, C. Majidi, and A. Gupta. Reskin: versatile, replaceable, lasting tactile skins, 2022. URL https://arxiv.org/abs/2111.00071.

[16] V. Pattabiraman, Z. Huang, D. Panozzo, D. Zorin, L. Pinto, and R. Bhirangi. eflesh: Highly customizable magnetic touch sensing using cut-cell microstructures, 2025. URL https://arxiv.org/abs/2506.09994.

[17] B. Huang and Y. Li. Flexitac: A low-cost, open-source, scalable tactile sensing solution for robotic systems, 2026. URL https://arxiv.org/abs/2604.28156.

[18] H. Khamis, R. Albero, M. Salerno, A. Shah Idil, and A. Loizou. Papillarray: An incipient slip sensor for dexterous robotic or prosthetic manipulation – design and prototype validation. *Sensors and Actuators A: Physical*, 270, 12 2017. doi:10.1016/j.sna.2017.12.058.

[19] A. van den Oord, Y. Li, and O. Vinyals. Representation learning with contrastive predictive coding, 2019. URL https://arxiv.org/abs/1807.03748.

[20] T. Z. Zhao, V. Kumar, S. Levine, and C. Finn. Learning fine-grained bimanual manipulation with low-cost hardware, 2023. URL https://arxiv.org/abs/2304.13705.

[21] H. Qi, B. Yi, S. Suresh, M. Lambeta, Y. Ma, R. Calandra, and J. Malik. General in-hand object rotation with vision and touch, 2023. URL https://arxiv.org/abs/2309.09979.

[22] T. Lin, Y. Zhang, Q. Li, H. Qi, B. Yi, S. Levine, and J. Malik. Learning visuotactile skills with two multifingered hands, 2024. URL https://arxiv.org/abs/2404.16823.

[23] Z.-H. Yin, C. Wang, L. Pineda, F. Hogan, K. Bodduluri, A. Sharma, P. Lancaster, I. Prasad, M. Kalakrishnan, J. Malik, M. Lambeta, T. Wu, P. Abbeel, and M. Mukadam. Dexteritygen: Foundation controller for unprecedented dexterity, 2025. URL https://arxiv.org/abs/2502.04307.

[24] X. Liu, H. Wang, and L. Yi. Dexndm: Closing the reality gap for dexterous in-hand rotation via joint-wise neural dynamics model, 2025. URL https://arxiv.org/abs/2510.08556.

[25] S. Jiang, S. Zhao, Y. Fan, and P. Yin. Gelfusion: Enhancing robotic manipulation under visual constraints via visuotactile fusion, 2025. URL https://arxiv.org/abs/2505.07455.

[26] Y. She, S. Wang, S. Dong, N. Sunil, A. Rodriguez, and E. Adelson. Cable manipulation with a tactile-reactive gripper, 2020. URL https://arxiv.org/abs/1910.02860.

[27] F. R. Hogan, M. Bauza, O. Canal, E. Donlon, and A. Rodriguez. Tactile regrasp: Grasp adjustments via simulated tactile transformations, 2018. URL https://arxiv.org/abs/1803.01940.

[28] E. Donlon, S. Dong, M. Liu, J. Li, E. Adelson, and A. Rodriguez. Gelslim: A high-resolution, compact, robust, and calibrated tactile-sensing finger, 2018. URL https://arxiv.org/abs/1803.00628.

[29] Daimon Robotics. DM-Tac W: High-resolution vision-based tactile sensor. https://www.dmrobot.com/en/, 2025. Accessed: 2026-05-28.

[30] A. Alspach, K. Hashimoto, N. Kuppuswamy, and R. Tedrake. Soft-bubble: A highly compliant dense geometry tactile sensor for robot manipulation, 2019. URL https://arxiv.org/abs/1904.02252.

[31] W. K. Do and M. K. III. Densetact: Optical tactile sensor for dense shape reconstruction, 2022. URL https://arxiv.org/abs/2201.01367.

[32] R. Bhirangi, V. Pattabiraman, E. Erciyes, Y. Cao, T. Hellebrekers, and L. Pinto. Anyskin: Plug-and-play skin sensing for robotic touch, 2024. URL https://arxiv.org/abs/2409.08276.

[33] N. Wettels, V. Santos, R. Johansson, and G. Loeb. Biomimetic tactile sensor array. *Advanced Robotics*, 22:829–849, 08 2008. doi:10.1163/156855308X314533.

[34] M. S. Li and H. S. Stuart. Acoustac: Tactile sensing with acoustic resonance for electronics-free soft skin, 2023. URL https://arxiv.org/abs/2307.09730.

[35] K. Zhang, D.-G. Kim, E. T. Chang, H.-H. Liang, Z. He, K. Lampo, P. Wu, I. Kymissis, and M. Ciocarlie. Vibecheck: Using active acoustic tactile sensing for contact-rich manipulation, 2025. URL https://arxiv.org/abs/2504.15535.

[36] M. A. Lee, Y. Zhu, K. Srinivasan, P. Shah, S. Savarese, L. Fei-Fei, A. Garg, and J. Bohg. Making sense of vision and touch: Self-supervised learning of multimodal representations for contact-rich tasks, 2019. URL https://arxiv.org/abs/1810.10191.

[37] 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, 2025. URL https://arxiv.org/abs/2505.11420.

[38] C. Higuera, A. Sharma, T. Fan, C. K. Bodduluri, B. Boots, M. Kaess, M. Lambeta, T. Wu, Z. Liu, F. R. Hogan, and M. Mukadam. Tactile beyond pixels: Multisensory touch representations for robot manipulation, 2025. URL https://arxiv.org/abs/2506.14754.

[39] Z. Xu, R. Uppuluri, X. Zhang, C. Fitch, P. G. Crandall, W. Shou, D. Wang, and Y. She. Unit: Data efficient tactile representation with generalization to unseen objects. 2025. URL https://arxiv.org/abs/2408.06481.

[40] A. Zorin, Z. Si, M. Park, J. Park, A. Buynitsky, S. Bhadang, T. Park, S. J. Yoon, Y.-L. Park, O. Kroemer, Z. Temel, M. T. Tolley, S. Yi, and X. Wang. Taco: Benchmarking tactile sensors for object manipulation, 2026. URL https://arxiv.org/abs/2605.21976.

[41] G. Jiang, Y. Liang, J. Ye, J.-Y. Huang, C. Jing, R. Duan, P. Abbeel, X. Wang, and X. Zou. Cross-hand latent representation for vision-language-action models, 2026. URL https://arxiv.org/abs/2603.10158.

[42] E. Bauer, E. Nava, and R. K. Katzschmann. Latent action diffusion for cross-embodiment manipulation, 2026. URL https://arxiv.org/abs/2506.14608.

[43] T. Wang, D. Bhatt, X. Wang, and N. Atanasov. Cross-embodiment robot manipulation skill transfer using latent space alignment, 2024. URL https://arxiv.org/abs/2406.01968.

[44] A. Dastider, H. Fang, and M. Lin. Cross-embodiment robotic manipulation synthesis via guided demonstrations through cyclevae and human behavior transformer, 2025. URL https://arxiv.org/abs/2503.08622.

[45] Q. Bu, Y. Yang, J. Cai, S. Gao, G. Ren, M. Yao, P. Luo, and H. Li. Univla: Learning to act anywhere with task-centric latent actions, 2025. URL https://arxiv.org/abs/2505.06111.

[46] H. Gupta, Y. Mo, S. Jin, and W. Yuan. Sensor-invariant tactile representation, 2025. URL https://arxiv.org/abs/2502.19638.

[47] R. Feng, Y. Zhou, S. Mei, D. Zhou, P. Wang, S. Cui, B. Fang, G. Yao, and D. Hu. Anytouch 2: General optical tactile representation learning for dynamic tactile perception, 2026. URL https://arxiv.org/abs/2602.09617.

[48] S. Rodriguez, Y. Dou, W. van den Bogert, M. Oller, K. So, A. Owens, and N. Fazeli. Contrastive touch-to-touch pretraining, 2024. URL https://arxiv.org/abs/2410.11834.

[49] Z. Chen, F. Ni, K. Luo, Z. Wu, X. Zhang, E. Spyrakos-Papastavridis, L. Jamone, N. F. Lepora, J. Deng, and S. Luo. Uniforce: A unified latent force model for robot manipulation with diverse tactile sensors, 2026. URL https://arxiv.org/abs/2602.01153.

[50] Z. Chen, N. Ou, X. Zhang, Z. Wu, Y. Zhao, Y. Wang, E. S. Papastavridis, N. Lepora, L. Jamone, J. Deng, and S. Luo. Training tactile sensors to learn force sensing from each other, 2025. URL https://arxiv.org/abs/2503.01058.

[51] J. Hou, X. Zhou, Q. Yang, and A. J. Spiers. Unitac-nv: A unified tactile representation for non-vision-based tactile sensors, 2025. URL https://arxiv.org/abs/2506.19699.

[52] Z. Chen, N. Ou, X. Zhang, and S. Luo. Transforce: Transferable force prediction for vision-based tactile sensors with sequential image translation, 2025. URL https://arxiv.org/abs/2409.09870.

[53] Y. Wi, J. Yin, E. Xiang, A. Sharma, J. Malik, M. Mukadam, N. Fazeli, and T. Hellebrekers. Tactalign: Human-to-robot policy transfer via tactile alignment, 2026. URL https://arxiv.org/abs/2602.13579.

[54] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton. A simple framework for contrastive learning of visual representations, 2020. URL https://arxiv.org/abs/2002.05709.

[55] P. Wu, Y. Shentu, Z. Yi, X. Lin, and P. Abbeel. Gello: A general, low-cost, and intuitive teleoperation framework for robot manipulators, 2024. URL https://arxiv.org/abs/2309.13037.

A Data Collection Details

Sensors. To prevent too much visual change the eFlesh housing is 3D-printed in black TPU and the FlexiTac surface is covered with black anti-slip tape, matching Daimon's black elastomer; the three active sensing areas are roughly commensurate so a contact is captured by every sensor.

(a) Daimon (vision-based) (b) eFlesh (magnetic) (c) FlexiTac (resistive)

Figure 7: The three tactile sensors used in TACTX, each spanning a different transduction modality. All three are visually matched (black TPU/tape/elastomer) to remove cosmetic shortcuts and have roughly commensurate active sensing areas.

Mounting and pairing. Two sensors are mounted on opposing fingers of a Franka parallel-jaw gripper; the third is swapped in for separate runs. We cover all $\binom{3}{2} \times 2 = 6$ configurations: each unordered pair is recorded twice with sensors swapped between the left and right fingers, doubling the effective pair-dataset size and removing left/right asymmetry. This yields three pair-datasets $\mathcal{D}_{DE}, \mathcal{D}_{EF}, \mathcal{D}_{FD}$. Because the sensors sit at structurally different positions on their respective housings, we apply a one-time 180° rotation to one sensor of each pair at load time so contact regions align across paired observations.

Objects and protocol. Pretraining data uses 10 3D-printed objects (Fig. 8) spanning point contacts (small ball), edge contacts (line, triangle, circle circumference), and area contacts (large ball, ellipse), sized to fit within the shared sensing area. Each grasp follows a scripted approach–contact–stable-grasp–release–withdraw protocol; automation ensures consistent contact conditions across the many repetitions needed for paired supervision. All three sensors are sampled at a matched rate and paired frames are aligned by nearest timestamp.

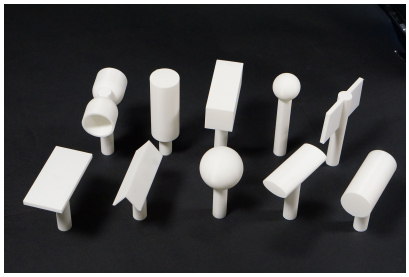


Figure 8: The 10 3D-printed pretraining objects used for paired data collection, spanning point, edge, and area contact geometries.

Contact gating. At preprocessing, frames are labeled *contact* if Daimon depth abs-mean exceeds 0.003 or FlexiTac pressure grid mean exceeds 0.01; otherwise no-contact.

Table 3: Pretraining data composition.

Objects (10)	circle, cylinder, cylinder_vertical, ellipse, large_ball, plane, plane_horizontal, small_ball, square, triangle
Pair-datasets	$\mathcal{D}_{DE}$, $\mathcal{D}_{EF}$, $\mathcal{D}_{FD}$ (6 mounting configs, L/R-swapped)
Trajectories	2,670 total ($\sim$442–448 / config, across 10 objects)
Frames	145k total ($\sim$20k–31k / config, $\sim$54 / trajectory)
Train/val split	episode-level, $\sim$20% val, sampled across the recording order

B TACTX Architecture

Each encoder maps its native signal to a 512-D feature, then a shared-form projection head ($512 \to 512 \to 2d$, Linear–ReLU–Linear) outputs $(\mu, \log \sigma^2)$ over the d=16 latent. At training $z = \mu + \sigma \odot \epsilon$, $\epsilon \sim \mathcal{N}(0, I)$; at inference $z = \mu$. Decoders are sensor-specific (no shared decoder, since output spaces differ too much) and each is used for both self- and cross-reconstruction. All modules train from scratch.

Table 4: Per-sensor encoder / decoder architectures. Shared latent d=16.

Sensor	Modality	Native input	Encoder backbone	Decoder
Daimon (D)	vision-based	224×224×3 (depth+shear)	ResNet-18 → 512	linear → transposed conv
eFlesh (E)	magnetic	15-D field vector	MLP $[64, 128, 256] \to 512$	reversed MLP
FlexiTac (F)	resistive	12×16 pressure grid	residual CNN → 512	mirrored conv

C TACTX Training

Table 5: TACTX representation-learning hyperparameters.

Latent dim d	16
Batch size	64
Learning rate	1×10^{-4} (Adam, weight decay 1×10^{-4})
Epochs	300
Seed	42
λ_{recon}	1.0
KL weight β	$0 \to 0.1$ linear warmup over 30 epochs
InfoNCE temperature τ	0.01 (NT-Xent variant: 0.03)
Alignment weight λ_{align}	$0 \to 1$ warmup (start 0)

D Downstream Policy Details

Overview. All manipulation policies are Action Chunking Transformers (ACT) with a DETR-style CVAE decoder. Demonstrations are collected via GELLO teleoperation (∼50 episodes per task, ∼10k frames each). Unless noted, training uses learning rate 1×10^{-5}, batch size 8, and 50,000 training steps with action-chunk length 64. Robot state and actions are padded to 128 dimensions; only the first 8 joint/gripper dimensions are used at deployment. Two RGB cameras (cam0 wrist, cam1 third-person or side-view (for board wiping)) and two tactile fingers (tac0, tac1) are always logged; how tactile data enter ACT depends on the variant below.

Shared ACT backbone. Visual observations are encoded with an ImageNet-pretrained ResNet-18 and ACT_linear projection into hidden dimension 512. Images are resized to 224×308 ($16 \times 14 \times 22 \times 14$ patches) and ImageNet-normalized. The transformer uses 4 encoder layers, 7 decoder layers, 8 attention heads, hidden dimension 512, and feed-forward width 3200. The training objective is an L1 loss on predicted action chunks plus λ_{KL} times the KL divergence on ACT's internal CVAE latent $z \in R^{32}$, which encodes action-sequence diversity and is set to zero at inference. This CVAE latent is distinct from any tactile representation (raw sensor data, binary contact, or TACTX latents).

D.1 Raw tactile policies

Raw policies consume sensor-native measurements per finger with sensor-specific encoders but otherwise share the hyperparameters in Table 6. We set $\lambda_{\text{KL}} = 10$ (ACT default) for all raw runs.

- **Daimon (D, vision-style).** Modality image: each finger provides a $240 \times 320 \times 3$ composite visualization (depth, deformation, and shear panels). A second ResNet-18 tactile backbone

Table 6: Shared ACT hyperparameters (all tactile variants).

Policy class	ACT (DETR-style action chunking)
State / action dim	128 / 128 (padded; first 8 used)
Enc. / dec. layers	4 / 7 nheads 8 hidden dim 512 FFN 3200
Image backbone	ResNet-18 (ImageNet-pretrained), ACT_linear features
Cameras	cam0 (wrist), cam1 (third-person or side-view)
Tactile fingers	tac0, tac1
Chunk size / queries	64
Learning rate	1×10^{-5}
Batch size / steps	8 / 50,000
Demonstrations	$\sim$50 / task via GELLO teleoperation [55]

($\ell_{\text{tactile}} = 10^{-4}$) encodes these images; projected features are concatenated with camera feature maps along the spatial width axis, matching the layout of additional camera views. The camera backbone learning rate is 10^{-5}.

- **FlexiTac (F, array).** Modality array: a $12 \times 16 = 192$-dimensional resistive grid per finger ($\sim$30 Hz). Each finger is mapped by an MLP adapter $192 \to 64 \to 128 \to 512$ to one tactile token prepended to the transformer (alongside proprioception and the CVAE token). The camera backbone learning rate is 10^{-5}.
- **eFlesh (E, array).** Modality array: a 15-dimensional magnetic vector per finger (5 magnets $\times$ 3 axes). MLP adapter $15 \to 64 \to 128 \to 512$ per finger with the same token injection as FlexiTac. The camera backbone learning rate is 10^{-5}.

Vision-only ablations set tactile modality to none (cameras and proprioception only), using the same task datasets with tactile channels ignored.

D.2 Latent tactile policies (TACTX)

A frozen cross-modal VAE encoder maps each sensor's raw tactile observations to a shared 16-dimensional latent mean vector, μ, offline before ACT training. Daimon latents are computed from tactile RGB composites, eFlesh latents from 15-dimensional magnetic readings, and FlexiTac latents from 12×16 resistive tactile grids. These precomputed latents replace the raw tactile inputs in the policy datasets, and the encoder remains frozen during policy learning. For each dataset, the latent representations are normalized using the corresponding latent mean and standard deviation statistics before ACT training.

A lightweight MLP adapter is applied independently to each finger:

$$16 \to 64 \to 128 \to 512,$$

mapping each TACTX latent into a single tactile token for the transformer. Consequently, ACT does not require a tactile CNN encoder. We further reduce λ_{KL} from 10 to 1 to mitigate mode collapse on the relatively small per-task datasets. During deployment, only the sensor-specific encoder branch changes across tactile hardware, while the ACT policy weights and the shared 16-dimensional latent space remain fixed.

D.3 Binary contact policies

Contact policies encode a single binary contact bit per finger. Offline, each timestep is labeled $1[\text{mean}(|x - x_{\text{baseline}}|) > \tau_s]$ using a per-sensor baseline x_{baseline} and threshold τ_s fit on training data (Daimon uses depth and shear channels before scoring). ACT uses modality array with $tactile\_dim = 1$ and MLP adapter $1 \to 64 \to 128 \to 512$. The hidden layers $[64, 128]$ match the TACTX run so that only input dimension (1 vs. 16) differs between contact and latent policies. We set $\lambda_{\text{KL}} = 1$, as for latent tactile. The same ACT checkpoint can be served on different tactile hardware by swapping only $(x_{\text{baseline}}, \tau_s)$ at inference.

Table 7: Tactile-specific ACT settings. Shared backbone hyperparameters are in Table 6.

Variant	Modality	Per-finger input	Tactile encoder in ACT
Raw Daimon (D)	image	$240\times320\times3$ composite	ResNet-18 $\rightarrow$ spatial tokens
Raw FlexiTac (F)	array	192-D resistive grid	MLP $192\rightarrow64\rightarrow128\rightarrow512$
Raw eFlesh (E)	array	15-D magnetic vector	MLP $15\rightarrow64\rightarrow128\rightarrow512$
Latent TACTX	array	16-D μ (frozen VAE, offline)	MLP $16\rightarrow64\rightarrow128\rightarrow512$
Binary contact	array	1-D thresholded bit (offline)	MLP $1\rightarrow64\rightarrow128\rightarrow512$

Table 8: **In-domain policy performance.** Same-sensor train and test. TACTX matches or exceeds both baselines on pick-and-place and recovers most of the in-domain insertion gain provided by raw tactile. Success rates over 10 rollouts; bold marks the best per task per sensor.

Method	**Sensor**	**P&P**	**P&P (OOD)**	**Insertion**	**Wiping**	**Reorient.**
Vision only	Daimon	9/10	8/10	2/10	4/10	9/10
	eFlesh	9/10	6/10	2/10	0/10	8/10
	FlexiTac	7/10	6/10	8/10	9/10	7/10
Vision + Ground Truth Tactile	Daimon	9/10	7/10	7/10	10/10	8/10
	eFlesh	**10/10**	**9/10**	7/10	6/10	10/10
	FlexiTac	9/10	8/10	8/10	9/10	10/10
Vision + TACTX (Ours)	Daimon	**10/10**	5/10	5/10	5/10	10/10
	eFlesh	**10/10**	8/10	5/10	8/10	9/10
	FlexiTac	**10/10**	**9/10**	8/10	9/10	10/10

E Full Policy Evaluation Results

E.1 In-domain Policy Evaluation

We report the full in-domain policy results for each tactile sensor and task in Table 8. These experiments evaluate whether tactile-conditioned ACT policies benefit from tactile input when trained and evaluated with the same sensor. Raw tactile observations provide a strong same-sensor baseline, while TACTX tests whether the shared latent preserves enough task-relevant contact information for downstream control.