BIDETA: BRAIN-INSPIRED DATA-EFFICIENT TACTILE ADAPTATION FOR UNSEEN SENSORS

Anonymous authors
Paper under double-blind review

ABSTRACT

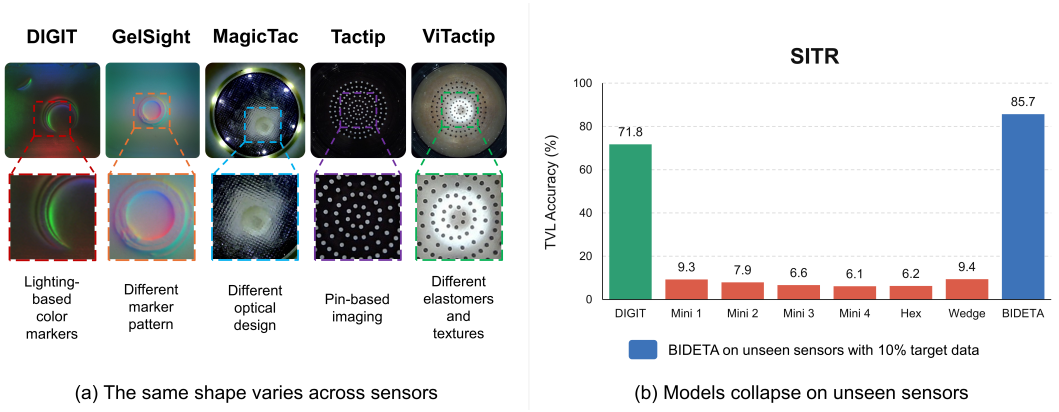
Vision-based tactile sensors provide high-resolution contact information for robotic perception and contact-rich manipulation, advancing embodied intelligence through more reliable physical interaction. However, device-specific sensing mechanisms cause tactile foundation models to degrade on unfamiliar hardware. Existing cross-sensor methods often require calibration data, paired observations, or iterative training. To address this problem, we propose Brain-Inspired Data-Efficient Tactile Adaptation (BIDETA), a gradient-free framework that uses a frozen tactile encoder and a few labeled target contacts to jointly predict labels for an unlabeled query batch. Inspired by the brain's rapid sensory adaptation, BIDETA combines rapid support memory, support-conditioned spectral graphs, and reliability-gated recurrence to preserve pretrained representations, repair sensor-dependent feature neighborhoods, and integrate reliable cross-query evidence. Experiments on SITR, TacVerse Shape, and TacQuad show that BIDETA substantially improves adaptation to unknown sensors: with only 10% labeled target data on SITR, it raises mean Sparsh accuracy from 6.86% for the frozen source classifier to 87.09%, exceeding the strongest implemented prior comparison by 47.22 percentage points, and these gains generalize across datasets, pretrained backbones, and tactile tasks. In the SITR timing benchmark with TVL, BIDETA also achieves approximately 20× faster target-sensor adaptation than the best baseline. BIDETA thus offers a gradient-free, data-efficient route to deploying tactile models on new hardware.

1 INTRODUCTION

With rapid progress in embodied intelligence, touch is becoming a core capability for agents that must perceive and understand the physical world through contact (Luo et al., 2025). Vision-based tactile sensors convert contact-induced deformation of a soft skin into high-resolution images, giving robots access to contact geometry, force, and slip that external vision cannot observe directly (Yuan et al., 2017; Lambeta et al., 2020). However, differences in optical design, elastomer mechanics, marker or pin structure, and field of view can cause the same contact to appear substantially different across sensors (Wei et al., 2026; Cong et al., 2026). Despite substantial advances in tactile foundation models, their performance often drops sharply on unseen sensors introduced by sensor replacement, hardware upgrades, or deployment across heterogeneous robot fleets. Rapid adaptation to unseen tactile sensors is therefore essential for advancing embodied intelligence, allowing existing tactile models to serve diverse hardware.

Recent tactile foundation models reduce the cost of learning transferable features. TVL aligns touch with vision and language to learn semantically grounded tactile representations, while Sparsh uses self-supervised learning on heterogeneous tactile data to support transfer across downstream tasks (Fu et al., 2024; Higuera et al., 2025a). However, the same contact can produce markedly different images across sensors, causing source-trained models to fail on unseen hardware. Figure 1 shows this failure: a classifier trained with frozen TVL features performs well on its source sensor but falls close to chance on unknown sensors. Existing cross-sensor methods mitigate this shift, but often rely on calibration data, paired observations, or multi-sensor collection, limiting rapid adaptation to a new device (Gupta et al., 2025; Feng et al., 2025; Zhang et al., 2026). Lighter-weight alternatives use prototypes, support caches, or query relations without updating the tactile encoder (Wang et al., 2019; Zhang et al., 2022; Ziko et al., 2020). Yet their effectiveness depends on pretrained feature geometry remaining meaningful after sensor shift; distorted target neighborhoods can make both support-based

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.



(a) The same shape varies across sensors

(b) Models collapse on unseen sensors

Figure 1: **The same shape produces markedly different tactile observations across sensors, degrading model performance on unseen sensors.** (a) The same shape produces distinct tactile images across sensor designs. (b) A linear classifier trained on the source sensor is evaluated using frozen TVL features. On SITR (Gupta et al., 2025), accuracy falls from 71.8% on DIGIT to 9.3%, 7.9%, 6.6%, 6.1%, 6.2%, and 9.4% on Mini 1, Mini 2, Mini 3, Mini 4, Hex, and Wedge, respectively. With 10% labeled target data, BIDETA reaches 85.7% averaged across all six unknown sensors.

predictions and query relations unreliable. This exposes the central question of our work: *how can a tactile foundation model use a few labeled target contacts to transfer its perceptual capability rapidly and reliably across unseen sensors?*

Even when pretrained tactile features retain useful class information, rapid and reliable transfer to unseen sensors remains challenging: sensor shift changes how classes are distinguished, distorts feature neighborhoods, and makes query relations unreliable. Data-efficient tactile adaptation should therefore satisfy three requirements. First, it should recover class evidence from a few labeled target contacts while preserving the pretrained representation. Second, it should use variation in those contacts to recalibrate feature neighborhoods, so that query relations remain meaningful on the new sensor. Third, it should determine how much relational evidence each query incorporates according to prediction reliability, without allowing unreliable relations to override support-derived evidence. These requirements call for a data-efficient method that rapidly adapts to unseen sensors with few labeled examples, without retraining the tactile foundation model.

To meet these requirements, we draw inspiration from the brain's rapid sensory adaptation mechanism (McClelland et al., 1995; Carandini & Heeger, 2012; Ernst & Banks, 2002; Khona & Fiete, 2022) and propose Brain-Inspired Data-Efficient Tactile Adaptation (BIDETA). First, *rapid support memory* fits discriminant memories to two readouts from the same frozen encoder; it then fuses them into a stable probability anchor. Second, *support-conditioned spectral geometry* uses labeled target variation to suppress unstable feature directions; it then constructs a cross-readout consensus graph with more reliable neighborhoods. Third, *reliability-gated anchored recurrence* uses prediction ambiguity and readout disagreement; these signals determine how strongly each query should absorb graph evidence while retaining its support-derived anchor. By separating stable memory from rapid geometric recalibration and reliability-controlled integration, BIDETA converts unknown-sensor transfer into a support-conditioned inference problem without gradient-based encoder updates, bridging the gap between pretrained sensor experience and previously unseen hardware.

Experiments across three tactile datasets, using two frozen backbones, demonstrate that BIDETA substantially improves perception on unknown sensors. With only 10% labeled target data on SITR, BIDETA raises TVL accuracy from the frozen source classifier's 7.58% to 85.71% and exceeds the strongest external comparison by 38.24 percentage points. On TacVerse Shape, the same target-data budget raises Sparsh accuracy from 13.67% to 83.58%, exceeding the strongest external comparison by 27.60 percentage points. The gains extend across datasets, pretrained backbones, and tactile classification and retrieval tasks. These results show that brain-inspired rapid adaptation can strengthen cross-sensor tactile generalization, thereby offering a promising direction for robust embodied perception across evolving hardware.

2 Related Work

Tactile foundation models. Vision-based tactile learning began with sensors such as GelSight and DIGIT, which convert contact deformation into high-resolution images (Yuan et al., 2017; Lambeta et al., 2020). Tactile foundation models progressed from early paired vision–touch learning with Touch and Go toward reusable multimodal and self-supervised representations (Yang et al., 2022). TVL aligns touch with vision and language; Sparsh learns transferable features from unlabeled tactile data; UniTouch and ViT-Lens connect heterogeneous sensors to shared multimodal representation spaces through sensor-specific tokens or pretrained visual encoders (Fu et al., 2024; Higuera et al., 2025a; Yang et al., 2024; Lei et al., 2024). Recent systems further extend tactile learning to multisensory manipulation, distributed tactile skin, and visual–tactile material localization (Higuera et al., 2025b; Sharma et al., 2025; Kim et al., 2026). These advances improve tactile perception, but the resulting models remain difficult to adapt efficiently to other sensors across diverse downstream tasks.

Data-Efficient Adaptation to Unseen Tactile Sensors. Data-efficient adaptation methods broadly follow optimization-based or relation-based strategies. Optimization-based methods fit a lightweight classifier, adapter, or test-time state from the support set, directly adjusting the decision boundary but requiring iterative updates and risking overfitting to scarce examples (Karmanov et al., 2024; Boudiaf et al., 2020; Singh et al., 2026). Relation-based methods retain frozen features and classify through normalized prototypes, support caches, probability objectives, or query graphs (Wang et al., 2019; Zhang et al., 2022; Ziko et al., 2020; Zhou et al., 2003; Martin et al., 2024). Data-efficient adaptation methods designed specifically for tactile hardware remain sparse. The closest cross-sensor methods use simulated variation and calibration in SITR, cross-sensor matching in AnyTouch, or synthetic transfer and sensor-conditioned modulation in CTSRL (Gupta et al., 2025; Feng et al., 2025; Zhang et al., 2026). These methods require specific calibration or paired data, or struggle to adapt when an unseen sensor differs substantially from the training sensors. Such dependence on predefined acquisition protocols consequently limits rapid deployment with heterogeneous tactile hardware.

Brain-inspired methods. The human brain adapts efficiently to changing environments by combining specialized neural processes, flexible memory, and dynamic regulation, providing computational principles for more robust and efficient algorithm design (Zador et al., 2023). Recent brain-inspired methods model prefrontal component processes for language-model planning, cortico-hippocampal dual-memory circuits for continual learning, and adaptive excitation–inhibition balance for reservoir computing (Webb et al., 2025; Shi et al., 2025; Srinivasan et al., 2025). By translating neural organization, memory, and self-regulation into computational mechanisms, these methods improve planning, continual adaptation, and efficiency while addressing the rigidity of uniformly structured learning systems under rapidly changing conditions and distribution shifts.

3 Method

3.1 Problem formulation

We consider an unseen target tactile sensor with a labeled support set $\mathcal{S} = \{(x_i, y_i)\}_{i=1}^{n}$ and an unlabeled query batch $\mathcal{Q} = \{x_j\}_{j=1}^{m}$, where x denotes a tactile observation, $y_i \in \{1, \dots, C\}$ is its label, C is the number of classes, and every class appears in $\mathcal{S}$. A pretrained encoder f_θ with fixed parameters θ provides two feature readouts $h^{(v)}(x) \in \mathbb{R}^{d_v}$, where $v \in \{1, 2\}$ indexes the readout and d_v is its dimension. Given $\mathcal{S}$ and $\mathcal{Q}$, our goal is to predict the m query labels jointly without query annotations, source-data replay, or encoder updates.

3.2 Brain-inspired design overview

Human perception remains stable under changing sensory conditions, as the brain combines long-term knowledge with rapid sensory adaptation. Complementary learning protects established representations while incorporating new experience, sensory normalization recalibrates responses to current input statistics, and reliability-weighted recurrent integration accumulates uncertain evidence toward a stable interpretation (McClelland et al., 1995; Carandini & Heeger, 2012; Ernst & Banks, 2002; Khona & Fiete, 2022). Together, these processes allow perception to adjust quickly without discarding previously acquired knowledge.

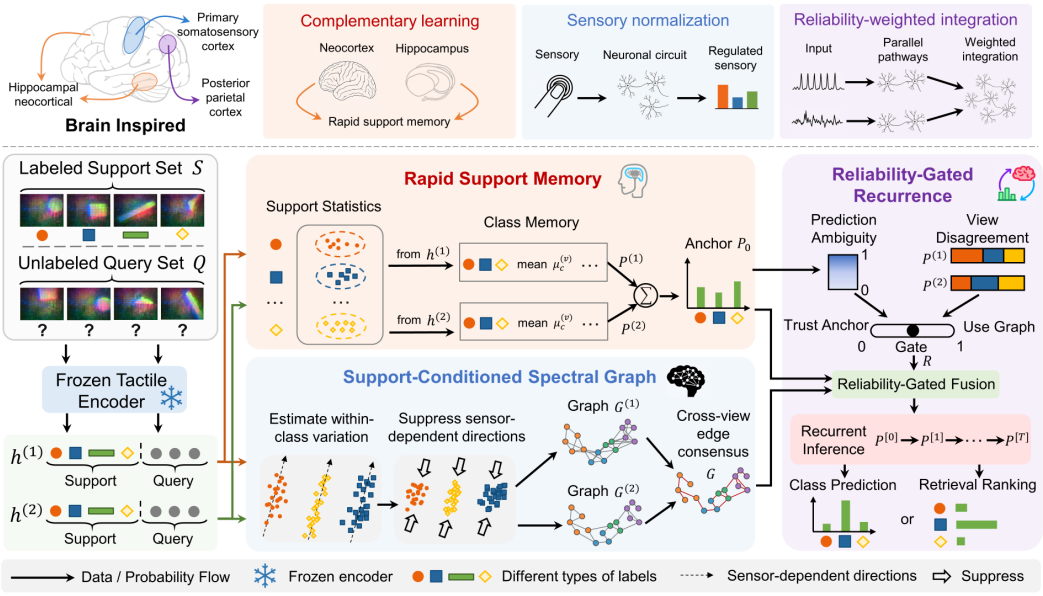


Figure 2: **Information flow through BIDETA.** A frozen encoder produces two readouts. Rapid support memory forms the probability anchor P_0, while the support-conditioned spectral graph reshapes query geometry and constructs the cross-readout affinity matrix G. Prediction ambiguity and readout disagreement determine the query-wise reliability gate R. Anchored recurrent inference combines P_0, G, and R to produce class probabilities or identity rankings.

Inspired by this mechanism, BIDETA couples three modules, as shown in Figure 2. *Rapid support memory* converts the two frozen readouts into a support-supervised probability anchor $P_0 \in [0,1]^{m\times C}$. A *support-conditioned spectral graph* suppresses unstable feature directions and produces a query affinity matrix $G \in \mathbb{R}_+^{m\times m}$, where $\mathbb{R}_+$ denotes the nonnegative real numbers. A *reliability-gated recurrence* forms a diagonal matrix $R \in [0,1]^{m\times m}$ that controls how strongly each query uses relational evidence. Their joint action preserves stable class evidence while recalibrating sensor-dependent geometry and evidence flow for the unknown sensor.

3.3 Rapid Support Memory

The brain combines stable long-term representations with rapid learning from limited new experience, allowing perception to adapt without discarding prior knowledge. Inspired by this complementary learning principle, we design rapid support memory to preserve the frozen encoder while extracting class evidence from a small target support set. For each readout v, we compute the support mean $b^{(v)} \in \mathbb{R}^{d_v}$ and coordinate-wise standard deviation $s^{(v)} \in \mathbb{R}^{d_v}$, then standardize every support or query feature as

$$z^{(v)}(x) = \left(h^{(v)}(x) - b^{(v)}\right) \oslash s^{(v)}, \tag{1}$$

where $\oslash$ denotes coordinate-wise division and constant coordinates use unit scale. Let $n_c = |\{i : y_i = c\}|$, $\pi_c = n_c/n$, and $\mu_c^{(v)}$ denote the support count, support prior, and standardized mean of class c in readout v. We fit a shrinkage linear discriminant analysis memory (Fisher, 1936) with pooled within-class covariance $\Sigma_\rho^{(v)} \in \mathbb{R}^{d_v\times d_v}$ and shrinkage parameter $\rho \in [0,1]$. Its class score for a standardized feature z is

$$\ell_c^{(v)}(z) = z^\top (\Sigma_\rho^{(v)})^{-1}\mu_c^{(v)} - \frac{1}{2}(\mu_c^{(v)})^\top (\Sigma_\rho^{(v)})^{-1}\mu_c^{(v)} + \log \pi_c. \tag{2}$$

For query x_j, a temperature $\tau_m > 0$ converts the scores into the class distribution $P_j^{(v)} \in [0,1]^C$ through

$$P_{jc}^{(v)} = \mathrm{softmax}_c\left(\ell_c^{(v)}(z^{(v)}(x_j))/\tau_m\right), \qquad P_0 = \alpha P^{(1)} + (1-\alpha)P^{(2)}, \tag{3}$$

where $\mathrm{softmax}_c$ normalizes over the C class scores, $\alpha \in [0,1]$ is the readout weight, and P_0 is the fused probability anchor. This anchor carries direct support-supervised class evidence into the final inference stage.

3.4 Support-conditioned spectral graph

Neural sensory systems normalize responses according to current input statistics, reducing the influence of dominant but uninformative variation. Drawing on this adaptive regulation principle, we design the support-conditioned spectral graph to use within-class target variation to reshape feature geometry and obtain more reliable query neighborhoods. Sensor changes can enlarge feature directions that vary within a class, causing them to dominate query similarity. For each support example, we define the class residual $e_i^{(v)} = z^{(v)}(x_i) - \mu_{y_i}^{(v)}$ and estimate its covariance with Ledoit–Wolf shrinkage (Ledoit & Wolf, 2004):

$$\widehat{\Sigma}_w^{(v)} = \mathrm{LW}\Big(\{e_i^{(v)}\}_{i=1}^n\Big) = U^{(v)}\,\mathrm{diag}(\lambda_1^{(v)},\dots,\lambda_{d_v}^{(v)})(U^{(v)})^\top, \tag{4}$$

where $U^{(v)}$ contains the eigenvectors and $\lambda_a^{(v)}$ is the eigenvalue of direction a. Given a spectral exponent $\gamma \geq 0$ and numerical floor $\epsilon = 10^{-6}$, we downweight high-variance directions through

$$g_a^{(v)} = \frac{\max(\lambda_a^{(v)},\epsilon)^{-\gamma}}{\mathrm{median}_b \max(\lambda_b^{(v)},\epsilon)^{-\gamma}}, \qquad A^{(v)} = U^{(v)}\,\mathrm{diag}(g^{(v)})(U^{(v)})^\top,$$
$$\widetilde{z}_j^{(v)} = A^{(v)} z^{(v)}(x_j), \tag{5}$$

where $g_a^{(v)}$ is the gain for direction a, $A^{(v)}$ is the spectral transform, and $\widetilde{z}_j^{(v)}$ is the transformed query feature. The transform changes only the geometry used to connect queries and leaves the probability anchor in Eq. (3) unchanged.

We normalize $q_j^{(v)} = \widetilde{z}_j^{(v)}/\|\widetilde{z}_j^{(v)}\|_2$ and construct a directed k-nearest-neighbor graph. Let $\mathcal{N}_k^{(v)}(j)$ be the neighbors of query j in readout v, excluding j, and let $\tau_g > 0$ be the graph temperature. The directed edge from query j to query l is

$$W_{jl}^{(v)} = \frac{\mathbf{1}[l \in \mathcal{N}_k^{(v)}(j)]\exp((q_j^{(v)})^\top q_l^{(v)}/\tau_g)}{\sum_{a\in\mathcal{N}_k^{(v)}(j)}\exp((q_j^{(v)})^\top q_a^{(v)}/\tau_g)}. \tag{6}$$

Here $\mathbf{1}[\cdot]$ is the indicator function. We symmetrize each readout graph and retain relations supported by both readouts:

$$G^{(v)} = \mathrm{RowNorm}\left(\frac{W^{(v)} + (W^{(v)})^\top}{2}\right), \qquad G = \mathrm{RowNorm}\left(\sqrt{G^{(1)} \odot G^{(2)}}\right), \tag{7}$$

where RowNorm divides each nonzero row by its sum, $\odot$ is entry-wise multiplication, and the square root is entry-wise. If a query has no shared edge, its row is replaced by the corresponding row of $(G^{(1)} + G^{(2)})/2$ before normalization. The resulting row-stochastic matrix G encodes the support-adapted query geometry.

3.5 Reliability-gated recurrence

The brain integrates multiple sensory cues according to their confidence, relying more on contextual evidence when the current percept is uncertain. Motivated by this reliability-aware integration, we design the reliability gate to use prediction ambiguity and cross-readout disagreement to regulate each query's access to graph evidence.

The gate measures whether each query has a stable initial prediction. For class distributions $u, v \in [0,1]^C$, define entropy $H(u) = -\sum_{c=1}^C u_c \log u_c$ and divergence $\mathrm{KL}(u\|v) = \sum_{c=1}^C u_c \log(u_c/v_c)$. For query j, let $P_{0,j}$ be row j of P_0 and $\overline{P}_j = (P_j^{(1)} + P_j^{(2)})/2$. We measure anchor ambiguity $a_j \in [0,1]$ and cross-readout disagreement $d_j \in [0,1]$ as (Shannon, 1948; Lin, 1991)

$$a_j = \frac{H(P_{0,j})}{\log C}, \qquad d_j = \sqrt{\frac{\frac{1}{2}\,\mathrm{KL}(P_j^{(1)}\|\overline{P}_j) + \frac{1}{2}\,\mathrm{KL}(P_j^{(2)}\|\overline{P}_j)}{\log 2}}. \tag{8}$$

Given a disagreement weight $\lambda \in [0,1]$, gate exponent $p > 0$, and recurrence bounds $0 \le r_{\min} \le r_{\max} < 1$, the combined uncertainty s_j and recurrence weight r_j are

$$s_j = \mathrm{clip}((1-\lambda)a_j + \lambda d_j, 0, 1)\,, \qquad r_j = r_{\min} + (r_{\max} - r_{\min})s_j^p, \qquad R = \mathrm{diag}(r_1, \ldots, r_m). \tag{9}$$

Here $\mathrm{clip}(u, 0, 1)$ truncates u to $[0,1]$. A larger r_j gives query j greater access to graph evidence, while a smaller value keeps its prediction closer to the support-derived anchor.

Perceptual decisions are gradually refined through the interaction between retained knowledge and newly accumulated sensory evidence. Inspired by this recurrent integration process, BIDETA iteratively combines the stable support-derived anchor with reliability-controlled graph information, refining query predictions while retaining their initial class evidence. Starting from $P^{[0]} = P_0$, BIDETA combines the anchor, spectral graph, and reliability gate for $T \in \mathbb{N}_+$ iterations:

$$P^{[t+1]} = \mathcal{N}_\pi\left[(I_m - R)P_0 + RGP^{[t]}\right], \qquad t = 0, \ldots, T-1, \tag{10}$$

where $P^{[t]} \in [0,1]^{m \times C}$ is the query probability matrix at iteration t and I_m is the $m \times m$ identity matrix. The first term restores support-supervised evidence, while the second propagates neighborhood evidence in proportion to each query's reliability weight.

The operator $\mathcal{N}_\pi$ clips its nonnegative input below 10^{-8} and applies five alternating column and row normalizations (Cuturi, 2013). For an intermediate matrix $V \in \mathbb{R}_+^{m \times C}$, each normalization step is

$$V_{jc} \leftarrow V_{jc}\frac{m\pi_c}{\sum_{l=1}^{m} V_{lc} + 10^{-12}}, \qquad V_{jc} \leftarrow \frac{V_{jc}}{\sum_{b=1}^{C} V_{jb} + 10^{-12}}. \tag{11}$$

This operation aligns aggregate query mass with the support prior $\pi = (\pi_1, \ldots, \pi_C)$ while returning each row to a class distribution. After T iterations, classification outputs $\widehat{y}_j = \arg\max_c P_{jc}^{[T]}$, whereas retrieval ranks enrolled identities by sorting $P_{jc}^{[T]}$ over c in descending order. Appendix A.1 provides the complete pseudocode.

4 Experiments

4.1 Experimental setup

Datasets and label budgets. SITR (Gupta et al., 2025) contains seven sensors and 16 household-object classes. We use DIGIT as the source and four GelSight Mini devices, GelSight Hex, and GelSight Wedge as six targets; all are unseen by TVL, while Hex and Wedge are unseen sensor families for Sparsh. From 800 training images per target class, 1%, 5%, and 10% budgets provide 8, 40, and 80 support images, and the 200-image evaluation split gives 3,200 queries per target. TacVerse Shape (Wei et al., 2026) contains seven sensors and nine shape classes. We use GelSightNoMarker as the source and MagicGripper, MagicTac, TacTip, ViTac, and ViTacTip as five targets, all unseen by both backbones. Each sensor–class pair has 300 training, 100 validation, and 100 test images; the three budgets sample 3, 15, and 30 support images per class and use 900 test queries per target. TacQuad (Feng et al., 2025) contains 56 enrolled trial identities, each a 20-frame object-contact sequence, across three RGB sensors. DIGIT is the source, while GelSight Mini and DuraGel are two targets; both are unseen by TVL, and DuraGel is unseen by Sparsh. We sample 2, 4, or 6 support frames per identity at 10%, 20%, and 30%, reserve four frames as a temporal gap, and use the final four frames to form 224 queries per target. Appendix A.2 lists the label spaces and exact splits.

Encoders and sensor exposure. Both TVL ViT-Small and Sparsh-DINO Small remain frozen (Fu et al., 2024; Higuera et al., 2025a). Their architectures follow the Vision Transformer (Dosovitskiy et al., 2021), and Sparsh uses DINO self-distillation (Caron et al., 2021). TVL takes 224×224 RGB images, whereas Sparsh concatenates two RGB frames along the channel dimension. SITR and TacVerse duplicate the same frame for Sparsh, while TacQuad pairs each frame with an earlier neighboring frame from the same trial. TVL was pretrained with DIGIT; Sparsh's pretraining includes DIGIT, GelSight 2017, and GelSight Mini. Appendix A.3 details preprocessing.

Table 1: **Data-efficient classification on six unseen SITR sensors with two pretrained backbones under different label budgets.** Accuracy (%) is averaged across the unknown sensors and reported as mean ± sample standard deviation over three random seeds. I/T denotes inductive/transductive inference. Best results are in bold and second-best results are underlined.

Method	Mode	TVL 1%	TVL 5%	TVL 10%	Sparsh 1%	Sparsh 5%	Sparsh 10%
Frozen backbone	I	7.58 ± 0.14	7.58 ± 0.14	7.58 ± 0.14	6.86 ± 0.33	6.86 ± 0.33	6.86 ± 0.33
Tip-Adapter	I	17.21 ± 1.63	23.51 ± 0.51	28.56 ± 1.37	14.85 ± 0.34	21.90 ± 0.92	22.30 ± 0.04
SimpleShot	I	37.98 ± 0.69	44.47 ± 0.90	44.55 ± 0.62	32.08 ± 2.45	38.64 ± 0.96	39.21 ± 0.81
LaplacianShot	T	38.55 ± 0.38	45.84 ± 1.65	44.11 ± 0.60	32.78 ± 3.52	39.94 ± 1.29	39.87 ± 0.77
SITR-Calib	I	38.00 ± 1.56	46.58 ± 0.08	47.47 ± 1.44	30.88 ± 1.99	34.46 ± 1.00	34.17 ± 0.35
AnyTouch Match	I	36.27 ± 1.55	44.30 ± 0.12	45.32 ± 1.00	32.34 ± 0.99	35.36 ± 0.56	35.81 ± 0.64
CTSRL CSM	I	37.42 ± 1.77	43.87 ± 1.29	44.99 ± 0.85	31.98 ± 1.27	34.94 ± 0.25	34.31 ± 0.39
BIDETA	T	$\mathbf{62.22 \pm 1.11}$	$\mathbf{83.59 \pm 0.81}$	$\mathbf{85.71 \pm 0.12}$	$\mathbf{62.12 \pm 1.43}$	$\mathbf{83.81 \pm 0.80}$	$\mathbf{87.09 \pm 0.61}$

Table 2: **Data-efficient classification on five unseen TacVerse Shape sensors with two pretrained backbones under different label budgets.** Accuracy (%) is averaged across the unknown sensors and reported as mean ± sample standard deviation over three random seeds. I/T denotes inductive/transductive inference. SITR-Calib is unavailable because standard calibration images are absent. Best results are in bold and second-best results are underlined.

Method	Mode	TVL 1%	TVL 5%	TVL 10%	Sparsh 1%	Sparsh 5%	Sparsh 10%
Frozen backbone	I	9.56 ± 0.48	9.56 ± 0.48	9.56 ± 0.48	13.67 ± 0.47	13.67 ± 0.47	13.67 ± 0.47
Tip-Adapter	I	22.79 ± 4.24	39.45 ± 2.15	45.52 ± 0.73	14.24 ± 1.07	25.61 ± 0.74	35.43 ± 0.72
SimpleShot	I	44.57 ± 2.70	51.13 ± 1.73	53.35 ± 0.76	46.61 ± 0.54	53.80 ± 0.91	55.98 ± 0.86
LaplacianShot	T	43.57 ± 4.28	49.66 ± 1.61	51.61 ± 1.26	44.96 ± 2.31	51.47 ± 0.60	52.24 ± 1.17
SITR-Calib	I	–	–	–	–	–	–
AnyTouch Match	I	42.44 ± 1.93	51.61 ± 1.73	55.83 ± 0.37	41.33 ± 1.71	49.23 ± 1.16	49.82 ± 2.12
CTSRL CSM	I	43.26 ± 1.41	53.61 ± 3.41	56.67 ± 0.83	42.44 ± 1.93	49.50 ± 1.82	49.74 ± 1.97
BIDETA	T	$\mathbf{57.93 \pm 1.73}$	$\mathbf{75.47 \pm 2.03}$	$\mathbf{79.13 \pm 1.20}$	$\mathbf{59.66 \pm 2.61}$	$\mathbf{80.65 \pm 0.43}$	$\mathbf{83.58 \pm 1.05}$

Comparison methods. We use Tip-Adapter (Zhang et al., 2022), SimpleShot (Wang et al., 2019), and LaplacianShot (Ziko et al., 2020) as representative frozen-feature data-efficient methods. They transfer support information through cache affinity, normalized class prototypes, or transductive query relations without updating the tactile encoder. We further include three recent tactile cross-sensor adaptation methods: SITR-Calib (Gupta et al., 2025), AnyTouch Match (Feng et al., 2025), and CTSRL CSM (Zhang et al., 2026). These methods retain frozen features while training a lightweight feature module and classification head. Appendix A.4 gives their implementations.

Metrics and experimental settings. Classification uses accuracy and macro-F1 (Sokolova & Lapalme, 2009); ranking uses mean reciprocal rank (MRR) (Craswell, 2009) and recall at one (R@1), following standard ranked-retrieval evaluation (Manning et al., 2008). We report the mean and sample standard deviation over three random seeds. All experiments run on an NVIDIA H800 GPU with 80 GB memory. Appendix A.5 reports the complete implementation and parameters.

4.2 Cross-sensor classification

Table 1 shows that BIDETA improves unseen-sensor performance at every backbone–budget setting on SITR. With TVL and only 1% labeled target data, BIDETA raises accuracy from 7.58% for the frozen source classifier to 62.22%, outperforming the strongest comparison by 23.67 percentage points. At the 10% budget, it raises accuracy from 7.58% to 85.71% and exceeds the strongest comparison by 38.24 percentage points. With Sparsh, BIDETA reaches 62.12%, 83.81%, and 87.09% at the three budgets, outperforming LaplacianShot by 29.34, 43.87, and 47.22 percentage points, respectively. Table 2 evaluates whether BIDETA's performance advantage transfers to another dataset. With TVL at the 10% budget, BIDETA reaches 79.13% accuracy and outperforms the strongest comparison by 22.46 percentage points. With Sparsh, BIDETA achieves 83.58% and improves over the strongest comparison by 27.60 percentage points. These results demonstrate that BIDETA robustly generalizes across datasets and improves unknown-sensor recognition with different pretrained backbones.

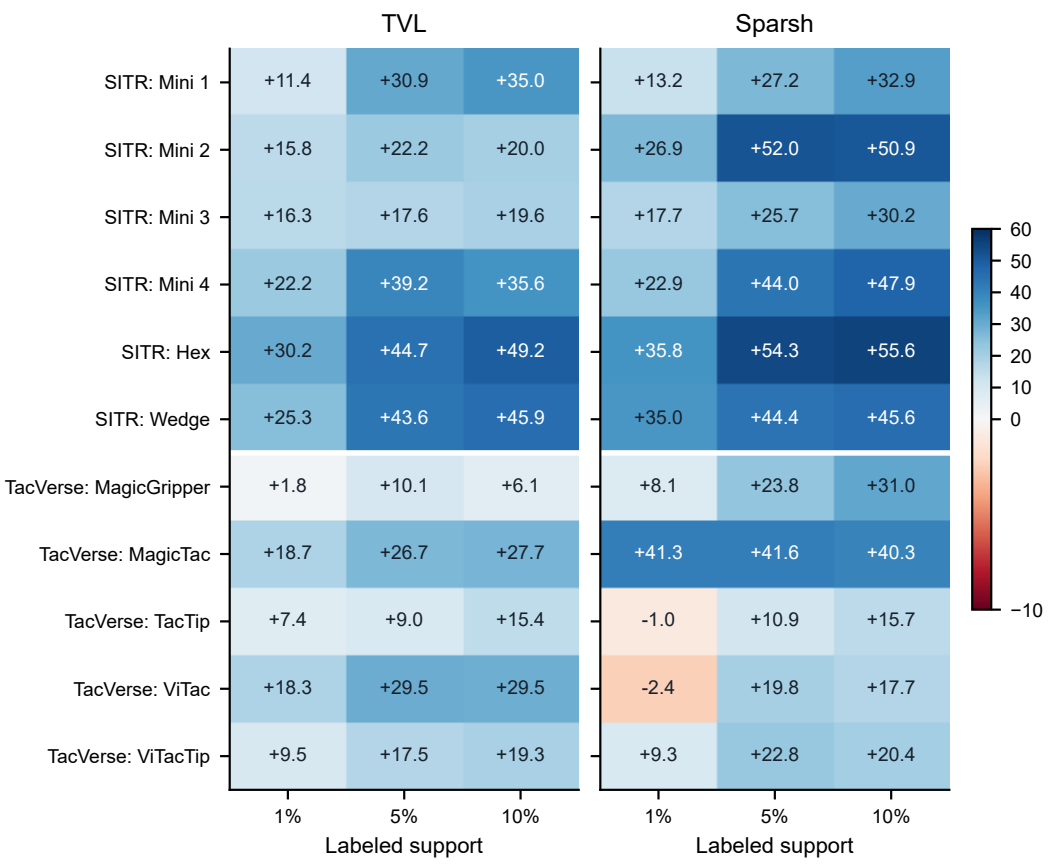


Figure 3: **Performance comparison with the strongest baseline across sensors on SITR and TacVerse.** Each cell reports the accuracy gain of BIDETA for the corresponding unseen sensor and label budget.

Figure 3 reports BIDETA's gain over the strongest comparison for every unseen sensor on SITR and TacVerse with both pretrained backbones. The largest gains occur on GelSight Hex in SITR and MagicTac in TacVerse, reaching 55.6 and 41.6 percentage points, respectively. On TacVerse, negative gains occur only with Sparsh on TacTip and ViTac at the 1% budget. We suspect that the three randomly sampled examples per class may sometimes be unrepresentative, leaving insufficient target variation for BIDETA to reliably recalibrate feature geometry. Additional sensor-wise, class-wise, 2-shot, 5-shot, 10-shot and macro-F1 results are reported in Appendix B.

Table 3: **MRR results for closed-set identity ranking on TacQuad.** MRR is reported as a percentage, with mean ± sample standard deviation over three random seeds. SITR-Support uses the support–source mean difference as its condition. I/T denotes inductive/transductive inference. Best results are in bold and second-best results are underlined.

Method	Mode	TVL 10%	TVL 20%	TVL 30%	Sparsh 10%	Sparsh 20%	Sparsh 30%
Frozen backbone	I	8.76 ± 0.00	8.76 ± 0.00	8.76 ± 0.00	10.01 ± 0.00	10.01 ± 0.00	10.01 ± 0.00
Tip-Adapter	I	36.68 ± 3.70	47.80 ± 12.36	53.80 ± 9.55	45.80 ± 0.69	61.42 ± 1.17	64.57 ± 0.73
SimpleShot	I	$\underline{59.05 \pm 0.78}$	64.65 ± 1.65	66.57 ± 1.06	60.36 ± 0.85	65.83 ± 1.20	68.75 ± 0.78
LaplacianShot	T	45.11 ± 3.03	48.78 ± 2.27	47.64 ± 1.44	43.32 ± 3.10	43.61 ± 2.61	45.82 ± 0.88
SITR-Support	I	57.94 ± 2.00	64.54 ± 0.38	65.49 ± 1.42	61.49 ± 1.62	67.90 ± 0.63	70.72 ± 0.44
AnyTouch Match	I	58.29 ± 1.37	$\underline{66.15 \pm 2.49}$	$\underline{68.49 \pm 1.10}$	$\underline{61.73 \pm 0.93}$	$\underline{67.93 \pm 0.85}$	$\underline{71.00 \pm 0.51}$
CTSRL CSM	I	57.49 ± 2.16	63.88 ± 0.88	65.38 ± 1.35	58.88 ± 0.78	64.99 ± 1.11	68.14 ± 0.23
BIDETA	T	$\mathbf{61.16 \pm 0.95}$	$\mathbf{76.98 \pm 0.20}$	$\mathbf{80.90 \pm 1.89}$	$\mathbf{69.28 \pm 0.72}$	$\mathbf{80.82 \pm 2.67}$	$\mathbf{85.86 \pm 1.80}$

Table 4: **Ablation results on SITR.** Accuracy (%) is reported as mean ± sample standard deviation over three random seeds. Best results are in bold and second-best results are underlined.

Variant	TVL 1%	TVL 5%	TVL 10%	Sparsh 1%	Sparsh 5%	Sparsh 10%
Feature-memory ablations						
Memory: view 1 only	50.37 ± 0.85	72.49 ± 0.99	75.85 ± 0.32	51.65 ± 3.06	75.13 ± 0.23	77.65 ± 0.59
Memory: view 2 only	49.84 ± 1.67	70.82 ± 1.01	76.00 ± 0.15	54.54 ± 1.38	71.38 ± 0.20	77.62 ± 0.82
Dual memory; no graph	51.71 ± 0.88	72.74 ± 1.02	76.07 ± 0.23	52.88 ± 2.88	74.56 ± 0.20	77.92 ± 0.80
Graph-geometry ablations						
+ Consensus recurrence	56.42 ± 1.00	77.41 ± 0.57	80.10 ± 0.21	55.53 ± 2.57	78.62 ± 0.63	80.86 ± 0.67
+ Support standardization	57.01 ± 1.02	77.99 ± 0.58	80.49 ± 0.21	56.18 ± 2.50	79.19 ± 0.63	81.51 ± 0.63
+ Spectral normalization	58.73 ± 1.11	81.48 ± 0.77	82.68 ± 0.11	59.17 ± 2.25	81.70 ± 0.90	84.14 ± 0.83
Full: arithmetic graph fusion	61.76 ± 1.17	83.24 ± 0.79	$\underline{85.70 \pm 0.15}$	$\underline{61.89 \pm 1.46}$	83.61 ± 0.78	86.88 ± 0.67
Full: no spectral normalization	58.04 ± 1.02	76.53 ± 0.44	81.50 ± 0.11	56.27 ± 2.10	77.74 ± 0.34	82.29 ± 0.75
Reliability-gate ablations						
Full: disagreement only	$\mathbf{62.22 \pm 1.11}$	$\underline{83.53 \pm 0.81}$	$\mathbf{85.71 \pm 0.12}$	$\mathbf{62.12 \pm 1.43}$	83.79 ± 0.79	$\mathbf{87.09 \pm 0.64}$
Full: entropy only	$\underline{62.04 \pm 1.07}$	$\mathbf{83.59 \pm 0.80}$	$\underline{85.70 \pm 0.22}$	61.80 ± 1.46	$\mathbf{83.85 \pm 0.83}$	$\underline{87.02 \pm 0.50}$
Complete framework						
Full BIDETA	$\mathbf{62.22 \pm 1.11}$	$\mathbf{83.59 \pm 0.81}$	$\mathbf{85.71 \pm 0.12}$	$\mathbf{62.12 \pm 1.43}$	$\underline{83.81 \pm 0.80}$	$\mathbf{87.09 \pm 0.61}$

4.3 Closed-set identity ranking

Table 3 reports closed-set identity-ranking performance on TacQuad with different pretrained backbones. Using the TVL backbone, BIDETA improves MRR over the frozen backbone by an average of 64.25 percentage points and over the strongest comparison by an average of 8.45 percentage points. Using Sparsh, BIDETA outperforms the strongest external comparison by 7.55, 12.89, and 14.86 percentage points at the 10%, 20%, and 30% support budgets, respectively. These results demonstrate BIDETA's cross-task gains, validating the framework's adaptability beyond classification. Appendix C reports additional R@1 results.

4.4 Ablation Study

Table 4 reports the component ablations on SITR. Among the sequential component additions, consensus recurrence produces the largest gain, improving accuracy by 2.65–4.71 percentage points across the six backbone–budget settings. By repeatedly aggregating cross-view-consistent evidence over the query graph, it strengthens neighborhood coherence and class discrimination. Within the full model, removing spectral normalization causes the largest performance loss, reducing accuracy by 5.36 percentage points on average and by up to 7.06 points. Its support-conditioned covariance transformation suppresses sensor-dependent variation and aligns target neighborhoods with class structure, making it central to robust cross-sensor propagation. The view-1-only and view-2-only configurations also show a clear decline relative to full BIDETA, indicating that complementary hierarchical representations preserve richer class evidence and stabilize data-efficient adaptation. Together, these components provide complementary functions, and their integration achieves the strongest overall performance across the evaluated settings. Appendix D reports additional macro-F1 ablations, while Appendix E presents the parameter-sensitivity experiments.

5 Conclusion

Tactile foundation models suffer substantial performance degradation when deployed with sensors that differ from those seen during pretraining. To address this problem, we proposed BIDETA, a brain-inspired data-efficient adaptation framework that adapts frozen tactile encoders to unseen sensors through rapid support memory, support-conditioned spectral geometry, and reliability-gated recurrence. Experiments across multiple benchmarks show that BIDETA effectively improves the performance of tactile foundation models on unseen sensors and transfers these gains across datasets, foundation models, and tasks. Future work will extend BIDETA to online robotic manipulation, visual–tactile material understanding, and distributed tactile-skin perception. These advances can promote sensor-agnostic tactile foundation models and address the rapid transfer of embodied agents to newly deployed, heterogeneous tactile hardware.

LARGE LANGUAGE MODEL USAGE STATEMENT

In this work, generative AI tools were used only to improve the language and readability of the manuscript. All research ideas, methodology, algorithm design, implementation, experiments, result analysis, and scientific claims were developed and verified by the authors. The authors reviewed all AI-assisted edits and take full responsibility for the final content of this work.

REFERENCES

Malik Boudiaf, Imtiaz Ziko, Jérôme Rony, Jose Dolz, Pablo Piantanida, and Ismail Ben Ayed. Information Maximization for Few-Shot Learning. In *Advances in Neural Information Processing Systems*, volume 33, pp. 2445–2457. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/196f5641aa9dc87067da4ff90fd81e7b-Paper.pdf.

Matteo Carandini and David J. Heeger. Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, 13(1):51–62, 2012. doi: 10.1038/nrn3136.

Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging Properties in Self-Supervised Vision Transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9650–9660, 2021. URL https://openaccess.thecvf.com/content/ICCV2021/html/Caron_Emerging_Properties_in_Self-Supervised_Vision_Transformers_ICCV_2021_paper.html.

Qingzheng Cong, Steven Oh, Wen Fan, Shan Luo, Kaspar Althoefer, and Dandan Zhang. Taceva: A performance evaluation framework for vision-based tactile sensors. *Advanced Intelligent Systems*, 8(4):e202501179, 2026.

Nick Craswell. Mean Reciprocal Rank. In *Encyclopedia of Database Systems*, pp. 1703. Springer US, 2009. doi: 10.1007/978-0-387-39940-9_488.

Marco Cuturi. Sinkhorn Distances: Lightspeed Computation of Optimal Transport. In *Advances in Neural Information Processing Systems*, volume 26, pp. 2292–2300. Curran Associates, Inc., 2013. URL https://proceedings.neurips.cc/paper_files/paper/2013/file/af21d0c97db2e27e13572cbf59eb343d-Paper.pdf.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=YicbFdNTTy.

Marc O. Ernst and Martin S. Banks. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870):429–433, 2002. doi: 10.1038/415429a.

Ruoxuan Feng, Jiangyu Hu, Wenke Xia, Tianci Gao, Ao Shen, Yuhao Sun, Bin Fang, and Di Hu. AnyTouch: Learning Unified Static-Dynamic Representation across Multiple Visuo-tactile Sensors. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/file/4d893f766ab60e5337659b9e71883af4-Paper-Conference.pdf.

R. A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(2): 179–188, 1936. doi: 10.1111/j.1469-1809.1936.tb02137.x.

Letian Fu, Gaurav Datta, Huang Huang, William Chung-Ho Panitch, Jaimyn Drake, Joseph Ortiz, Mustafa Mukadam, Mike Lambeta, Roberto Calandra, and Ken Goldberg. A Touch, Vision, and Language Dataset for Multimodal Alignment. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 14080–14101. PMLR, 2024. URL https://proceedings.mlr.press/v235/fu24b.html.

Harsh Gupta, Yuchen Mo, Shengmiao Jin, and Wenzhen Yuan. Sensor-Invariant Tactile Representation. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/file/f05a11e682bf9c57ab62f7cff08fbe5b-Paper-Conference.pdf.

Carolina Higuera, Akash Sharma, Chaithanya Krishna Bodduluri, Taosha Fan, Patrick Lancaster, Mrinal Kalakrishnan, Michael Kaess, Byron Boots, Mike Lambeta, Tingfan Wu, and Mustafa Mukadam. Sparsh: Self-supervised touch representations for vision-based tactile sensing. In *Proceedings of The 8th Conference on Robot Learning*, volume 270 of *Proceedings of Machine Learning Research*, pp. 885–915. PMLR, 2025a. URL https://proceedings.mlr.press/v270/higuera25a.html.

Carolina Higuera, Akash Sharma, Taosha Fan, Chaithanya Krishna Bodduluri, Byron Boots, Michael Kaess, Mike Lambeta, Tingfan Wu, Zixi Liu, Francois Robert Hogan, and Mustafa Mukadam. Tactile Beyond Pixels: Multisensory Touch Representations for Robot Manipulation. In *Proceedings of The 9th Conference on Robot Learning*, volume 305 of *Proceedings of Machine Learning Research*, pp. 105–123. PMLR, 2025b. URL https://proceedings.mlr.press/v305/higuera25a.html.

Adilbek Karmanov, Dayan Guan, Shijian Lu, Abdulmotaleb El Saddik, and Eric Xing. Efficient Test-Time Adaptation of Vision-Language Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14162–14171, 2024. URL https://openaccess.thecvf.com/content/CVPR2024/html/Karmanov_Efficient_Test-Time_Adaptation_of_Vision-Language_Models_CVPR_2024_paper.html.

Mikail Khona and Ila R. Fiete. Attractor and integrator networks in the brain. *Nature Reviews Neuroscience*, 23(12):744–766, 2022. doi: 10.1038/s41583-022-00642-0.

Seongyu Kim, Seungwoo Lee, Hyeonggon Ryu, Joon Son Chung, and Arda Senocak. Seeing Through Touch: Tactile-Driven Visual Localization of Material Regions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8717–8726, 2026. URL https://openaccess.thecvf.com/content/CVPR2026/html/Kim_Seeing_Through_Touch_Tactile-Driven_Visual_Localization_of_Material_Regions_CVPR_2026_paper.html.

Mike Lambeta, Po-Wei Chou, Stephen Tian, Brian Yang, Benjamin Maloon, Victoria Rose Most, Dave Stroud, Raymond Santos, Ahmad Byagowi, Gregg Kammerer, Dinesh Jayaraman, and Roberto 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. doi: 10.1109/lra.2020.2977257.

Olivier Ledoit and Michael Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365–411, 2004. doi: 10.1016/s0047-259x(03)00096-4.

Weixian Lei, Yixiao Ge, Kun Yi, Jianfeng Zhang, Difei Gao, Dylan Sun, Yuying Ge, Ying Shan, and Mike Zheng Shou. ViT-Lens: Towards Omni-modal Representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 26647–26657, 2024. URL https://openaccess.thecvf.com/content/CVPR2024/html/Lei_ViT-Lens_Towards_Omni-modal_Representations_CVPR_2024_paper.html.

Jianhua Lin. Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory*, 37(1):145–151, 1991. doi: 10.1109/18.61115.

Ilya Loshchilov and Frank Hutter. Decoupled Weight Decay Regularization. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=Bkg6RiCqY7.

Shan Luo, Nathan F Lepora, Wenzhen Yuan, Kaspar Althoefer, Gordon Cheng, and Ravinder Dahiya. Tactile robotics: An outlook. *IEEE Transactions on Robotics*, 2025.

Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to Information Retrieval*. Cambridge University Press, 2008. URL https://nlp.stanford.edu/IR-book/.

Ségolène Martin, Yunshi Huang, Fereshteh Shakeri, Jean-Christophe Pesquet, and Ismail Ben Ayed. Transductive Zero-Shot and Few-Shot CLIP. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 28816–28826, 2024. URL https://openaccess.thecvf.com/content/CVPR2024/html/Martin_Transductive_Zero-Shot_and_Few-Shot_CLIP_CVPR_2024_paper.html.

James L. McClelland, Bruce L. McNaughton, and Randall C. O'Reilly. Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102(3):419–457, 1995. doi: 10.1037/0033-295x.102.3.419.

C. E. Shannon. A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3): 379–423, 1948. doi: 10.1002/j.1538-7305.1948.tb01338.x.

Akash Sharma, Carolina Higuera, Chaithanya Krishna Bodduluri, Zixi Liu, Taosha Fan, Tess Hellebrekers, Mike Lambeta, Byron Boots, Michael Kaess, Tingfan Wu, Francois Robert Hogan, and Mustafa Mukadam. Self-supervised perception for tactile skin covered dexterous hands. In *Proceedings of The 9th Conference on Robot Learning*, volume 305 of *Proceedings of Machine Learning Research*, pp. 2311–2328. PMLR, 2025. URL https://proceedings.mlr.press/v305/sharma25a.html.

Qianqian Shi, Faqiang Liu, Hongyi Li, Guangyu Li, Luping Shi, and Rong Zhao. Hybrid Neural Networks for Continual Learning Inspired by Corticohippocampal Circuits. *Nature Communications*, 16:1272, 2025. doi: 10.1038/s41467-025-56405-9.

Kunal Pratap Singh, Ali Garjani, Rishubh Singh, Muhammad Uzair Khattak, Jason Toskov, Efe Tarhan, Andrei Atanov, Oğuzhan Kar, and Amir Zamir. Multimodality as Supervision: Self-Supervised Specialization to the Test Environment via Multimodality. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/file/517241fc7c6396bf5694822447008b64-Paper-Conference.pdf.

Marina Sokolova and Guy Lapalme. A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4):427–437, 2009. doi: 10.1016/j.ipm.2009.03.002.

Keshav Srinivasan, Dietmar Plenz, and Michelle Girvan. Boosting Reservoir Computing with Brain-Inspired Adaptive Control of E–I Balance. *Nature Communications*, 16:10212, 2025. doi: 10.1038/s41467-025-64978-8.

Yan Wang, Wei-Lun Chao, Kilian Q. Weinberger, and Laurens van der Maaten. SimpleShot: Revisiting Nearest-Neighbor Classification for Few-Shot Learning. arXiv:1911.04623, 2019. URL https://arxiv.org/abs/1911.04623.

Taylor Webb, Shanka Subhra Mondal, and Ida Momennejad. A Brain-Inspired Agentic Architecture to Improve Planning with LLMs. *Nature Communications*, 16:8633, 2025. doi: 10.1038/s41467-025-63804-5.

Lan Wei, Gurmeher Khurana, Sirine Bhouri, Wenhao Hong, Zeyuan Xin, Qingzheng Cong, Wen Fan, Yanzheng Xiang, and Dandan Zhang. TacVerse: A Multi-Sensor Dataset and Benchmark for Cross-Sensor Vision-Based Tactile Perception. arXiv:2606.25877, 2026. URL https://arxiv.org/abs/2606.25877.

Fengyu Yang, Chenyang Ma, Jiacheng Zhang, Jing Zhu, Wenzhen Yuan, and Andrew Owens. Touch and Go: Learning from Human-Collected Vision and Touch. In *Advances in Neural Information Processing Systems*, volume 35, pp. 8081–8103. Curran Associates, Inc., 2022. doi: 10.52202/068431-0587. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/354892587fe39b17c2b727af02abff4a-Paper-Datasets_and_Benchmarks.pdf.

Fengyu Yang, Chao Feng, Ziyang Chen, Hyoungseob Park, Daniel Wang, Yiming Dou, Ziyao Zeng, Xien Chen, Rit Gangopadhyay, Andrew Owens, and Alex Wong. Binding Touch to Everything: Learning Unified Multimodal Tactile Representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 26340–26353, 2024. URL https://openaccess.thecvf.com/content/CVPR2024/html/Yang_Binding_Touch_to_Everything_Learning_Unified_Multimodal_Tactile_Representations_CVPR_2024_paper.html.

Wenzhen Yuan, Siyuan Dong, and Edward H. Adelson. GelSight: High-Resolution Robot Tactile Sensors for Estimating Geometry and Force. *Sensors*, 17(12):2762, 2017. doi: 10.3390/s17122762.

Anthony Zador, Sean Escola, Blake Richards, Bence Ölveczky, Yoshua Bengio, Kwabena Boahen, Matthew Botvinick, Dmitri Chklovskii, Anne Churchland, Claudia Clopath, James DiCarlo, Surya Ganguli, Jeff Hawkins, Konrad Körding, Alexei Koulakov, Yann LeCun, Timothy Lillicrap, Adam Marblestone, Bruno Olshausen, Alexandre Pouget, Cristina Savin, Terrence Sejnowski, Eero Simoncelli, Sara Solla, David Sussillo, Andreas S. Tolias, and Doris Tsao. Catalyzing Next-Generation Artificial Intelligence through NeuroAI. *Nature Communications*, 14:1597, 2023. doi: 10.1038/s41467-023-37180-x.

Renrui Zhang, Wei Zhang, Rongyao Fang, Peng Gao, Kunchang Li, Jifeng Dai, Yu Qiao, and Hongsheng Li. Tip-Adapter: Training-Free Adaption of CLIP for Few-Shot Classification. In *Computer Vision – ECCV 2022*, volume 13695 of *Lecture Notes in Computer Science*, pp. 493–510. Springer, 2022. doi: 10.1007/978-3-031-19833-5_29. URL https://www.ecva.net/papers/eccv_2022/papers_ECCV/html/154_ECCV_2022_paper.php.

Yan Zhang, Zheng Wang, Pengpeng Zeng, Xing Xu, Jingkuan Song, and Heng Tao Shen. Cross-Tactile Sensor Representation Learning. In *International Conference on Machine Learning*, 2026. URL https://icml.cc/virtual/2026/poster/66793.

Dengyong Zhou, Olivier Bousquet, Thomas N. Lal, Jason Weston, and Bernhard Schölkopf. Learning with Local and Global Consistency. In *Advances in Neural Information Processing Systems*, volume 16, pp. 321–328. MIT Press, 2003. URL https://proceedings.neurips.cc/paper_files/paper/2003/file/87682805257e619d49b8e0dfdc14affa-Paper.pdf.

Imtiaz Ziko, Jose Dolz, Eric Granger, and Ismail Ben Ayed. Laplacian Regularized Few-Shot Learning. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 11660–11670. PMLR, 2020. URL https://proceedings.mlr.press/v119/ziko20a.html.

A Protocol and implementation details

A.1 Algorithm Procedure

Algorithm 1 BIDETA adaptation for one target sensor

Require: Frozen encoder f_θ, labeled support $\mathcal{S}$, unlabeled queries $\mathcal{Q}$, and fixed hyperparameters

Extract two support and query readouts with f_θ
for $v \in \{1, 2\}$ **do**
    Fit the support scaler and shrinkage-LDA memory
    Compute query probabilities $P^{(v)}$ and spectral transform $A^{(v)}$
    Transform query features and construct the readout graph $G^{(v)}$
end for
Fuse the two memories into P_0 and the two graphs into G
Compute the reliability gate R and initialize $P \leftarrow P_0$
for $t = 1, \ldots, T$ **do**
    $P \leftarrow \mathcal{N}_\pi[(I_m - R)P_0 + RGP]$
end for
return class predictions or identity rankings from P

A.2 Dataset splits and sample accounting

SITR's original classification archive contains 20 classes across seven sensors. We follow the supplied 16-class subset, mapped consecutively to the classifier label space. This yields 112,000 images: 12,800 training and 3,200 validation images per sensor. The six target episodes each use 128, 640, or 1,280 support images. The source is DIGIT, and targets correspond to Mini_1, Mini_2, Mini_3, Mini_4, Hex, and Wedge. The train-only inner split is distinct from this evaluation set; the final support sample is drawn from the full 800-per-class training pool.

The TacVerse Shape archive contains 30,094 images across seven sensors. The experiment uses the source plus five targets specified in the main text; GelSightMarker is not part of the five-target average. The nine labels are column, cuboid, dots, edge, hexagon, moon, ring, triangles, and wave. For the included sensor–class pairs, the 300/100/100 ordered split yields 2,700 training images, 900 validation images, and 900 test images per sensor. The support sizes are 27, 135, and 270 per target.

The final TacQuad task retains 56 trial identities with complete 20-frame sequences across all three RGB sensors. The force-field sensor Tac3D is excluded from the RGB encoder experiment. The source contains 12 frames per identity (672 total); target support contains 112/224/336 frames, and each target query set contains 224 frames. The support and query positions are disjoint within the sampled sequence, but they share identity and can share a physical trial. Table 5 summarizes the label spaces used by all three benchmarks.

A.3 Preprocessing and Feature Extraction

SITR and TacVerse images are converted to RGB and resized to 224×224 with PIL bicubic interpolation. TVL uses channel means $(0.291746, 0.297133, 0.291040)$ and standard deviations $(0.187645, 0.194677, 0.218716)$ after conversion to $[0, 1]$. Sparsh uses no additional channel normalization and concatenates the image with itself to obtain six channels. TacQuad uses torchvision resizing with bilinear interpolation and antialiasing.

The TVL wrapper loads the tactile-encoder state from the released checkpoint and returns global average-pooled and projected features. The first readout is not a penultimate transformer block or a CLS token. For Sparsh, the wrapper mean-pools normalized patch tokens from the final two blocks, excluding register tokens. The archived frozen parameter counts are 21,961,344 for TVL and 21,894,144 for Sparsh. Encoders run in evaluation mode without gradient updates.

Table 5: **Label spaces used by the three benchmarks.** The table summarizes the classification labels and enrolled identity definition used in our experiments.

Dataset	Classes or identities used
SITR	16 household-object classes selected from the 20-object classification release.
TacVerse Shape	Nine local-shape classes: column, cuboid, dots, edge, hexagon, moon, ring, triangles, and wave.
TacQuad	56 trial-folder identities. Each identity denotes one enrolled object-contact trial represented by a 20-frame sequence for each RGB sensor.

Table 6: Baseline implementations.

Method	Readout	Target fitting	Joint query inference
Frozen backbone	Final	None	No
Tip-Adapter	Final	Support cache	No
SimpleShot	Earlier	Class prototypes	No
LaplacianShot	Earlier	Class prototypes	Yes
SITR-Calib	Final	Lightweight head	No
AnyTouch Match	Final	Lightweight head	No
CTSRL CSM	Final	Lightweight head	No
BIDETA	Both	Statistical memory and graph	Yes

Table 7: **Hyperparameter configurations used in the experiments.** SITR and TacVerse share one configuration, while TacQuad uses a task-specific configuration.

Hyperparameter	Value
SITR and TacVerse	
Spectral exponent (γ)	0.6
Neighborhood size (k)	40
Graph temperature (τ_g)	0.2
Recurrence range ($r_{\min}$–$r_{\max}$)	0.70–0.90
Iterations (T)	10
Disagreement weight (λ)	1.0
Gate exponent (p)	0.05
TacQuad	
Spectral exponent (γ)	0.25
Neighborhood size (k)	80
Graph temperature (τ_g)	0.07
Recurrence range ($r_{\min}$–$r_{\max}$)	0.60–0.80
Iterations (T)	20
Disagreement weight (λ)	0.50
Gate exponent (p)	0.15

A.4 Baseline Implementations

Methods with lightweight classification heads keep the tactile encoder frozen and optimize compact feature modules and classifiers with AdamW (Loshchilov & Hutter, 2019), class-balanced target sampling, target cross-entropy, source-prototype preservation, and within-class compactness. The shared settings use a weight decay of 0.0005, at most 300 epochs, and 16 sampled examples per class in each training step. Checkpoints are selected using the training loss. Frozen-feature methods use normalized source or support features, class prototypes, support caches, or query graphs without updating the backbone.

A.5 Experimental Settings and Hyperparameters

SITR and TacVerse use the same hyperparameter configuration, while TacQuad uses a separate configuration because it evaluates identity ranking. Table 7 lists the fixed values used for these two settings.

B Classification Results

B.1 Per-Sensor Accuracy Results

Tables 8–11 report all main accuracy results by sensor. The gains generally increase with the support budget, and BIDETA achieves the strongest overall performance across sensors and backbones, showing that support-conditioned memory and graph inference effectively adapt pretrained representations to sensor-specific shifts.

Table 8: **Per-sensor accuracy on SITR with TVL.** Results (%) are reported as mean ± sample standard deviation over three random seeds. Best results are in bold and second-best results are underlined.

Method	Labels	Mini 1	Mini 2	Mini 3	Mini 4	Hex	Wedge	Mean
Frozen backbone	0%	9.27 ± 0.81	7.89 ± 0.13	6.62 ± 0.16	6.10 ± 0.05	6.22 ± 0.05	9.39 ± 0.25	7.58 ± 0.14
Tip-Adapter	1%	12.40 ± 5.43	24.88 ± 6.23	37.03 ± 1.09	13.27 ± 0.21	6.22 ± 0.05	9.48 ± 0.34	17.21 ± 1.63
Tip-Adapter	5%	23.33 ± 0.24	31.39 ± 1.26	45.15 ± 0.21	23.54 ± 2.32	6.25 ± 0.00	11.43 ± 3.16	23.51 ± 0.51
Tip-Adapter	10%	25.74 ± 2.32	34.50 ± 0.68	44.39 ± 1.87	28.26 ± 0.82	7.54 ± 1.12	30.96 ± 4.38	28.56 ± 1.37
SimpleShot	1%	33.14 ± 3.93	32.93 ± 3.49	50.80 ± 1.01	34.61 ± 5.27	34.28 ± 2.15	42.15 ± 0.85	37.98 ± 0.69
SimpleShot	5%	41.25 ± 2.24	41.29 ± 1.32	58.63 ± 1.52	41.00 ± 1.81	37.88 ± 0.68	46.75 ± 1.26	44.47 ± 0.90
SimpleShot	10%	39.81 ± 0.89	43.28 ± 2.18	58.11 ± 2.02	40.65 ± 1.17	37.27 ± 0.95	48.19 ± 0.46	44.55 ± 0.62
LaplacianShot	1%	29.73 ± 3.37	36.61 ± 4.43	55.74 ± 1.39	32.80 ± 5.88	35.30 ± 2.63	41.08 ± 1.12	38.55 ± 0.38
LaplacianShot	5%	42.02 ± 3.48	44.86 ± 0.83	61.02 ± 2.57	40.72 ± 1.37	38.05 ± 1.47	48.36 ± 1.55	45.84 ± 1.65
LaplacianShot	10%	37.23 ± 2.03	44.25 ± 3.34	60.97 ± 3.64	38.07 ± 0.83	37.11 ± 2.06	47.02 ± 1.50	44.11 ± 0.60
SITR-Calib	1%	40.03 ± 1.89	36.83 ± 2.08	55.78 ± 1.54	39.97 ± 4.84	17.81 ± 0.20	37.56 ± 3.98	38.00 ± 1.56
SITR-Calib	5%	46.85 ± 3.45	56.40 ± 0.97	68.15 ± 4.64	41.46 ± 1.85	23.66 ± 3.72	42.95 ± 1.47	46.58 ± 0.08
SITR-Calib	10%	46.10 ± 1.75	64.46 ± 1.31	68.43 ± 2.26	41.97 ± 1.88	21.19 ± 6.03	42.65 ± 2.22	47.47 ± 1.44
AnyTouch Match	1%	41.98 ± 1.81	35.20 ± 2.10	55.30 ± 1.44	38.81 ± 5.66	6.03 ± 1.22	40.27 ± 2.70	36.27 ± 1.55
AnyTouch Match	5%	47.41 ± 2.19	56.49 ± 0.66	70.41 ± 2.61	42.44 ± 1.49	6.58 ± 0.73	42.50 ± 0.78	44.30 ± 0.12
AnyTouch Match	10%	44.07 ± 3.38	63.97 ± 1.00	68.68 ± 1.64	44.22 ± 1.90	6.07 ± 1.71	44.90 ± 2.89	45.32 ± 1.00
CTSRL CSM	1%	39.93 ± 0.89	36.71 ± 2.04	55.59 ± 2.86	36.60 ± 5.43	17.61 ± 3.08	38.09 ± 0.23	37.42 ± 1.77
CTSRL CSM	5%	44.99 ± 1.42	56.72 ± 1.11	66.64 ± 3.83	41.82 ± 2.05	15.79 ± 2.71	37.28 ± 1.41	43.87 ± 1.29
CTSRL CSM	10%	45.24 ± 1.83	63.53 ± 1.02	65.88 ± 2.70	41.40 ± 1.75	15.72 ± 4.68	38.21 ± 3.24	44.99 ± 0.85
BIDETA	1%	$\mathbf{53.39 \pm 3.20}$	$\mathbf{52.66 \pm 0.65}$	$\mathbf{72.06 \pm 3.99}$	$\mathbf{62.18 \pm 1.84}$	$\mathbf{65.55 \pm 4.27}$	$\mathbf{67.47 \pm 2.32}$	$\mathbf{62.22 \pm 1.11}$
BIDETA	5%	$\mathbf{78.33 \pm 1.77}$	$\mathbf{78.88 \pm 1.06}$	$\mathbf{88.05 \pm 2.62}$	$\mathbf{81.63 \pm 1.21}$	$\mathbf{82.72 \pm 1.89}$	$\mathbf{91.93 \pm 1.58}$	$\mathbf{83.59 \pm 0.81}$
BIDETA	10%	$\mathbf{81.13 \pm 0.83}$	$\mathbf{84.41 \pm 2.67}$	$\mathbf{88.25 \pm 0.94}$	$\mathbf{79.85 \pm 1.99}$	$\mathbf{86.50 \pm 3.35}$	$\mathbf{94.10 \pm 0.83}$	$\mathbf{85.71 \pm 0.12}$

B.2 Macro-F1 and Sensor Dependence

Macro-F1 assigns equal weight to every class and complements accuracy. Tables 12–15 report the corresponding Macro-F1 results. BIDETA remains the strongest overall method on both datasets; compared with accuracy, Macro-F1 reveals a more balanced advantage across classes, indicating greater robustness to class-dependent sensor shifts.

B.3 Unified Shot Evaluation

Each shot denotes the number of labeled support images per target class under the unified 2-, 5-, and 10-shot protocols. Within each random seed, the support sets are nested as 2-shot $\subset$ 5-shot $\subset$ 10-shot. Tables 16–19 report the corresponding classification results.

Table 9: **Per-sensor accuracy on SITR with Sparsh.** Results (%) are reported as mean ± sample standard deviation over three random seeds. Best results are in bold and second-best results are underlined.

Method	Labels	Mini 1	Mini 2	Mini 3	Mini 4	Hex	Wedge	Mean
Frozen backbone	0%	6.25 ± 0.00	6.25 ± 0.00	6.25 ± 0.00	6.25 ± 0.00	6.26 ± 0.02	9.88 ± 1.99	6.86 ± 0.33
Tip-Adapter	1%	6.86 ± 1.06	27.56 ± 1.70	32.04 ± 2.51	6.25 ± 0.00	6.26 ± 0.02	10.13 ± 1.96	14.85 ± 0.34
Tip-Adapter	5%	20.83 ± 0.46	30.58 ± 2.86	38.05 ± 1.81	23.00 ± 2.89	6.27 ± 0.04	12.67 ± 0.69	21.90 ± 0.92
Tip-Adapter	10%	20.40 ± 0.93	34.23 ± 1.76	36.54 ± 1.19	22.44 ± 0.74	6.58 ± 0.58	13.59 ± 0.57	22.30 ± 0.04
SimpleShot	1%	25.53 ± 2.51	32.70 ± 3.68	37.35 ± 4.25	28.97 ± 5.44	29.85 ± 2.83	38.07 ± 1.88	32.08 ± 2.45
SimpleShot	5%	33.40 ± 2.23	36.18 ± 2.78	46.02 ± 1.49	35.98 ± 2.85	34.81 ± 4.33	45.46 ± 0.30	38.64 ± 0.96
SimpleShot	10%	31.57 ± 1.07	41.28 ± 3.01	43.52 ± 0.13	35.80 ± 1.72	37.23 ± 0.51	45.84 ± 0.66	39.21 ± 0.81
LaplacianShot	1%	25.30 ± 3.18	35.99 ± 5.79	39.10 ± 6.45	29.20 ± 6.45	29.82 ± 5.30	37.26 ± 2.77	32.78 ± 3.52
LaplacianShot	5%	32.82 ± 2.79	40.30 ± 2.88	47.49 ± 1.85	38.13 ± 4.49	34.68 ± 4.93	46.20 ± 0.69	39.94 ± 1.29
LaplacianShot	10%	31.72 ± 1.14	44.97 ± 4.40	43.71 ± 1.37	37.30 ± 3.74	36.08 ± 0.48	45.43 ± 0.34	39.87 ± 0.77
SITR-Calib	1%	33.19 ± 5.49	35.42 ± 1.87	44.82 ± 4.49	28.85 ± 1.40	13.96 ± 1.66	29.01 ± 2.12	30.88 ± 1.99
SITR-Calib	5%	36.48 ± 0.10	36.52 ± 3.99	52.08 ± 6.81	35.95 ± 0.81	14.70 ± 2.21	31.02 ± 0.44	34.46 ± 1.00
SITR-Calib	10%	36.07 ± 1.39	39.01 ± 1.25	50.68 ± 2.88	36.67 ± 2.00	13.41 ± 0.61	29.17 ± 2.82	34.17 ± 0.35
AnyTouch Match	1%	38.15 ± 4.88	36.82 ± 2.23	45.34 ± 4.25	32.96 ± 0.97	11.66 ± 2.00	29.09 ± 2.72	32.34 ± 0.99
AnyTouch Match	5%	40.17 ± 1.30	41.64 ± 2.01	52.00 ± 0.97	37.58 ± 1.82	12.28 ± 1.26	28.51 ± 4.11	35.36 ± 0.56
AnyTouch Match	10%	38.60 ± 2.37	43.94 ± 2.02	54.44 ± 4.56	38.32 ± 0.35	12.94 ± 1.73	26.63 ± 4.75	35.81 ± 0.64
CTSRL CSM	1%	34.83 ± 1.75	34.59 ± 2.57	44.85 ± 3.85	30.23 ± 1.27	16.40 ± 1.02	30.96 ± 1.13	31.98 ± 1.27
CTSRL CSM	5%	37.95 ± 0.23	38.85 ± 1.38	54.39 ± 2.85	34.11 ± 1.63	14.22 ± 1.30	30.10 ± 1.23	34.94 ± 0.25
CTSRL CSM	10%	36.16 ± 0.69	38.51 ± 1.95	51.66 ± 2.16	35.16 ± 1.34	13.70 ± 1.94	30.70 ± 1.81	34.31 ± 0.39
BIDETA	1%	$\mathbf{51.34 \pm 2.86}$	$\mathbf{63.73 \pm 3.73}$	$\mathbf{63.06 \pm 4.50}$	$\mathbf{55.88 \pm 3.19}$	$\mathbf{65.64 \pm 4.47}$	$\mathbf{73.05 \pm 2.49}$	$\mathbf{62.12 \pm 1.43}$
BIDETA	5%	$\mathbf{67.35 \pm 2.36}$	$\mathbf{93.66 \pm 1.13}$	$\mathbf{80.07 \pm 2.04}$	$\mathbf{82.10 \pm 1.13}$	$\mathbf{89.10 \pm 1.17}$	$\mathbf{90.59 \pm 1.69}$	$\mathbf{83.81 \pm 0.80}$
BIDETA	10%	$\mathbf{71.52 \pm 2.61}$	$\mathbf{95.89 \pm 0.10}$	$\mathbf{84.66 \pm 0.81}$	$\mathbf{86.21 \pm 1.11}$	$\mathbf{92.88 \pm 0.81}$	$\mathbf{91.42 \pm 1.17}$	$\mathbf{87.09 \pm 0.61}$

Table 10: **Per-sensor accuracy on TacVerse Shape with TVL.** Results (%) are reported as mean ± sample standard deviation over three random seeds. Best results are in bold and second-best results are underlined. SITR-Calib is unavailable because standard calibration images are absent.

Method	Labels	MagicGripper	MagicTac	TacTip	ViTac	ViTacTip	Mean
Frozen backbone	0%	8.67 ± 0.38	6.48 ± 1.80	10.41 ± 1.00	11.11 ± 0.00	11.11 ± 0.00	9.56 ± 0.48
Tip-Adapter	1%	23.07 ± 6.55	30.26 ± 13.45	10.74 ± 1.12	26.56 ± 14.11	23.30 ± 2.26	22.79 ± 4.24
Tip-Adapter	5%	39.78 ± 3.84	52.07 ± 0.84	17.48 ± 9.36	43.37 ± 1.99	44.56 ± 0.69	39.45 ± 2.15
Tip-Adapter	10%	42.41 ± 3.81	51.19 ± 0.53	28.93 ± 0.84	43.81 ± 1.53	61.26 ± 1.57	45.52 ± 0.73
SimpleShot	1%	33.78 ± 5.61	46.59 ± 4.33	29.56 ± 2.19	55.00 ± 6.47	57.93 ± 7.47	44.57 ± 2.70
SimpleShot	5%	40.78 ± 3.72	47.67 ± 0.19	33.78 ± 1.64	64.00 ± 2.98	69.41 ± 3.49	51.13 ± 1.73
SimpleShot	10%	48.85 ± 3.04	48.30 ± 1.67	36.11 ± 0.40	64.22 ± 3.61	69.26 ± 2.64	53.35 ± 0.76
LaplacianShot	1%	36.56 ± 8.76	47.52 ± 2.31	26.22 ± 5.46	52.15 ± 6.20	55.41 ± 4.01	43.57 ± 4.28
LaplacianShot	5%	41.85 ± 6.41	46.37 ± 1.68	31.07 ± 4.59	66.30 ± 4.07	62.70 ± 3.95	49.66 ± 1.61
LaplacianShot	10%	49.89 ± 3.01	47.67 ± 2.38	34.00 ± 0.56	65.04 ± 2.81	61.44 ± 4.46	51.61 ± 1.26
SITR-Calib	1%	–	–	–	–	–	–
SITR-Calib	5%	–	–	–	–	–	–
SITR-Calib	10%	–	–	–	–	–	–
AnyTouch Match	1%	29.67 ± 4.45	54.44 ± 11.26	20.59 ± 1.61	52.78 ± 6.44	54.70 ± 8.56	42.44 ± 1.93
AnyTouch Match	5%	41.37 ± 8.66	70.15 ± 3.02	22.41 ± 0.96	60.26 ± 4.72	63.85 ± 4.72	51.61 ± 1.73
AnyTouch Match	10%	50.89 ± 1.49	70.74 ± 2.84	22.48 ± 1.03	68.04 ± 0.23	67.00 ± 2.67	55.83 ± 0.37
CTSRL CSM	1%	30.33 ± 2.99	54.93 ± 14.75	22.56 ± 3.53	54.63 ± 5.14	53.85 ± 8.46	43.26 ± 1.41
CTSRL CSM	5%	42.30 ± 8.79	72.70 ± 2.56	25.19 ± 1.67	62.19 ± 0.93	65.67 ± 6.01	53.61 ± 3.41
CTSRL CSM	10%	50.22 ± 2.83	71.74 ± 0.68	25.33 ± 1.07	65.26 ± 0.97	70.78 ± 0.69	56.67 ± 0.83
BIDETA	1%	$\mathbf{38.41 \pm 6.56}$	$\mathbf{73.59 \pm 7.58}$	$\mathbf{36.93 \pm 2.21}$	$\mathbf{73.30 \pm 3.22}$	$\mathbf{67.44 \pm 3.70}$	$\mathbf{57.93 \pm 1.73}$
BIDETA	5%	$\mathbf{52.44 \pm 5.17}$	$\mathbf{99.44 \pm 0.87}$	$\mathbf{42.74 \pm 1.39}$	$\mathbf{95.85 \pm 0.65}$	$\mathbf{86.89 \pm 4.26}$	$\mathbf{75.47 \pm 2.03}$
BIDETA	10%	$\mathbf{57.04 \pm 6.40}$	$\mathbf{99.48 \pm 0.71}$	$\mathbf{51.52 \pm 3.21}$	$\mathbf{97.52 \pm 1.07}$	$\mathbf{90.11 \pm 1.16}$	$\mathbf{79.13 \pm 1.20}$

Table 11: **Per-sensor accuracy on TacVerse Shape with Sparsh.** Results (%) are reported as mean ± sample standard deviation over three random seeds. Best results are in bold and second-best results are underlined. SITR-Calib is unavailable because standard calibration images are absent.

Method	Labels	MagicGripper	MagicTac	TacTip	ViTac	ViTacTip	Mean
Frozen backbone	0%	10.59 ± 0.63	11.11 ± 0.00	11.11 ± 0.00	17.26 ± 1.58	18.26 ± 1.98	13.67 ± 0.47
Tip-Adapter	1%	13.19 ± 3.39	11.11 ± 0.00	11.11 ± 0.00	17.48 ± 1.45	18.33 ± 2.04	14.24 ± 1.07
Tip-Adapter	5%	43.70 ± 2.84	11.11 ± 0.00	11.11 ± 0.00	31.15 ± 0.93	30.96 ± 0.50	25.61 ± 0.74
Tip-Adapter	10%	43.93 ± 1.62	57.33 ± 2.44	11.11 ± 0.00	32.48 ± 0.51	32.30 ± 0.45	35.43 ± 0.72
SimpleShot	1%	38.56 ± 2.11	48.04 ± 4.45	$\mathbf{40.00 \pm 5.14}$	57.96 ± 6.42	48.48 ± 1.01	46.61 ± 0.54
SimpleShot	5%	48.70 ± 2.41	58.15 ± 2.06	41.00 ± 0.87	68.48 ± 1.10	52.67 ± 1.95	53.80 ± 0.91
SimpleShot	10%	48.85 ± 2.39	59.70 ± 2.45	46.78 ± 2.35	69.93 ± 1.56	54.63 ± 1.17	55.98 ± 0.86
LaplacianShot	1%	38.41 ± 2.93	46.74 ± 5.47	38.93 ± 4.14	56.22 ± 8.11	44.52 ± 4.39	44.96 ± 2.31
LaplacianShot	5%	48.07 ± 3.03	58.37 ± 2.17	41.44 ± 2.70	64.22 ± 0.69	45.26 ± 5.20	51.47 ± 0.60
LaplacianShot	10%	47.56 ± 2.22	58.00 ± 3.07	44.48 ± 2.97	68.81 ± 3.11	42.37 ± 2.45	52.24 ± 1.17
SITR-Calib	1%	–	–	–	–	–	–
SITR-Calib	5%	–	–	–	–	–	–
SITR-Calib	10%	–	–	–	–	–	–
AnyTouch Match	1%	41.15 ± 2.44	45.63 ± 6.30	11.11 ± 0.00	58.22 ± 3.66	50.52 ± 1.69	41.33 ± 1.71
AnyTouch Match	5%	58.00 ± 3.29	54.52 ± 1.12	11.11 ± 0.00	64.37 ± 1.83	58.15 ± 2.02	49.23 ± 1.16
AnyTouch Match	10%	58.74 ± 3.11	57.70 ± 3.24	11.11 ± 0.00	66.11 ± 4.03	55.44 ± 4.35	49.82 ± 2.12
CTSRL CSM	1%	40.67 ± 6.30	44.74 ± 6.53	12.96 ± 0.65	$\mathbf{61.15 \pm 3.73}$	52.67 ± 1.95	42.44 ± 1.93
CTSRL CSM	5%	53.96 ± 5.11	52.89 ± 1.15	14.81 ± 3.61	68.04 ± 1.34	57.81 ± 1.39	49.50 ± 1.82
CTSRL CSM	10%	56.63 ± 3.11	52.67 ± 3.42	13.59 ± 0.93	68.19 ± 0.34	57.63 ± 3.34	49.74 ± 1.97
BIDETA	1%	$\mathbf{49.22 \pm 8.19}$	$\mathbf{89.37 \pm 5.11}$	39.00 ± 2.44	58.74 ± 2.98	$\mathbf{61.96 \pm 6.10}$	$\mathbf{59.66 \pm 2.61}$
BIDETA	5%	$\mathbf{81.78 \pm 7.62}$	$\mathbf{100.00 \pm 0.00}$	$\mathbf{52.30 \pm 1.88}$	$\mathbf{88.26 \pm 1.98}$	$\mathbf{80.93 \pm 3.37}$	$\mathbf{80.65 \pm 0.43}$
BIDETA	10%	$\mathbf{89.74 \pm 3.26}$	$\mathbf{100.00 \pm 0.00}$	$\mathbf{62.48 \pm 1.80}$	$\mathbf{87.63 \pm 1.62}$	$\mathbf{78.04 \pm 5.65}$	$\mathbf{83.58 \pm 1.05}$

Table 12: **Per-sensor Macro-F1 on SITR with TVL.** Results (%) are reported as mean ± sample standard deviation over three random seeds. Best results are in bold and second-best results are underlined.

Method	Labels	Mini 1	Mini 2	Mini 3	Mini 4	Hex	Wedge	Mean
Frozen backbone	0%	2.16 ± 0.23	5.49 ± 0.11	2.27 ± 0.11	0.81 ± 0.15	0.73 ± 0.00	2.51 ± 0.13	2.33 ± 0.01
Tip-Adapter	1%	6.29 ± 7.33	22.92 ± 6.58	32.88 ± 0.44	5.58 ± 0.49	0.73 ± 0.00	2.55 ± 0.09	11.82 ± 1.98
Tip-Adapter	5%	18.74 ± 0.16	31.02 ± 1.61	41.26 ± 0.83	19.03 ± 0.80	0.74 ± 0.00	5.99 ± 5.79	19.46 ± 0.66
Tip-Adapter	10%	21.67 ± 2.95	34.30 ± 0.65	40.36 ± 1.82	23.51 ± 1.63	2.74 ± 1.73	26.07 ± 3.28	24.77 ± 1.57
SimpleShot	1%	31.31 ± 3.60	31.71 ± 3.19	48.72 ± 0.63	32.77 ± 4.18	32.65 ± 2.74	40.07 ± 0.91	36.20 ± 0.88
SimpleShot	5%	39.67 ± 2.49	40.28 ± 1.24	57.92 ± 1.69	40.67 ± 1.04	36.68 ± 0.95	45.18 ± 1.06	43.40 ± 1.12
SimpleShot	10%	37.79 ± 1.47	42.05 ± 2.39	57.29 ± 2.28	40.16 ± 1.21	36.21 ± 1.30	46.70 ± 0.48	43.37 ± 0.74
LaplacianShot	1%	27.49 ± 2.45	33.96 ± 4.01	53.22 ± 3.00	30.26 ± 4.81	32.53 ± 3.67	37.03 ± 0.61	35.75 ± 0.96
LaplacianShot	5%	39.79 ± 3.25	43.25 ± 1.20	60.15 ± 2.82	40.23 ± 1.60	36.03 ± 1.44	45.75 ± 1.34	44.20 ± 1.82
LaplacianShot	10%	34.39 ± 2.44	41.85 ± 3.43	60.05 ± 3.79	37.29 ± 0.75	35.23 ± 3.10	44.07 ± 1.78	42.15 ± 0.78
SITR-Calib	1%	38.88 ± 2.16	36.35 ± 1.68	54.40 ± 1.46	38.50 ± 3.87	13.71 ± 0.48	35.11 ± 3.87	36.16 ± 1.54
SITR-Calib	5%	44.75 ± 4.48	55.76 ± 0.77	67.34 ± 4.99	40.44 ± 1.90	18.11 ± 3.32	39.18 ± 2.87	44.26 ± 0.51
SITR-Calib	10%	43.58 ± 0.24	64.00 ± 1.50	67.73 ± 2.42	40.41 ± 2.59	15.07 ± 6.43	40.99 ± 2.23	45.30 ± 1.70
AnyTouch Match	1%	40.28 ± 1.72	34.27 ± 1.74	54.27 ± 1.58	37.48 ± 4.39	3.61 ± 0.43	38.10 ± 3.69	34.67 ± 1.41
AnyTouch Match	5%	46.13 ± 2.64	55.85 ± 0.47	69.61 ± 2.46	41.65 ± 1.24	3.96 ± 0.13	39.42 ± 1.78	42.77 ± 0.30
AnyTouch Match	10%	42.53 ± 4.37	63.11 ± 1.09	67.91 ± 1.71	43.53 ± 2.11	3.59 ± 0.61	41.68 ± 4.23	43.72 ± 1.64
CTSRL CSM	1%	38.24 ± 1.26	36.09 ± 1.50	54.68 ± 2.48	35.56 ± 4.72	14.38 ± 1.39	34.58 ± 1.43	35.59 ± 1.44
CTSRL CSM	5%	43.99 ± 1.67	56.23 ± 0.98	66.15 ± 3.97	40.41 ± 2.23	12.68 ± 2.63	34.33 ± 2.94	42.30 ± 1.59
CTSRL CSM	10%	44.33 ± 1.11	62.64 ± 1.58	65.42 ± 2.56	40.75 ± 2.33	11.43 ± 3.11	35.53 ± 4.13	43.35 ± 0.57
BIDETA	1%	$\mathbf{52.23 \pm 2.90}$	$\mathbf{52.44 \pm 1.04}$	$\mathbf{71.70 \pm 4.23}$	$\mathbf{61.28 \pm 1.21}$	$\mathbf{63.97 \pm 4.66}$	$\mathbf{67.08 \pm 3.07}$	$\mathbf{61.45 \pm 1.36}$
BIDETA	5%	$\mathbf{77.12 \pm 1.58}$	$\mathbf{78.27 \pm 1.03}$	$\mathbf{87.80 \pm 2.48}$	$\mathbf{80.82 \pm 1.19}$	$\mathbf{81.93 \pm 1.98}$	$\mathbf{91.69 \pm 1.73}$	$\mathbf{82.94 \pm 0.79}$
BIDETA	10%	$\mathbf{80.69 \pm 0.66}$	$\mathbf{84.17 \pm 2.79}$	$\mathbf{88.13 \pm 0.89}$	$\mathbf{79.35 \pm 2.23}$	$\mathbf{86.32 \pm 3.52}$	$\mathbf{94.05 \pm 0.83}$	$\mathbf{85.45 \pm 0.19}$

Table 13: **Per-sensor Macro-F1 on SITR with Sparsh.** Results (%) are reported as mean ± sample standard deviation over three random seeds. Best results are in bold and second-best results are underlined.

Method	Labels	Mini 1	Mini 2	Mini 3	Mini 4	Hex	Wedge	Mean
Frozen backbone	0%	0.74 ± 0.00	0.74 ± 0.00	0.74 ± 0.00	0.74 ± 0.00	0.76 ± 0.04	3.36 ± 0.64	1.18 ± 0.11
Tip-Adapter	1%	1.08 ± 0.60	25.36 ± 3.13	27.30 ± 1.25	0.74 ± 0.00	0.76 ± 0.04	3.46 ± 0.60	9.78 ± 0.49
Tip-Adapter	5%	16.58 ± 0.37	27.62 ± 3.09	32.34 ± 1.62	16.36 ± 2.71	0.78 ± 0.07	4.34 ± 0.16	16.34 ± 1.12
Tip-Adapter	10%	15.23 ± 0.51	30.77 ± 2.41	31.26 ± 1.16	14.33 ± 1.00	1.08 ± 0.60	4.68 ± 0.20	16.23 ± 0.27
SimpleShot	1%	23.80 ± 2.41	31.30 ± 4.30	35.04 ± 3.40	27.74 ± 5.13	28.68 ± 2.80	36.97 ± 2.28	30.59 ± 2.71
SimpleShot	5%	31.79 ± 1.87	34.96 ± 2.79	45.11 ± 0.98	35.02 ± 3.05	33.44 ± 4.00	44.82 ± 0.42	37.52 ± 0.80
SimpleShot	10%	29.31 ± 0.91	39.91 ± 3.24	42.23 ± 0.15	34.46 ± 1.81	36.37 ± 0.95	45.40 ± 1.12	37.94 ± 0.88
LaplacianShot	1%	22.96 ± 2.98	34.88 ± 6.70	36.23 ± 6.12	27.23 ± 5.25	27.96 ± 5.89	35.00 ± 3.63	30.71 ± 3.82
LaplacianShot	5%	31.19 ± 2.78	38.59 ± 3.01	46.49 ± 1.24	36.98 ± 4.63	32.96 ± 4.68	45.76 ± 1.14	38.66 ± 1.20
LaplacianShot	10%	29.27 ± 1.51	42.89 ± 4.45	42.21 ± 1.64	35.55 ± 3.77	34.55 ± 0.81	44.47 ± 0.36	38.16 ± 0.91
SITR-Calib	1%	31.03 ± 5.98	32.01 ± 2.13	42.80 ± 5.61	26.67 ± 1.31	9.02 ± 1.67	24.69 ± 0.84	27.70 ± 1.75
SITR-Calib	5%	33.61 ± 0.20	33.78 ± 3.45	49.96 ± 7.24	32.92 ± 1.57	10.74 ± 2.58	26.82 ± 0.41	31.31 ± 1.04
SITR-Calib	10%	33.36 ± 1.78	35.72 ± 2.07	48.82 ± 3.09	34.31 ± 1.19	8.47 ± 1.10	24.44 ± 2.72	30.85 ± 0.83
AnyTouch Match	1%	35.09 ± 5.25	34.05 ± 2.28	43.52 ± 3.75	31.09 ± 2.19	8.95 ± 1.74	26.76 ± 2.81	29.91 ± 0.60
AnyTouch Match	5%	37.40 ± 0.90	38.43 ± 2.37	50.74 ± 1.41	35.10 ± 2.67	9.01 ± 1.57	24.59 ± 4.96	32.54 ± 0.55
AnyTouch Match	10%	35.87 ± 2.18	40.44 ± 1.29	52.89 ± 5.38	36.32 ± 0.59	9.64 ± 1.69	24.68 ± 4.97	33.31 ± 0.78
CTSRL CSM	1%	32.84 ± 2.40	32.38 ± 3.07	43.11 ± 3.45	28.67 ± 0.45	14.27 ± 1.30	28.08 ± 1.32	29.89 ± 1.12
CTSRL CSM	5%	35.20 ± 0.70	36.42 ± 1.47	52.98 ± 2.31	31.91 ± 2.09	11.98 ± 2.64	26.12 ± 1.71	32.43 ± 0.35
CTSRL CSM	10%	33.62 ± 0.80	36.01 ± 2.03	49.72 ± 2.39	32.99 ± 1.09	11.20 ± 1.54	27.61 ± 1.26	31.86 ± 0.50
BIDETA	1%	$\mathbf{50.16 \pm 2.49}$	$\mathbf{61.49 \pm 4.48}$	$\mathbf{62.41 \pm 5.01}$	$\mathbf{55.37 \pm 3.25}$	$\mathbf{65.14 \pm 4.44}$	$\mathbf{72.42 \pm 2.82}$	$\mathbf{61.16 \pm 1.79}$
BIDETA	5%	$\mathbf{66.04 \pm 2.45}$	$\mathbf{93.45 \pm 1.17}$	$\mathbf{79.73 \pm 1.90}$	$\mathbf{81.70 \pm 1.31}$	$\mathbf{89.05 \pm 1.16}$	$\mathbf{90.47 \pm 1.70}$	$\mathbf{83.41 \pm 0.79}$
BIDETA	10%	$\mathbf{71.21 \pm 2.59}$	$\mathbf{95.80 \pm 0.09}$	$\mathbf{84.25 \pm 0.82}$	$\mathbf{85.91 \pm 1.26}$	$\mathbf{92.84 \pm 0.80}$	$\mathbf{91.20 \pm 1.18}$	$\mathbf{86.87 \pm 0.58}$

Table 14: **Per-sensor Macro-F1 on TacVerse Shape with TVL.** Results (%) are reported as mean ± sample standard deviation over three random seeds. Best results are in bold and second-best results are underlined. SITR-Calib is unavailable because standard calibration images are absent.

Method	Labels	MagicGripper	MagicTac	TacTip	ViTac	ViTacTip	Mean
Frozen backbone	0%	6.91 ± 0.60	2.51 ± 0.35	3.89 ± 0.72	2.22 ± 0.00	2.22 ± 0.00	3.55 ± 0.13
Tip-Adapter	1%	21.20 ± 6.34	23.88 ± 15.70	3.95 ± 0.65	15.72 ± 12.21	12.29 ± 3.31	15.41 ± 4.29
Tip-Adapter	5%	38.41 ± 3.42	46.44 ± 0.51	9.47 ± 8.10	33.06 ± 1.40	34.99 ± 0.38	32.47 ± 1.71
Tip-Adapter	10%	39.73 ± 4.36	46.48 ± 0.44	20.97 ± 0.52	34.40 ± 3.49	59.45 ± 1.23	40.21 ± 0.48
SimpleShot	1%	30.86 ± 6.58	43.25 ± 5.01	26.11 ± 3.10	51.00 ± 7.27	57.09 ± 7.62	41.66 ± 2.31
SimpleShot	5%	40.03 ± 3.05	44.70 ± 0.64	29.40 ± 1.59	61.37 ± 3.13	68.62 ± 3.98	48.82 ± 1.67
SimpleShot	10%	46.83 ± 3.08	45.62 ± 1.86	31.90 ± 0.67	62.01 ± 3.42	68.34 ± 3.09	50.94 ± 0.81
LaplacianShot	1%	31.59 ± 8.16	43.70 ± 3.28	19.57 ± 6.81	43.57 ± 6.93	49.56 ± 3.49	37.60 ± 4.02
LaplacianShot	5%	41.05 ± 5.03	43.13 ± 2.32	23.74 ± 4.55	59.28 ± 5.49	58.25 ± 5.91	45.09 ± 1.82
LaplacianShot	10%	48.12 ± 3.50	44.67 ± 1.86	25.62 ± 1.25	57.82 ± 4.58	55.61 ± 6.29	46.37 ± 1.78
SITR-Calib	1%	–	–	–	–	–	–
SITR-Calib	5%	–	–	–	–	–	–
SITR-Calib	10%	–	–	–	–	–	–
AnyTouch Match	1%	27.54 ± 3.23	50.96 ± 12.71	17.02 ± 1.82	48.27 ± 8.14	52.97 ± 8.30	39.35 ± 1.66
AnyTouch Match	5%	39.04 ± 9.14	67.45 ± 3.64	19.46 ± 1.39	55.47 ± 9.09	61.81 ± 5.03	48.65 ± 1.36
AnyTouch Match	10%	49.34 ± 2.41	68.30 ± 3.57	18.62 ± 1.18	65.39 ± 0.66	65.60 ± 3.38	53.45 ± 0.35
CTSRL CSM	1%	28.19 ± 2.03	52.24 ± 16.38	18.79 ± 3.60	49.85 ± 7.48	52.37 ± 8.66	40.29 ± 1.28
CTSRL CSM	5%	40.23 ± 8.47	70.22 ± 4.02	22.41 ± 3.50	56.99 ± 1.85	64.59 ± 6.66	50.89 ± 4.24
CTSRL CSM	10%	49.03 ± 3.48	69.68 ± 1.48	21.20 ± 1.44	61.80 ± 1.65	70.24 ± 1.34	54.39 ± 0.66
BIDETA	1%	$\mathbf{36.53 \pm 3.79}$	$\mathbf{71.47 \pm 8.13}$	$\mathbf{35.67 \pm 2.94}$	$\mathbf{71.06 \pm 2.09}$	$\mathbf{67.32 \pm 4.00}$	$\mathbf{56.41 \pm 1.84}$
BIDETA	5%	$\mathbf{51.88 \pm 5.09}$	$\mathbf{99.44 \pm 0.87}$	$\mathbf{41.12 \pm 1.28}$	$\mathbf{95.77 \pm 0.68}$	$\mathbf{86.93 \pm 4.44}$	$\mathbf{75.03 \pm 1.98}$
BIDETA	10%	$\mathbf{56.32 \pm 6.90}$	$\mathbf{99.48 \pm 0.71}$	$\mathbf{50.39 \pm 3.48}$	$\mathbf{97.49 \pm 1.09}$	$\mathbf{90.22 \pm 1.30}$	$\mathbf{78.78 \pm 1.15}$

Table 15: **Per-sensor Macro-F1 on TacVerse Shape with Sparsh.** Results (%) are reported as mean ± sample standard deviation over three random seeds. Best results are in bold and second-best results are underlined. SITR-Calib is unavailable because standard calibration images are absent.

Method	Labels	MagicGripper	MagicTac	TacTip	ViTac	ViTacTip	Mean
Frozen backbone	0%	5.47 ± 0.45	2.22 ± 0.00	2.22 ± 0.00	6.58 ± 0.80	6.68 ± 1.00	4.64 ± 0.19
Tip-Adapter	1%	7.03 ± 1.78	2.22 ± 0.00	2.22 ± 0.00	6.68 ± 0.74	6.70 ± 1.02	4.97 ± 0.52
Tip-Adapter	5%	36.49 ± 2.63	2.22 ± 0.00	2.22 ± 0.00	16.32 ± 0.43	16.99 ± 0.38	14.85 ± 0.66
Tip-Adapter	10%	37.81 ± 1.37	56.27 ± 3.17	2.22 ± 0.00	16.92 ± 0.24	18.34 ± 0.54	26.31 ± 0.54
SimpleShot	1%	34.68 ± 1.33	45.42 ± 3.84	35.36 ± 4.27	54.47 ± 6.34	44.60 ± 4.05	42.91 ± 0.49
SimpleShot	5%	44.55 ± 2.74	54.89 ± 2.81	37.79 ± 0.90	66.01 ± 2.53	48.71 ± 1.80	50.39 ± 0.26
SimpleShot	10%	43.95 ± 3.53	58.55 ± 2.83	43.49 ± 2.88	68.17 ± 2.01	52.38 ± 1.48	53.31 ± 1.25
LaplacianShot	1%	32.50 ± 2.81	43.16 ± 4.91	30.67 ± 3.43	49.51 ± 6.99	37.99 ± 9.04	38.77 ± 2.84
LaplacianShot	5%	43.62 ± 3.44	55.16 ± 2.42	36.15 ± 5.13	60.20 ± 0.33	37.75 ± 8.50	46.57 ± 1.62
LaplacianShot	10%	41.26 ± 2.54	56.24 ± 3.20	36.15 ± 3.23	63.94 ± 5.98	34.55 ± 4.16	46.43 ± 2.14
SITR-Calib	1%	–	–	–	–	–	–
SITR-Calib	5%	–	–	–	–	–	–
SITR-Calib	10%	–	–	–	–	–	–
AnyTouch Match	1%	37.72 ± 2.87	43.10 ± 7.27	2.22 ± 0.00	54.40 ± 4.61	46.12 ± 1.92	36.71 ± 2.29
AnyTouch Match	5%	57.65 ± 2.90	50.32 ± 1.26	2.22 ± 0.00	58.69 ± 1.34	55.58 ± 2.65	44.89 ± 0.83
AnyTouch Match	10%	57.94 ± 2.41	55.03 ± 4.87	2.22 ± 0.00	63.61 ± 4.27	49.98 ± 4.48	45.76 ± 2.00
CTSRL CSM	1%	37.83 ± 6.49	42.01 ± 6.24	5.54 ± 1.11	$\mathbf{58.70 \pm 4.67}$	50.44 ± 2.78	38.91 ± 2.21
CTSRL CSM	5%	52.67 ± 5.77	48.53 ± 2.66	5.70 ± 3.34	65.23 ± 0.62	52.86 ± 1.00	45.00 ± 1.79
CTSRL CSM	10%	54.50 ± 3.45	49.62 ± 4.58	6.03 ± 1.10	65.63 ± 1.96	51.98 ± 2.43	45.55 ± 2.04
BIDETA	1%	$\mathbf{48.45 \pm 8.25}$	$\mathbf{88.95 \pm 5.31}$	$\mathbf{37.75 \pm 2.06}$	56.30 ± 3.04	$\mathbf{59.17 \pm 6.55}$	$\mathbf{58.12 \pm 2.49}$
BIDETA	5%	$\mathbf{81.81 \pm 7.34}$	$\mathbf{100.00 \pm 0.00}$	$\mathbf{50.84 \pm 1.50}$	$\mathbf{86.77 \pm 3.44}$	$\mathbf{80.14 \pm 3.75}$	$\mathbf{79.91 \pm 0.16}$
BIDETA	10%	$\mathbf{89.67 \pm 3.41}$	$\mathbf{100.00 \pm 0.00}$	$\mathbf{61.92 \pm 2.19}$	$\mathbf{86.88 \pm 1.88}$	$\mathbf{77.14 \pm 5.91}$	$\mathbf{83.12 \pm 1.19}$

Table 16: **Unified-shot classification on SITR with the frozen TVL ViT-Small backbone.** Accuracy and Macro-F1 (%) are averaged over the five target sensors excluded from hyperparameter development and reported as mean ± sample standard deviation over three random seeds. Best results are in bold and second-best results are underlined.

	2-shot		5-shot		10-shot	
Method	Acc.	Macro-F1	Acc.	Macro-F1	Acc.	Macro-F1
Frozen + source head	7.77±0.18	2.34±0.03	7.77±0.18	2.34±0.03	7.77±0.18	2.34±0.03
Tip-Adapter	8.20±0.70	2.80±0.79	11.53±1.18	5.52±1.19	14.97±2.07	9.47±2.12
SimpleShot	28.49±2.13	26.36±1.86	33.49±1.35	31.58±1.20	36.68±1.26	35.05±1.04
LaplacianShot	28.55±3.57	25.38±3.23	34.33±0.74	31.02±0.58	37.77±1.26	34.70±1.18
SITR-Calib	26.58±1.77	24.95±2.05	32.45±2.47	31.03±2.84	37.15±2.40	35.06±2.61
AnyTouch Match	24.92±1.53	23.22±1.89	29.35±1.64	28.38±1.69	33.30±0.58	31.61±0.37
CTSRL CSM	25.95±1.27	24.22±1.37	30.57±2.13	29.30±2.01	34.33±1.22	32.56±1.41
BIDETA	**36.40±1.46**	**35.00±1.38**	**54.90±2.91**	**53.96±3.10**	**65.83±1.84**	**64.85±1.67**

Table 17: **Unified-shot classification on SITR with the frozen Sparsh-DINO Small backbone.** Accuracy and Macro-F1 (%) are averaged over the five target sensors excluded from hyperparameter development and reported as mean ± sample standard deviation over three random seeds. Best results are in bold and second-best results are underlined.

	2-shot		5-shot		10-shot	
Method	Acc.	Macro-F1	Acc.	Macro-F1	Acc.	Macro-F1
Frozen + source head	6.98±0.40	1.26±0.14	6.98±0.40	1.26±0.14	6.98±0.40	1.26±0.14
Tip-Adapter	6.99±0.40	1.27±0.14	9.69±1.20	4.05±1.56	12.57±0.42	7.11±0.33
SimpleShot	24.39±0.62	23.11±0.96	30.87±1.54	29.83±1.78	34.45±0.98	33.41±0.98
LaplacianShot	25.17±0.71	23.39±1.13	32.26±1.46	30.58±1.64	34.94±1.51	33.44±1.47
SITR-Calib	21.79±1.66	19.41±1.38	29.20±0.41	26.21±0.89	29.08±1.85	26.42±1.57
AnyTouch Match	22.66±1.22	20.58±0.72	29.23±0.33	27.14±0.35	29.81±1.34	27.31±1.78
CTSRL CSM	23.17±0.67	21.34±1.29	29.67±0.35	27.29±0.67	30.55±1.22	28.43±1.22
BIDETA	**35.15±0.55**	**34.54±0.45**	**57.52±0.54**	**56.87±0.65**	**70.06±0.48**	**69.58±0.72**

Table 18: **Unified-shot classification on TacVerse Shape with the frozen TVL ViT-Small backbone.** Accuracy and Macro-F1 (%) are averaged over five target sensors and reported as mean ± sample standard deviation over three random seeds. SITR-Calib is unavailable because standard calibration images are absent. Best results are in bold and second-best results are underlined.

	2-shot		5-shot		10-shot	
Method	Acc.	Macro-F1	Acc.	Macro-F1	Acc.	Macro-F1
Frozen + source head	9.56±0.48	3.55±0.13	9.56±0.48	3.55±0.13	9.56±0.48	3.55±0.13
Tip-Adapter	15.47±3.11	9.20±3.18	28.64±1.87	22.20±1.89	36.35±1.15	29.93±1.36
SimpleShot	40.61±1.81	38.21±2.13	46.07±1.76	43.71±1.77	49.41±2.44	47.16±2.68
LaplacianShot	40.96±2.88	34.78±3.99	43.90±2.91	37.65±3.58	46.82±2.22	41.11±2.93
AnyTouch Match	36.66±1.89	35.07±1.47	43.76±2.32	41.19±2.38	50.07±1.64	47.09±1.58
CTSRL CSM	36.79±2.17	35.12±1.42	45.41±2.25	43.15±2.54	49.77±2.20	46.40±2.62
BIDETA	**47.90±1.08**	**46.41±1.24**	**67.04±2.80**	**66.42±2.77**	**72.99±0.92**	**72.23±1.15**

Table 19: **Unified-shot classification on TacVerse Shape with the frozen Sparsh-DINO Small backbone.** Accuracy and Macro-F1 (%) are averaged over five target sensors and reported as mean ± sample standard deviation over three random seeds. SITR-Calib is unavailable because standard calibration images are absent. Best results are in bold and second-best results are underlined.

	2-shot		5-shot		10-shot	
Method	Acc.	Macro-F1	Acc.	Macro-F1	Acc.	Macro-F1
Frozen + source head	13.67±0.47	4.64±0.19	13.67±0.47	4.64±0.19	13.67±0.47	4.64±0.19
Tip-Adapter	14.28±1.17	5.00±0.69	21.21±0.43	10.26±0.84	23.66±2.07	12.91±2.49
SimpleShot	41.33±1.43	38.70±3.28	51.21±2.44	48.23±4.01	53.20±0.89	50.15±1.41
LaplacianShot	39.83±3.21	34.89±5.21	50.50±4.01	45.46±5.91	50.85±2.21	44.86±3.66
AnyTouch Match	37.01±3.71	32.85±4.12	43.79±1.69	39.04±2.53	48.01±0.37	43.82±0.80
CTSRL CSM	39.30±3.42	35.06±4.05	45.44±1.40	41.27±2.35	49.27±0.84	44.49±1.51
BIDETA	**48.55±4.73**	**47.52±5.72**	**70.73±1.62**	**69.46±1.56**	**77.59±1.67**	**76.78±1.86**

B.4 Class-wise prediction structure

Figure 4 compares BIDETA with SimpleShot on the same TacVerse queries using 10% support. BIDETA provides stronger class discrimination and cross-class balance, with predictions concentrated more clearly along the diagonal than SimpleShot, producing more accurate and stable recognition.

Table 20: **TacQuad identity-ranking results with TVL.** MRR and R@1 (%) are reported as mean ± sample standard deviation over three random seeds. Best results are in bold and second-best results are underlined.

Method	R@1 10%	MRR 10%	R@1 20%	MRR 20%	R@1 30%	MRR 30%
Frozen backbone	2.46 ± 0.00	8.76 ± 0.00	2.46 ± 0.00	8.76 ± 0.00	2.46 ± 0.00	8.76 ± 0.00
Tip-Adapter	24.18 ± 3.93	36.68 ± 3.70	34.90 ± 11.12	47.80 ± 12.36	41.29 ± 11.62	53.80 ± 9.55
SimpleShot	48.07 ± 1.57	59.05 ± 0.78	53.42 ± 2.91	64.65 ± 1.65	54.54 ± 1.95	66.57 ± 1.06
LaplacianShot	24.93 ± 5.86	45.11 ± 3.03	27.08 ± 2.95	48.78 ± 2.27	23.96 ± 1.90	47.64 ± 1.44
SITR-Support	45.39 ± 2.91	57.94 ± 2.00	52.16 ± 0.52	64.54 ± 0.38	52.38 ± 1.52	65.49 ± 1.42
AnyTouch Match	45.76 ± 0.97	58.29 ± 1.37	54.09 ± 4.15	66.15 ± 2.49	56.32 ± 1.49	68.49 ± 1.10
CTSRL CSM	45.39 ± 3.19	57.49 ± 2.16	51.19 ± 1.79	63.88 ± 0.88	52.16 ± 2.34	65.38 ± 1.35
BIDETA	**50.97 ± 1.01**	**61.16 ± 0.95**	**69.35 ± 0.93**	**76.98 ± 0.20**	**74.85 ± 2.28**	**80.90 ± 1.89**

C Additional identity-ranking results

Tables 20 and 21 present identity-ranking results on TacQuad with different pretrained backbones. R@1 is the percentage of queries for which the correct identity ranks first and is equivalent to closed-set identity classification accuracy. Compared with the other methods, BIDETA consistently achieves higher first-rank accuracy and MRR across both backbones and all support budgets, demonstrating stronger cross-sensor identity discrimination.

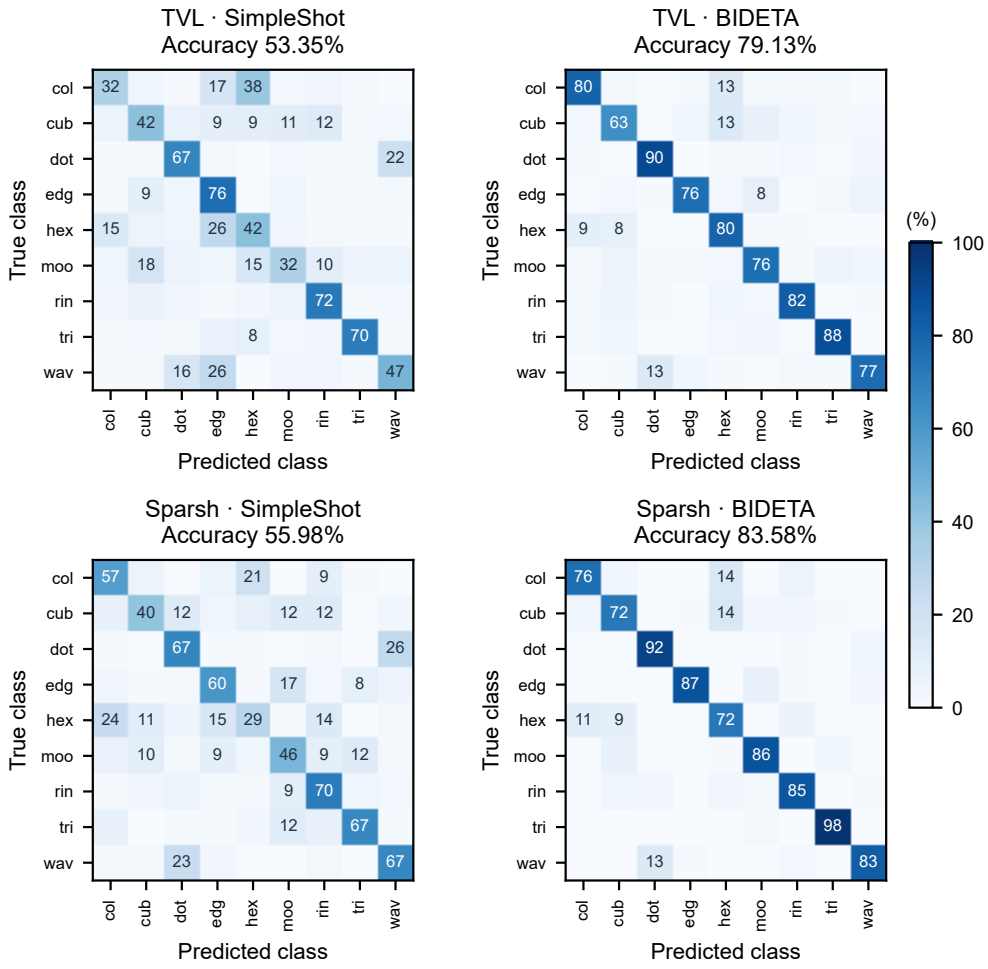


Figure 4: **TacVerse class predictions with 10% support.** Row-normalized confusion matrices compare SimpleShot and BIDETA on identical queries. Each matrix aggregates 13,500 predictions from five target sensors, three random seeds, and 900 queries per target.

Table 21: **TacQuad identity-ranking results with Sparsh.** MRR and R@1 (%) are reported as mean ± sample standard deviation over three random seeds. Best results are in bold and second-best results are underlined.

Method	R@1 10%	MRR 10%	R@1 20%	MRR 20%	R@1 30%	MRR 30%
Frozen backbone	1.34 ± 0.00	10.01 ± 0.00	1.34 ± 0.00	10.01 ± 0.00	1.34 ± 0.00	10.01 ± 0.00
Tip-Adapter	36.90 ± 0.34	45.80 ± 0.69	51.04 ± 1.23	61.42 ± 1.17	54.39 ± 1.23	64.57 ± 0.73
SimpleShot	51.26 ± 0.93	60.36 ± 0.85	56.55 ± 1.61	65.83 ± 1.20	60.27 ± 0.80	68.75 ± 0.78
LaplacianShot	21.65 ± 3.72	43.32 ± 3.10	18.15 ± 3.16	43.61 ± 2.61	20.16 ± 0.85	45.82 ± 0.88
SITR-Support	51.79 ± 1.83	61.49 ± 1.62	57.66 ± 0.78	67.90 ± 0.63	60.49 ± 0.59	70.72 ± 0.44
AnyTouch Match	51.86 ± 0.68	61.73 ± 0.93	57.07 ± 1.92	67.93 ± 0.85	60.64 ± 1.36	71.00 ± 0.51
CTSRL CSM	48.29 ± 1.01	58.88 ± 0.78	53.79 ± 1.36	64.99 ± 1.11	57.37 ± 0.67	68.14 ± 0.23
BIDETA	$\mathbf{62.05 \pm 1.36}$	$\mathbf{69.28 \pm 0.72}$	$\mathbf{74.85 \pm 2.83}$	$\mathbf{80.82 \pm 2.67}$	$\mathbf{80.73 \pm 2.19}$	$\mathbf{85.86 \pm 1.80}$

D COMPLETE COMPONENT EVIDENCE

Table 22: **Complete component ablation on SITR.** Macro-F1 (%) is reported as mean ± sample standard deviation over three random seeds. Best results are in bold and second-best results are underlined.

Type	Variant	TVL 1%	TVL 5%	TVL 10%	Sparsh 1%	Sparsh 5%	Sparsh 10%
Memory	View 1 only	49.61 ± 0.97	71.88 ± 0.94	75.25 ± 0.28	50.18 ± 3.23	74.41 ± 0.14	76.89 ± 0.63
	View 2 only	48.62 ± 1.79	70.26 ± 0.92	75.45 ± 0.13	53.06 ± 1.63	70.68 ± 0.19	76.91 ± 0.86
	Dual-view memory	50.84 ± 0.98	72.13 ± 0.95	75.48 ± 0.21	51.38 ± 3.11	73.84 ± 0.13	77.16 ± 0.83
Geometry	+ Consensus recurrence	55.69 ± 1.06	76.76 ± 0.61	79.72 ± 0.22	54.36 ± 2.65	78.18 ± 0.72	80.43 ± 0.73
	+ Support standardization	56.35 ± 1.09	77.37 ± 0.62	80.14 ± 0.23	55.14 ± 2.56	78.78 ± 0.72	81.12 ± 0.65
	+ Spectral normalization	58.13 ± 1.26	80.98 ± 0.78	82.40 ± 0.08	58.16 ± 2.50	81.33 ± 0.95	83.84 ± 0.83
Fusion and gate	Arithmetic graph fusion	60.96 ± 1.39	82.55 ± 0.76	85.43 ± 0.23	60.88 ± 1.82	83.19 ± 0.76	86.65 ± 0.64
	Without spectral normalization	57.02 ± 1.14	75.43 ± 0.52	81.10 ± 0.13	55.27 ± 2.20	77.16 ± 0.42	81.98 ± 0.73
	Disagreement only	$\mathbf{61.45 \pm 1.36}$	82.88 ± 0.79	$\mathbf{85.45 \pm 0.19}$	$\mathbf{61.16 \pm 1.79}$	$\underline{83.38 \pm 0.79}$	$\underline{86.86 \pm 0.61}$
	Entropy only	$\underline{61.31 \pm 1.32}$	$\mathbf{82.94 \pm 0.78}$	$\underline{85.44 \pm 0.32}$	$\underline{60.85 \pm 1.84}$	$\mathbf{83.44 \pm 0.83}$	86.80 ± 0.48
	Full BIDETA	$\mathbf{61.45 \pm 1.36}$	$\mathbf{82.94 \pm 0.79}$	$\mathbf{85.45 \pm 0.19}$	$\mathbf{61.16 \pm 1.79}$	$\underline{83.41 \pm 0.79}$	$\mathbf{86.87 \pm 0.58}$

Table 22 reports Macro-F1 for 11 BIDETA variants on SITR. Spectral normalization reshapes the standardized feature geometry and consistently strengthens class-balanced discrimination, while the uncertainty gate regulates anchor and graph evidence to reduce unreliable propagation. Removing spectral normalization produces the largest decline among the full-model variants; separating the gate branches shows that disagreement matches the complete gate in several settings, whereas entropy contributes more strongly with Sparsh at the 5% budget. Overall, the complete framework provides the strongest and most consistent performance pattern, indicating that memory construction, spectral geometry, and reliability-aware inference form a tightly coordinated adaptation process.

E PARAMETER SENSITIVITY

E.1 TRAINING-ONLY INNER VALIDATION ON SITR

Tables 23–29 evaluate parameter sensitivity on Mini_3 with a 5% support budget. Selecting parameters on one training-only sensor separates configuration selection from the six-sensor evaluation and improves experimental fairness. The disagreement weight and iteration count remain comparatively stable around their selected values, whereas graph temperature and recurrence range are more sensitive because they directly control edge concentration and the strength of recurrent evidence propagation.

Table 23: **Neighborhood-size sensitivity on SITR.** Accuracy (%) on the Mini_3 inner validation split with 5% support is reported as mean ± sample standard deviation over three random seeds. * indicates the selected value.

Value	TVL	Sparsh
10	89.74 ± 2.69	87.25 ± 3.67
20	91.17 ± 2.11	88.01 ± 4.92
40	$92.38 \pm 1.91^{*}$	$88.39 \pm 4.82^{*}$
80	92.06 ± 2.28	87.28 ± 5.16
120	91.45 ± 2.30	85.49 ± 4.26
160	90.57 ± 2.37	84.71 ± 4.22

Table 24: **Graph-temperature sensitivity on SITR.** Accuracy (%) on the Mini_3 inner validation split with 5% support is reported as mean ± sample standard deviation over three random seeds. * indicates the selected value.

Value	TVL	Sparsh
0.03	84.92 ± 2.29	86.03 ± 4.05
0.05	86.52 ± 2.72	86.55 ± 4.24
0.07	88.16 ± 2.59	86.68 ± 4.88
0.1	90.00 ± 2.48	87.38 ± 4.91
0.2	$92.38 \pm 1.91^*$	$88.39 \pm 4.82^*$

Table 25: **Spectral-exponent sensitivity on SITR.** Accuracy (%) on the Mini_3 inner validation split with 5% support is reported as mean ± sample standard deviation over three random seeds. * indicates the selected value.

Value	TVL	Sparsh
0.0	88.61 ± 2.23	83.66 ± 5.34
0.05	89.40 ± 2.07	84.43 ± 5.36
0.15	90.39 ± 1.81	85.59 ± 5.41
0.25	91.25 ± 1.46	86.63 ± 5.68
0.4	92.03 ± 1.76	87.77 ± 5.21
0.6	$92.38 \pm 1.91^*$	$88.39 \pm 4.82^*$

Table 26: **Disagreement-weight sensitivity on SITR.** Accuracy (%) on the Mini_3 inner validation split with 5% support is reported as mean ± sample standard deviation over three random seeds. * indicates the selected value.

Value	TVL	Sparsh
0.0	92.38 ± 1.94	88.37 ± 4.83
0.25	92.38 ± 1.91	88.36 ± 4.84
0.5	$92.38 \pm 1.91^*$	$88.39 \pm 4.82^*$
0.75	92.38 ± 1.91	88.35 ± 4.86
1.0	92.37 ± 1.89	88.31 ± 4.90

Table 27: **Gate-exponent sensitivity on SITR.** Accuracy (%) on the Mini_3 inner validation split with 5% support is reported as mean ± sample standard deviation over three random seeds. * indicates the selected value.

Value	TVL	Sparsh
0.05	$92.38 \pm 1.91^*$	$88.39 \pm 4.82^*$
0.15	92.06 ± 2.05	88.18 ± 4.73
0.3	91.48 ± 2.20	87.97 ± 4.64
0.5	91.09 ± 2.37	87.96 ± 4.48
1.0	90.47 ± 2.48	87.51 ± 4.25

Table 28: **Recurrence-range sensitivity on SITR.** Accuracy (%) on the Mini_3 inner validation split with 5% support is reported as mean ± sample standard deviation over three random seeds. * indicates the selected value.

Value	TVL	Sparsh
0.20:0.40	85.17 ± 2.07	83.61 ± 3.52
0.40:0.60	88.05 ± 2.44	85.51 ± 3.88
0.50:0.70	89.51 ± 2.51	86.68 ± 4.32
0.60:0.80	90.98 ± 2.48	87.81 ± 4.70
0.70:0.90	$92.38 \pm 1.91^*$	$88.39 \pm 4.82^*$

Table 29: **Iteration-count sensitivity on SITR.** Accuracy (%) on the Mini_3 inner validation split with 5% support is reported as mean ± sample standard deviation over three random seeds. * indicates the selected value.

Value	TVL	Sparsh
1	88.24 ± 2.88	84.96 ± 3.63
3	91.80 ± 2.13	87.67 ± 4.74
5	92.38 ± 1.91	88.15 ± 4.79
10	$92.38 \pm 2.08^{*}$	$88.39 \pm 4.82^{*}$
20	92.21 ± 2.33	88.18 ± 4.97
30	92.08 ± 2.28	87.57 ± 4.02
50	91.91 ± 2.20	86.45 ± 4.39

E.2 TacQuad ranking sensitivity

Tables 30–34 report parameter sensitivity for identity ranking on TacQuad. At the default spectral exponent of 0.25, spectral normalization improves MRR over an exponent of zero across all six backbone–budget settings. Neighborhood size and disagreement weight remain comparatively stable, whereas the spectral exponent and recurrence length have larger effects because they reshape inter-identity neighborhoods and regulate how strongly ranking evidence is propagated; this influence is most visible under the lowest support budget.

Table 30: **Neighborhood-size sensitivity on TacQuad.** MRR (%) for identity ranking is reported as mean ± sample standard deviation over three random seeds. * indicates the default value.

Value	TVL-10%	TVL-20%	TVL-30%	Sparsh-10%	Sparsh-20%	Sparsh-30%
10	61.38 ± 1.04	77.14 ± 0.30	80.88 ± 1.76	69.61 ± 0.60	80.82 ± 2.81	86.04 ± 1.45
20	61.23 ± 0.91	77.09 ± 0.36	81.10 ± 1.84	69.29 ± 0.60	80.86 ± 2.71	85.99 ± 1.66
40	61.08 ± 0.86	76.97 ± 0.19	80.96 ± 1.85	69.26 ± 0.66	80.54 ± 2.68	85.96 ± 1.71
80 *	61.16 ± 0.95	76.98 ± 0.20	80.90 ± 1.89	69.28 ± 0.72	80.82 ± 2.67	85.86 ± 1.80
120	61.17 ± 0.95	76.96 ± 0.24	80.85 ± 1.93	69.34 ± 0.73	80.83 ± 2.63	85.86 ± 1.81

Table 31: **Spectral-exponent sensitivity on TacQuad.** MRR (%) for identity ranking is reported as mean ± sample standard deviation over three random seeds. * indicates the default value.

Value	TVL-10%	TVL-20%	TVL-30%	Sparsh-10%	Sparsh-20%	Sparsh-30%
0.0	60.14 ± 0.31	75.63 ± 0.74	79.47 ± 1.46	68.22 ± 0.53	78.53 ± 3.38	84.00 ± 2.17
0.1	60.64 ± 0.50	76.31 ± 0.34	80.22 ± 1.67	68.59 ± 0.86	79.44 ± 3.20	84.65 ± 1.79
0.25 *	61.16 ± 0.95	76.98 ± 0.20	80.90 ± 1.89	69.28 ± 0.72	80.82 ± 2.67	85.86 ± 1.80
0.4	61.34 ± 1.15	77.26 ± 0.10	81.23 ± 1.61	69.47 ± 0.66	81.14 ± 2.34	86.22 ± 1.49
0.6	61.61 ± 1.10	77.35 ± 0.13	81.52 ± 1.72	69.43 ± 0.62	81.21 ± 2.31	86.10 ± 1.41

Table 32: **Recurrence-range sensitivity on TacQuad.** MRR (%) for identity ranking is reported as mean ± sample standard deviation over three random seeds. * indicates the default value.

Value	TVL-10%	TVL-20%	TVL-30%	Sparsh-10%	Sparsh-20%	Sparsh-30%
0.20:0.40	60.88 ± 0.41	76.51 ± 0.49	80.55 ± 1.09	68.33 ± 0.72	80.07 ± 2.70	85.36 ± 1.40
0.40:0.60	61.20 ± 0.82	76.83 ± 0.35	80.75 ± 1.31	68.86 ± 0.92	80.37 ± 2.57	85.57 ± 1.72
0.60:0.80 *	61.16 ± 0.95	76.98 ± 0.20	80.90 ± 1.89	69.28 ± 0.72	80.82 ± 2.67	85.86 ± 1.80
0.70:0.90	61.06 ± 1.01	76.98 ± 0.23	80.62 ± 1.93	69.71 ± 0.57	80.95 ± 2.81	86.15 ± 1.56

Table 33: **Disagreement-weight sensitivity on TacQuad.** MRR (%) for identity ranking is reported as mean ± sample standard deviation over three random seeds. * indicates the default value.

Value	TVL-10%	TVL-20%	TVL-30%	Sparsh-10%	Sparsh-20%	Sparsh-30%
0.0	61.24 ± 0.92	76.96 ± 0.34	80.78 ± 1.93	69.16 ± 0.75	80.82 ± 2.68	85.85 ± 1.77
0.25	61.21 ± 0.93	77.03 ± 0.29	80.85 ± 1.91	69.21 ± 0.78	80.84 ± 2.71	85.88 ± 1.79
0.5 *	61.16 ± 0.95	76.98 ± 0.20	80.90 ± 1.89	69.28 ± 0.72	80.82 ± 2.67	85.86 ± 1.80
0.75	61.16 ± 0.96	76.97 ± 0.17	80.93 ± 1.88	69.30 ± 0.70	80.76 ± 2.68	85.91 ± 1.75
1.0	61.10 ± 0.92	77.04 ± 0.19	80.94 ± 1.88	69.39 ± 0.66	80.77 ± 2.69	85.91 ± 1.75

Table 34: **Iteration-count sensitivity on TacQuad.** MRR (%) for identity ranking is reported as mean ± sample standard deviation over three random seeds. * indicates the default value.

Value	TVL-10%	TVL-20%	TVL-30%	Sparsh-10%	Sparsh-20%	Sparsh-30%
1	63.35 ± 1.77	78.01 ± 0.60	81.93 ± 2.32	69.49 ± 0.82	81.20 ± 2.54	85.82 ± 1.19
3	62.42 ± 1.41	77.87 ± 0.84	81.40 ± 2.20	69.41 ± 1.07	81.37 ± 2.83	86.02 ± 1.28
5	62.00 ± 1.09	77.49 ± 0.70	80.97 ± 1.86	69.42 ± 0.81	81.22 ± 2.62	85.76 ± 1.32
10	61.18 ± 0.89	77.10 ± 0.27	80.92 ± 1.82	69.29 ± 0.78	80.88 ± 2.77	85.95 ± 1.76
20 *	61.16 ± 0.95	76.98 ± 0.20	80.90 ± 1.89	69.28 ± 0.72	80.82 ± 2.67	85.86 ± 1.80
30	61.14 ± 0.90	76.88 ± 0.12	80.79 ± 1.94	69.27 ± 0.71	80.77 ± 2.66	85.86 ± 1.83

F Algorithm Cost Comparison

Table 35: **Target-sensor adaptation and inference costs on three tactile datasets.** Adaptation time (s) is reported as mean ± sample standard deviation over three random seeds. Query/s denotes the estimated end-to-end throughput including frozen-encoder inference. SITR and TacVerse use 10% support, while TacQuad uses 30% support. SITR-Calib is unavailable on TacVerse; its TacQuad counterpart is denoted SITR-Support. Best results are in bold and second-best results are underlined.

			SITR		TacVerse Shape		TacQuad	
Backbone	Method	Mode	Adapt. (s) ↓	Query/s ↑	Adapt. (s) ↓	Query/s ↑	Adapt. (s) ↓	Query/s ↑
TVL	Frozen backbone	I	$\mathbf{0.000 \pm 0.000}$	**7,168.1**	$\mathbf{0.000 \pm 0.000}$	**7,156.6**	$\mathbf{0.000 \pm 0.000}$	**7,082.3**
TVL	Tip-Adapter	I	0.033 ± 0.009	6,770.7	0.007 ± 0.003	6,932.0	0.007 ± 0.001	6,793.0
TVL	SimpleShot	I	0.009 ± 0.000	7,026.5	0.003 ± 0.000	7,016.2	$\underline{0.004 \pm 0.000}$	6,970.9
TVL	LaplacianShot	T	$\underline{0.008 \pm 0.000}$	5,256.2	$\underline{0.002 \pm 0.000}$	5,688.7	$\underline{0.004 \pm 0.000}$	4,697.8
TVL	SITR-Calib / SITR-Support	I	12.268 ± 0.807	7,118.3	–	–	25.269 ± 2.325	6,903.6
TVL	AnyTouch Match	I	10.249 ± 0.394	$\underline{7,130.8}$	8.183 ± 1.489	$\underline{7,112.1}$	22.330 ± 3.052	$\underline{6,989.5}$
TVL	CTSRL CSM	I	13.232 ± 0.223	7,115.8	10.623 ± 0.105	7,092.2	27.206 ± 0.590	6,880.3
TVL	BIDETA	T	0.636 ± 0.015	6,056.0	0.514 ± 0.031	5,835.1	0.846 ± 0.051	4,581.3
Sparsh	Frozen backbone	I	$\mathbf{0.000 \pm 0.000}$	**4,965.2**	$\mathbf{0.000 \pm 0.000}$	**4,958.5**	$\mathbf{0.000 \pm 0.000}$	**4,930.4**
Sparsh	Tip-Adapter	I	0.034 ± 0.011	4,845.3	0.005 ± 0.000	4,884.9	0.007 ± 0.001	4,830.3
Sparsh	SimpleShot	I	$\underline{0.009 \pm 0.001}$	4,891.2	$\underline{0.002 \pm 0.000}$	4,888.8	$\underline{0.004 \pm 0.000}$	4,866.6
Sparsh	LaplacianShot	T	0.009 ± 0.000	3,980.5	0.002 ± 0.000	4,167.6	0.004 ± 0.000	3,593.8
Sparsh	SITR-Calib / SITR-Support	I	12.600 ± 1.182	4,943.1	–	–	24.219 ± 0.175	4,840.1
Sparsh	AnyTouch Match	I	10.316 ± 0.110	$\underline{4,951.2}$	7.655 ± 0.882	$\underline{4,942.2}$	23.770 ± 0.570	$\underline{4,881.0}$
Sparsh	CTSRL CSM	I	13.688 ± 0.155	4,944.1	10.136 ± 0.659	4,929.0	24.917 ± 1.456	4,838.2
Sparsh	BIDETA	T	0.283 ± 0.034	4,545.1	0.160 ± 0.001	4,477.0	0.293 ± 0.015	3,747.4