TACO: A BENCHMARK FOR LOSSLESS AND LOSSY CODECS OF HETEROGENEOUS TACTILE DATA

Anonymous authors
Paper under double-blind review

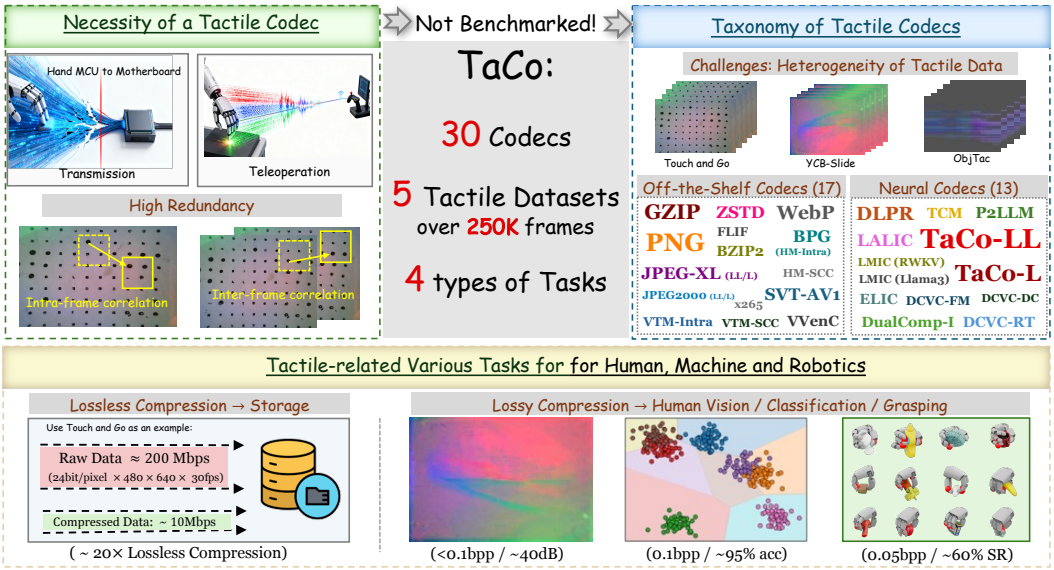


Figure 1: The motivation of our **TaCo** benchmark, established through an extensive evaluation on tactile codecs across multiple dimensions. First, we assess 30 off-the-shelf and neural codecs on 5 heterogeneous tactile datasets with more than 250K frames. Second, we introduce purely-trained TaCo-LL and TaCo-L codecs to explore the data-driven approaches in the field of lossless and lossy tactile data compression. Finally, we evaluate the coding performance on 4 distinct task types designed to serve for human, machine, and robotics.

ABSTRACT

Tactile sensing is crucial for embodied intelligence, providing fine-grained perception and control in complex environments. However, efficient tactile data compression, which is essential for real-time robotic applications under strict bandwidth constraints, remains underexplored. The inherent heterogeneity and spatiotemporal complexity of tactile data further complicate this challenge. To bridge this gap, we introduce **TaCo**, the first comprehensive benchmark for **Ta**ctile data **Co**decs. TaCo evaluates 30 compression methods, including off-the-shelf compression algorithms and neural codecs, across five diverse datasets from various sensor types. We systematically assess both lossless and lossy compression schemes on four key tasks: lossless storage, human visualization, material and object classification, and dexterous robotic grasping. Notably, we pioneer the development of data-driven codecs explicitly trained on tactile data, **TaCo-LL** (lossless) and **TaCo-L** (lossy). Results have validated the superior performance of our TaCo-LL and TaCo-L. This benchmark provides a foundational framework for understanding the critical trade-offs between compression efficiency and task performance, paving the way for future advances in tactile perception.

1 INTRODUCTION

The acquisition and interpretation of tactile data are paramount for advancing embodied AI and achieving sophisticated human-machine interaction, as they provide the rich, physical context neces-

sary for dexterous manipulation and awareness of the environment. However, the high-dimensional, spatio-temporally dense nature of tactile sensing results in rapidly growing data volumes, posing a significant bottleneck for real-time applications. Consequently, efficient tactile data compression is critical for real-time haptic feedback in dexterous hands, remote teleoperation, and large-scale storage of physical interactions for robotic model training.

While the imperative for efficient tactile data compression is well-established, current approaches remain diverse and fragmented. A corpus of existing research has explored this challenge through classical signal processing techniques (like dimensionality reduction and wavelet transforms), leveraging transforms and codecs designed for speech or image data. In recent years, data-driven methods have gradually gained popularity for their ability to learn optimal compact representations. Specifically, neural networks can learn compact latent representations in a data-driven manner, enabling efficient lossless or lossy compression (Liu et al., 2023b; Mentzer et al., 2019). Compared to traditional codecs, neural compression offers greater flexibility and adaptability, particularly in scenarios involving complex or irregular signal structures (Ma et al., 2019). These methods have been successfully applied in domains such as video and image compression (Zhao et al., 2025; Feng et al., 2025a), but they are still unexplored for tactile data. Another difficulty in generating a data-driven codec for tactile datasets is the heterogeneity, arising from different sensing principles: visuo-tactile sensors such as Gelsight (Yuan et al., 2017a) and DIGIT (Lambeta et al., 2020) capture surface transformation, while other force-based sensors (Paxini, 2025) measure force data. Therefore, the establishment of a comprehensive and open benchmark, comprising representative datasets, standardized evaluation metrics, and baseline models, is not merely beneficial but a necessary prerequisite for catalyzing advancements in this critical domain and enabling new research.

As illustrated in Fig. 1, we construct a comprehensive benchmark to evaluate the compression performance of various codecs on heterogeneous tactile datasets. First, we collect five diverse tactile datasets, and 30 representative codecs. They include off-the-shelf codecs designed for text, image and video, aiming to remove the 1D and 2D, inter-frame and intra-frame redundancies. We also incorporate neural codecs pretrained on other modalities to assess their cross-domain generalization on tactile data. Second, we propose two data-driven codecs, i.e. TaCo-LL and TaCo-L, which are trained from tactile datasets to learn intrinsic data patterns and exploit the redundancies in heterogeneous tactile data. Third, we evaluate tactile compression performance using four types of tasks: lossless compression, lossy compression for human perception, semantic classification, and robot grasping. Experimental results validate the superior performance of our proposed data-driven models, TaCo-LL and TaCo-L, and we hope our benchmark will inspire further research in this field.

In summary, our main contributions are as follows.

- We propose **TaCo**, the first comprehensive benchmark for tactile data codecs. It comprises five publicly tactile datasets, 30 codecs, and four tactile-related tasks: lossless compression, lossy compression for human visualization, tactile classification, and robotic grasping.
- We introduce **TaCo-LL** and **TaCo-L**, the first purely data-driven tactile codecs, trained end-to-end on heterogeneous tactile datasets to learn the intrinsic data distributions.
- Extensive experimental results demonstrate that our proposed TaCo-LL and TaCo-L models surpass existing methods across all four tasks, establishing a foundation for future research in the field of tactile data compression.

2 Related Work

Tactile Datasets. Tactile datasets play a key role in advancing robotic perception and manipulation, supporting tasks like grasping, object recognition, and material classification. Several recent datasets focus only on tactile signals (Liu & Ward-Cherrier, 2024; Zhao et al., 2024; Suresh et al., 2023; Yuan et al., 2018b; Higuera et al., 2024; Schneider et al., 2025). TIP Bench (Liu & Ward-Cherrier, 2024) converts sensor outputs into heatmaps and evaluates spatial acuity, stability, and generalization. FoTa (Zhao et al., 2024) merges multiple open datasets into a unified collection of over three million samples. YCB-Slide (Suresh et al., 2023) records both simulated and real sliding interactions between a DIGIT tactile sensor and YCB objects. TacBench (Higuera et al., 2024) comprises 180,000+ unlabeled tactile images from surface-sliding interactions, facilitating large-scale self-supervised learning. Tactile MNIST (Schneider et al., 2025) provides both simulated and

real GelSight interactions for MNIST digits, including 13,580 3D-printable meshes and 153,600 tactile recordings. ActiveCloth (Yuan et al., 2018b) comprises 6,616 robotic squeeze trials on 153 garments, recording synchronized GelSight tactile videos and Kinect depth with 11 attributes.

Beyond pure tactile sensing, some other recent works (Yang et al., 2022; Feng et al., 2025b; Liu et al.; Cheng et al., 2025a; Yu et al., 2024; Suresh et al., 2024; Kerr et al., 2022; Yuan et al., 2017b; Li et al., 2019; Gao et al., 2021; Cheng et al., 2025b) incorporate multi-modal signals, combining tactile with vision, language, or audio to support cross-modal learning. For instance, Touch and Go (Yang et al., 2022) is the first large-scale tactile dataset collected in outdoor environments, capturing human interactions with natural objects via synchronized tactile and video data. VTDexManip (Liu et al.) provides 565,000 frames of video–tactile data from human multi-finger manipulations across 10 tasks and 182 objects, filling a gap in dexterous interaction datasets. Touch100K (Cheng et al., 2025a) compiles and cleans TAG and VisGel data into 100,147 high-quality triplets, offering the first large-scale alignment across tactile, visual, and linguistic modalities. TacQuad (Feng et al., 2025b) integrates four visual–tactile sensors, recording aligned tactile signals, RGB frames, and GPT-generated textual descriptions for multimodal reasoning. ObjectForlder (Gao et al., 2021) provides 100 neural object representations that encode 3D shape, appearance, sound, and tactile properties, supporting on-demand multimodal data generation for unified perception and control.

Tactile Compression. While tactile sensing continues to advance rapidly in resolution, sampling rate, and coverage, compression research for tactile data remains under-explored. Existing work has proposed some sparse or task-specific compression approaches (Hollis et al., 2016; Bartolozzi et al., 2017; Hollis et al., 2018; Shao et al., 2020; Hassen et al., 2020; Seeling et al., 2021; Liu et al., 2023a; Slepyan et al., 2024; Li et al., 2025; Lu et al., 2025). For instance, Shao et al. (2020) exploits the propagation of mechanical waves during dynamic contact to enable compact tactile encoding. (Hassen et al., 2020) proposes a perceptual vibrotactile codec that combines sparse linear prediction with an acceleration sensitivity function. (Seeling et al., 2021) achieves real-time tactile compression by combining bit-level truncation with delta-coding driven by just-noticeable-difference thresholds. Others like (Liu et al., 2023a) and (Slepyan et al., 2024) investigate dimensionality reduction via stacked auto-encoders or wavelet sparsification. However, these methods typically focus on simple signal sparsity or quantization strategies, often lack rigorous compression metrics and are tailored to relatively narrow scenarios or limited generalizability. In fact, many common tactile signals can be naturally transformed into image-like formats, enabling the use of standard image or general-purpose compressors. This direction is appealing not only because these compressors are well-established and widely available, but also because they offer tunable configurations to trade off compression ratio and distortion, making them adaptable to diverse robotic applications. Yet, this perspective remains largely under-explored in the tactile domain. To fill this gap, this paper presents a comprehensive benchmark for tactile compression methods, aiming to provide practical guidance and spark future research into efficient tactile data compression.

3 Tactile Datasets and Compression Methods

3.1 Tactile Datasets

We benchmark tactile compression across five representative datasets: Touch and Go (Yang et al., 2022), ObjectFolder 1.0 (Gao et al., 2021), SSVTP (Kerr et al., 2022), YCB-Slide (Suresh et al., 2023), and ObjTac (Cheng et al., 2025b). These datasets span a range of sensor types (vision-based and force-based), resolutions (from 120×160 to 640×480), and data scales, as detailed in Table. 8. Depending on the sensor type, tactile data exhibit strong structural heterogeneity, along with complex spatiotemporal correlations and redundancy, as illustrated in Fig. 1.

Specifically, the GelSight-based datasets (Touch and Go and ObjectFolder) and DIGIT-based datasets (SSVTP and YCB-Slide) are collected using vision-based tactile sensors that operate by illuminating a deformable elastomer surface with micro-LED arrays and capturing its surface deformation through an internal camera. This process converts tactile interactions into sequences of RGB images or videos, enabling direct applications of image or video compression techniques. In contrast, the ObjTac dataset is collected using force-based tactile sensors. The sensor comprises $N = 60$ contact points across the contact surface, each measuring a 3D force vector. These measurements form a temporally structured sequence of force data. To enable efficient compression, we

Table 1: Introduction of the utilized tactile datasets.

Dataset	#Objects	#Frames	Resolution	Sensor
Touch and Go (Yang et al., 2022)	3971	13.9K	$640 \times 480 \times 30$Hz	GelSight (Yuan et al., 2017a)
ObjectFolder 1.0 (Gao et al., 2021)	100	100K	$120 \times 160 \times 30$Hz	GelSight (Yuan et al., 2017a)
SSVTP (Kerr et al., 2022)	10	4.5K	$240 \times 320 \times 30$Hz	DIGIT (Lambeta et al., 2020)
YCB-Slide Suresh et al. (2023)	10	4.5K	$240 \times 320 \times 30$Hz	DIGIT (Lambeta et al., 2020)
ObjTac (Cheng et al., 2025b)	56	135K	$5 \times 12 \times 200$Hz	Force Sensor (Paxini, 2025)

map each 3D force vector to an RGB pixel and temporally stack the force readings across a time duration T, generating images of resolution $T \times 60$.

3.2 Tactile Compression Methods

We establish a benchmark for two categories of tactile codecs: 1) *off-the-shelf codecs* based on conventional signal processing, originally designed for general-purpose or visual data, and 2) *neural codecs* that leverage neural networks to learn data patterns end-to-end. As tactile signals are natively or transformable into image or video formats (see Section 3.1), our evaluation of neural codecs includes both pretrained image codecs and, to our knowledge, the first data-driven codecs explicitly trained on tactile datasets.

3.2.1 Off-the-Shelf Compression Methods

Typically, off-the-shelf compression methods have been historically designed for text, image and video data. These classical techniques are fundamentally rooted in signal processing principles, aiming to eliminate statistical, spatial, or temporal redundancies present in 1D or 2D data.

General-Purpose Compression Methods. We evaluate three general-purpose lossless compressors: gzip (Pasco., 1996), zstd (Meta., 2015), and bzip2 (Seward, 2000), which are designed to exploit 1D symbol redundancy using techniques such as dictionary coding (e.g., LZ77 in gzip and zstd), block-sorting transforms (e.g., Burrows-Wheeler in bzip2), and entropy coding.

Image and Video Compression Methods. When treating tactile data as images, we evaluate six standard image lossless codecs: PNG (Boutell, 1997), FLIF (Sneyers, 2015), WebP (Google, 2010), JPEG-XL (Team, 2021), JPEG2000 (ISO/IEC, 2000), and BPG (Bellard, 2014) (the intra-mode of HEVC/265 codec). These image-specific compressors remove 2D spatial redundancy in images via predictive coding, transform coding (e.g., DCT or wavelets), and context-based entropy coding.

In addition, we also evaluate six lossy image codecs: JPEG-XL (Team, 2021), JPEG2000 (ISO/IEC, 2000), as well as the intra-frame and screen content coding (SCC) modes of HM Sullivan et al. (2012) and VTM (Bross et al., 2021) (i.e., HM-Intra, HM-SCC, VTM-Intra, VTM-SCC).

To further address the inter-frame redundancy in tactile data, we evaluate three off-the-shell video codecs, VVenC (Bross et al., 2021), x265 (Sullivan et al., 2012), and SVT-AV1 (Han et al., 2021). Due to the huge amount of video data, lossless video compression is rarely used in practice, so we only discuss lossy video compression methods.

3.2.2 Neural Compression Methods

Recently, neural codecs have surpassed conventional codecs on text, image and video, owing to powerful learning capabilities of neural networks to fit the latent data distribution. However, tactile data exhibit unique statistical patterns, and heterogeneous tactile datasets involve different distributions, potentially making pre-trained neural encoders less applicable. Herein, we briefly introduce the diagram of learning-based lossy and lossless neural codecs, as Fig. 2.

In the case of lossless compression, the tactile signal $\boldsymbol{x}$ is sequentially fed into a neural network f_a to predict the distribution of next symbol, $p(x_i|x_{<i})$, then it is followed by an arithmetic encoder (AE) to generate bitstream. The loss is the entropy, which is the minimal bound to encode $\boldsymbol{x}$:

$$\mathcal{L} = \mathbb{E}[-\log_2(p(x_i|x_{<i}))] \tag{1}$$

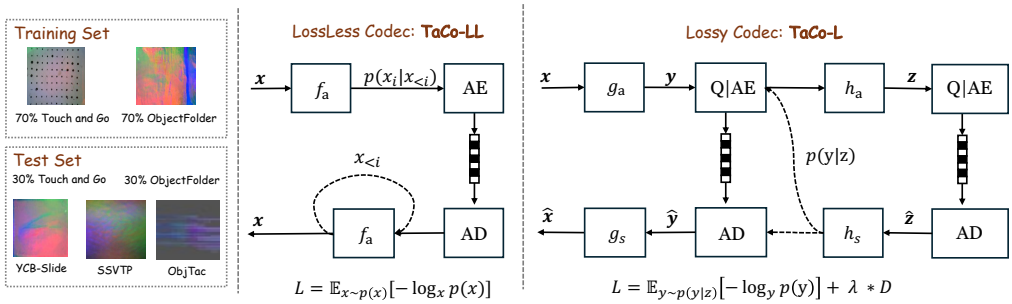


Figure 2: Diagram of data-driven compression methods and our proposed TaCo-L and TaCo-LL.

At the decoder side, the symbols can be lossless autoregressively decoded through an arithmetic decoder (AD) and the same network.

For neural lossy codecs, the tactile signal $\boldsymbol{x}$ is transformed through a transform function g_a into a latent presentation $\boldsymbol{y}$. Afterwards, $\boldsymbol{y}$ is quantized through Q to get discrete values $\hat{\boldsymbol{y}}$, then it is followed by AE to generate bitstream. At the decoder side, $\hat{\boldsymbol{y}}$ is decoded from bitstream using an AD and then transformed back to reconstructed images $\hat{\boldsymbol{x}}$ though an inverse transform function g_s. h_a and h_s denote analysis and synthesis transforms in the hyper autoencoder to generate side bits $\boldsymbol{z}$, as a prior to estimate density model of $\hat{\boldsymbol{y}}$. The loss is defined as a rate-distortion function:

$$\mathcal{L} = \lambda \times \mathcal{D}(\boldsymbol{x}, \hat{\boldsymbol{x}}) + \mathbb{E}[-\log_2(p_{\hat{\boldsymbol{y}}|\hat{\boldsymbol{z}}}(\hat{\boldsymbol{y}}|\hat{\boldsymbol{z}}))] \tag{2}$$

where λ is a hyper-parameter to control the bitrate, and all the parameters are learnable.

Pretrained Neural Codecs The lossless neural-based methods include 5 image compression models, DLPR (Bai et al., 2024), P2LLM (Chen et al., 2024), and DualComp-I (Zhao et al., 2025), as well as LMIC (Deletang et al., 2024), a multi-modality compressor based on pretrained large language models (specifically, in this paper we use RWKV-7B (Peng et al., 2025) and Llama3-8B (AI@Meta, 2024) as LLMs for implementation). It is worth noting that these models are pretrained on natural language or image data, and evaluated on tactile data without any domain-specific adaptations.

For the purpose of lossy neural-based compression approaches, we evaluate a total of 6 compression models, consisting of three recent neural-based image codecs (ELIC (He et al., 2022), TCM (Liu et al., 2023b)) and LALIC (Feng et al., 2025a), and three recent neural-based video compressors (DCVC-DC (Li et al., 2023), DCVC-FM (Li et al., 2024), and DCVC-RT (Jia et al., 2025)).

Tactile Data-Driven Neural Codecs To our knowledge, there have been yet no existing methods fully trained on tactile signals to explore the upper bound of tactile compression performance. To further explore the performance potential of data-driven codec, we retrain two state-of-the-art compression models, DualComp-I and LALIC, using tactile datasets. Specifically, DualComp-I operates lossless compression by tokenizing the input into discrete representations and applying an autoregressive model to predict each token's distributions, enabling efficient entropy coding. LALIC implements lossy compression and follows a variational auto-encoder (VAE) (Doersch, 2016) architecture. The two models are chosen for their competitive performance and efficiency in their respective domains. By retraining them on tactile data, we aim to assess the benefits of data-driven tactile compression. The retrained models are referred to as **TaCo-LL** (lossless) and **TaCo-L** (lossy), respectively, to distinguish them from their original pretrained versions.

Specifically, for **TaCo-LL**, the tokenization is conducted as shown in Fig. 3. We divide the input into $16 \times 16 \times 3$ patches to preserve local spatial correlations. We then flatten the data in a raster-scan order. For visuo-tactile data, including Touch and Go, YCB-slide, ObjectFolder, SSVTP, the RGB values are sequentially expanded as sub-pixels $(R_1, G_1, B_1, R_2, G_2, B_2, \cdots)$. For three-axis force signals, i.e. ObjTac, are treated as three color channels and expanded as $(x_1, y_1, z_1, x_2, y_2, z_2, \cdots)$. For **TaCo-L**, we follow the setup of LALIC [1] and randomly crop or zero-pad the input tactile image to 256×256 resolution. Since the input tensor has three channels for both visuo-tactile data and

[1] https://github.com/sjtu-medialab/RwkvCompress

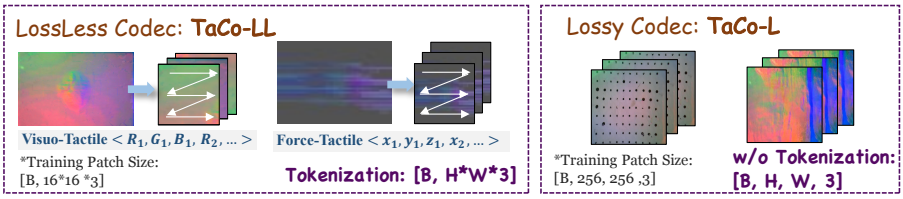


Figure 3: Detailed implementations of our proposed TaCo-L and TaCo-LL.

force-tactile data, no tokenization is needed, as shown in Fig. 3. The network architecture is adopted from the LALIC model (Feng et al., 2025a) and the g_a and g_s consist of four downsampling and upsampling operations, respectively.

To this end, we benchmark a total of **30 codecs** to evaluate the compressibility of tactile data. Among them, 14 codecs (9 off-the-shell codecs, 4 neural codecs and one proposed TaCo-LL) support lossless compression, aiming to preserve exact signal fidelity. The remaining codecs (9 off-the-shell, 6 neural codecs and one proposed TaCo-L) are lossy, targeting higher compression ratios at the cost of some reconstruction distortion. These methods can also be categorized by their training data domain: 28 codecs are existing methods originally developed for general-purpose or visual data and applied without any tactile-specific adaptation, while the remaining two (TaCo-LL, TaCo-L) are data-driven models explicitly trained on tactile datasets. Evaluating the pretrained models allows us to assess how well existing compression techniques generalize to tactile data, whereas the tactile data-driven methods help explore the potential of tactile-aware compression strategies.

4 Experiments

4.1 Experimental Setup

We benchmark the performance of tactile compression on five representative tactile datasets: Touch and Go, ObjectFolder 1.0, SSVTP, YCB-Slide, and ObjTac, as shown in Section 3.1. Specifically, we randomly select 70% of the data from the Touch and Go and ObjectFolder datasets to train the TaCo-LL and TaCo-L models. The remaining 30% of the two datasets, together with the entire SSVTP, YCB-Slide and ObjTac datasets, are used for all methods' compression evaluation. Following (Zhao et al., 2025), for TaCo-LL we use the FusedAdam optimizer (NVIDIA, 2018) with a cosine annealing learning rate schedule Loshchilov & Hutter (2016), starting at 1×10^{-4} and decaying to 2×10^{-5} over 20 epochs. Following (Feng et al., 2025a), we train TaCo-L using the Adam optimizer (Kingma & Ba, 2014). The learning rate is set to 1×10^{-4} for 40 epochs, then decayed to 1×10^{-5} for another 4 epochs. The training is performed on two NVIDIA A100 GPUs.

4.2 Lossless Compression

Evaluation Metrics. We evaluate lossless compression efficiency using bits per Byte, which quantifies the number of bits required to encode one byte of the original tactile data. Lower bits/Byte values indicate more effective compression, with uncompressed data corresponding to 8 bits/Byte. We also compare the complexity of different algorithms using four metrics, i.e. model parameters, MACs, inference speed (KB/s) on multiple devices (including NVIDIA A100 GPU, a MacBook Pro), and the frame per second (FPS) ranging with different spatial resolutions.

Results. Table. 2 benchmarks the lossless compression performance across five tactile datasets. As expected, all methods obviously reduce the data's storage cost, but the degree of compression varies across the compressors and datasets. Among off-the-shelf baselines, general-purpose compressors such as gzip and zstd achieve moderate compression ratios. Image-specific codecs like FLIF and JPEG-XL provide notably better results especially on vision-like tactile data, due to their ability to exploit spatial correlations. Learning-based methods pretrained on natural images, such as DLPR, P2LLM, and DualComp-I, effectively capture intra-frame correlations in tactile signals and generally provide pleasing results. However, their performance remains limited by domain mismatch, especially on non-visual or structurally different tactile datasets.

Table 2: Comparison of lossless compression performance (bits/Byte) on five tactile datasets. The best results are highlighted in **bold blue**, second-best in **bold**, and third to fifth in underline. For TaCo, 12M/48M/96M denotes the model parameter. To show the compression performance more clearly, we also list the compression ratios relative to the uncompressed data (8 bits/Byte) in parentheses only for the best and second best results.

	Compressor	bits/Byte↓				
		TouchandGo	ObjectFolder	SSVTP	ObjTac	YCB-Slide
	uncompressed	8 (1×)	8 (1×)	8 (1×)	8 (1×)	8 (1×)
Off-the-Shelf	gzip (Pasco., 1996)	2.298	3.969	2.234	0.571	2.185
	zstd (Meta., 2015)	2.263	3.966	2.233	0.568	2.184
	bzip2 (Seward, 2000)	2.288	4.031	2.255	0.594	2.205
	FLIF (Sneyers, 2015)	0.808 (10×)	3.765	1.567	**0.363** (22×)	1.489 (5×)
	BPG (Bellard, 2014)	1.293	3.726	2.000	0.513	1.922
	WebP (Google, 2010)	0.936	3.612	1.820	0.424 (19×)	1.767
	JPEG-XL (Team, 2021)	0.739 (11×)	3.657	1.478 (5×)	0.382 (21×)	1.431 (6×)
	JPEG2000 (ISO/IEC, 2000)	1.552	3.989	1.997	1.399	1.916
	PNG (Boutell, 1997)	2.500	3.964	2.233	0.579	2.183
Neural	DLPR (Bai et al., 2024)	1.082	3.774	1.539	0.522	1.503
	P2LLM (Chen et al., 2024)	1.212	3.400	1.804	0.546	1.512
	Llama3* (Deletang et al., 2024)	2.055	3.465	1.975	0.834	1.905
	RWKV* (Deletang et al., 2024)	2.223	3.718	2.010	0.540	1.880
	DualComp-I (Zhao et al., 2025)	0.948	3.126 (3×)	1.442 (6×)	0.540	1.388 (6×)
	TaCo-LL-12M (ours)	0.622 (13×)	3.098 (3×)	1.457 (6×)	0.569	1.520
	TaCo-LL-48M (ours)	**0.504** (16×)	**2.923** (3×)	**1.249** (6×)	0.411 (20×)	**1.321** (6×)
	TaCo-LL-96M (ours)	**0.447** (18×)	**2.709** (3×)	**1.066** (8×)	**0.360** (22×)	**1.073** (8×)

Table 3: The complexity of lossless compression algorithms on five tactile datasets. † and ‡ indicates speeds measured on MacBook Pro CPU and NVIDIA A100 GPU, respectively.

	Compressor	#Params↓	MACs↓	Speed (KB/s)↑	Speed (FPS)↑				
					TouchandGo	ObjectFolder	SSVTP	ObjTac	YCB-Slide
Off-the-Shelf	gzip (Pasco., 1996)	-	-	14500†	15.7†	252†	62.9†	190†	63.9†
	zstd (Meta., 2015)	-	-	11000†	11.9†	191†	47.7†	144†	47.4†
	bzip2 (Seward, 2000)	-	-	3300†	3.58†	57.3†	14.3†	43.3†	14.3†
	FLIF (Sneyers, 2015)	-	-	652†	0.71†	11.3†	2.84†	8.56†	2.84†
	BPG (Bellard, 2014)	-	-	180†	0.20†	3.13†	0.78†	2.36†	0.78†
	WebP (Google, 2010)	-	-	330†	0.36†	5.73†	1.43†	4.33†	1.43†
	JPEG-XL (Team, 2021)	-	-	970†	1.05†	16.8†	4.21†	12.7†	4.21†
	JPEG2000 (ISO/IEC, 2000)	-	-	5000†	5.43†	86.8†	21.7†	65.7†	21.7†
	PNG (Boutell, 1997)	-	-	200†	0.22†	3.47†	0.87†	2.63†	0.87†
Neural	DLPR (Bai et al., 2024)	22.3M	-	640‡	0.69‡	11.1‡	2.78‡	8.41‡	2.78‡
	P2LLM (Chen et al., 2024)	8B	-	20‡	0.02‡	0.35‡	0.09‡	0.26‡	0.09‡
	Llama3* (Deletang et al., 2024)	8B	7.8G	20‡	0.02‡	0.35‡	0.09‡	0.26‡	0.09‡
	RWKV* (Deletang et al., 2024)	7B	7.2G	86‡	0.09‡	1.49‡	0.37‡	1.13‡	0.37‡
	DualComp-I (Zhao et al., 2025)	96M	59.9M	317‡	0.34‡	5.50‡	1.38‡	4.16‡	1.38‡
	TaCo-LL-12M (ours)	12M	11.6M	614‡	0.67‡	10.7‡	2.66‡	8.06‡	2.66‡
	TaCo-LL-48M (ours)	48M	33.3M	360‡	0.39‡	6.25‡	1.56‡	4.73‡	1.56‡
	TaCo-LL-96M (ours)	96M	59.9M	317‡	0.34‡	5.50‡	1.38‡	4.16‡	1.38‡

To further explore the potential of data-driven compression, we retrain state-of-the-art lossless image compressor, DualComp-I, on tactile datasets and obtain our TaCo-LL model. The largest variant, TaCo-LL-96M, achieves the best performance across all five datasets, reaching 0.447 bits/Byte on TouchandGo, 2.709 bits/Byte on ObjectFolder, 1.066 on SSVTP, 0.360 bits/Byte on ObjTac, and 1.073 on TCB-Slide, corresponding to 18×, 3×, 8×, 22×, and 8× compression ratios, respectively. Table. 3 benchmarks the complexity of different compression algorithms. Off-the-shelf codecs can achieve relatively fast speed. Among neural codecs, TaCo-LL models achieve competitive compression performance with fewer parameters compared to P2LLM, Llama3-8B and RWKV-7B, and the encoding/decoding speed ranges from 317KB/s to 614KB/s.

Table 4: Evaluation of lossy compression performance on five tactile datasets leveraging intra-frame compressors. The best results are shown in **blue bold**, the second-best in **bold**, and the third-best in underline. For the reference, the bandwidth consumption of the anchor HEVC-intra is approximately 2Mbps at the quality of 40dB, which is calculated by 0.22 bit per pixel $\times 640 \times 480 \times 30\text{fps} \times 10^{-6}$ for Touch and Go dataset, as Fig. 4.

	Compressor	BD-Rate (%)↓				
		TouchandGo	ObjectFolder	SSVTP	YCB-Slide	ObjTac
Off-the-Shelf	HM-Intra (Sullivan et al., 2012)	0%	0%	0%	0%	0%
	HM-SCC (Sullivan et al., 2012)	-10.4%	2.0%	6.9%	7.2%	**-44.5%**
	VTM-Intra (Bross et al., 2021)	-21.7%	**-19.7%**	**-16.0%**	**-24.4%**	-22.0%
	VTM-SCC (Bross et al., 2021)	-23.7%	-18.0%	-13.7%	-19.1%	**-44.3%**
	JPEG-XL (Team, 2021)	66.7%	60.6%	77.5%	96.9%	99.4%
	JPEG2000 (ISO/IEC, 2000)	59.7%	69.9%	107.9%	89.1%	103.8%
Neural	ELIC (He et al., 2022)	-40.2%	0.6%	-5.8%	-9.2%	44.5%
	LALIC (Feng et al., 2025a)	**-51.6%**	0.2%	4.3%	-4.6%	32.8%
	TCM (Liu et al., 2023b)	-39.9%	23.7%	42.9%	30.5%	97.2%
	TaCo-L (Ours)	**-61.8%**	**-24.3%**	**-19.2%**	**-27.4%**	-27.0%

Table 5: The complexity of lossy compression algorithms on five tactile datasets leveraging intra-frame compressors. † and ‡ indicates speeds measured on MacBook Pro CPU and NVIDIA A100 GPU, respectively.

	Compressor	#Params↓	MACs↓	Speed (KB/s)↑	Speed (FPS)↑				
					TouchandGo	ObjectFolder	SSVTP	YCB-Slide	ObjTac
Off-the-Shelf	HM-Intra (Sullivan et al., 2012)	-	-	11.1†	0.12†	1.97†	0.50†	1.49†	0.50†
	HM-SCC (Sullivan et al., 2012)	-	-	31.3†	0.34†	0.56†	0.41†	0.42†	0.41†
	VTM-Intra (Bross et al., 2021)	-	-	9.22†	0.10†	1.63†	0.41†	1.23†	0.41†
	VTM-SCC (Bross et al., 2021)	-	-	4.61†	0.05†	0.72†	0.18†	0.55†	0.18†
	JPEG-XL (Team, 2021)	-	-	2305†	2.50†	40.0†	10.0†	30.2†	10.0†
	JPEG2000 (ISO/IEC, 2000)	-	-	13200†	14.3†	228†	57.1†	172†	57.1†
Neural	ELIC (He et al., 2022)	33.3M	0.9M	4075‡	4.42‡	70.7‡	17.7‡	53.5‡	17.7‡
	LALIC (Feng et al., 2025a)	63.2M	0.7M	3700‡	4.01‡	64.2‡	16.1‡	48.6‡	16.1‡
	TCM (Liu et al., 2023b)	75.9M	1.8M	5680‡	6.16‡	98.6‡	24.7‡	74.6‡	24.7‡
	TaCo-L (Ours)	63.2M	0.73M	3700‡	4.01‡	64.2‡	16.1‡	48.6‡	16.1‡

4.3 Lossy Compression for Human Vision

Evaluation Metrics. We evaluate lossy compression performance using the Bjøntegaard Delta Rate (BD-Rate) (Bjontegaard, 2001) metric, which quantifies the average bitrate savings at a given level of distortion. A lower BD-Rate indicates better compression efficiency. We measure the reconstruction distortion using Peak Signal-to-Noise Ratio (PSNR) (Rosenfeld & Kak, 1982). The bitrate is assessed in bits per pixel (BPP), where uncompressed data corresponds to 24 BPP.

Results. Table. 4 benchmarks the lossy compression performance when using intra-frame compressors. Off-the-shelf intra-frame codecs like HM-Intra and VTM-Intra provide strong baselines, consistently delivering competitive performance. General-purpose codecs like JPEG2000 and JPEG-XL are included as standard baselines, but their performance is relatively poor. Neural compression methods pretrained on natural images, such as ELIC, LALIC, and TCM, show promising results on some datasets, but they fail to generalize to more structurally distinct data like ObjTac. In contrast, our TaCo-L model, trained on tactile datasets, achieves the best performance across all five datasets. It outperforms all baselines with BD-Rate reductions of -61.8% (TouchandGo), -24.3% (ObjectFolder), -27.4% (YCB-Slide), and -27.0% (ObjTac). Further, the force-based ObjTac dataset, which is derived from 3D force signals and mapped into RGB images, exhibits characteristics similar to screen content (large uniform regions and repetitive patterns). This makes screen-content-optimized codecs, VTM-SCC and HM-SCC, particularly effective on this dataset, achieving BD-Rates of -44.3% and -44.5%, respectively, when taking HM-Intra as the anchor. Table. 5 benchmarks the complexity of different lossy compression algorithms. For off-the-shelf codecs, the complexity increases along with the development of newer generations. For neural codecs, TaCo-L,

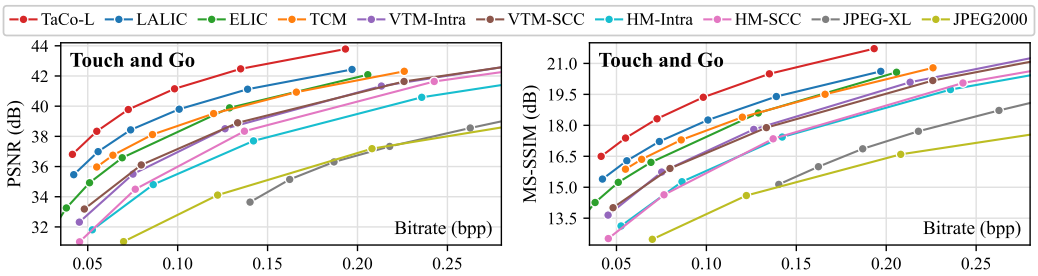


Figure 4: Rate-distortion curves on TouchandGo dataset, when applying intra-frame compression methods.

Table 6: Material classification results on TouchandGo, ObjectFolder-1.0 and object classification results on YCB-Slide. Best results are in **blue bold**, the second-best results are in bold, and the third-best in underline.

	Compressor	BPP	SVM	Random Forest	K-NN	Linear Regression
Touch and Go	Uncompressed	24 (1×)	**76.63%**	**74.88%**	**68.24%**	**73.51%**
	VTM-Intra	0.213	74.08%	72.53%	65.06%	70.87%
	JPEG-XL	0.218	70.67%	69.43%	61.87%	67.75%
	LALIC	0.196	74.70%	73.07%	65.43%	71.24%
	TaCo-L (ours)	0.193 (124×)	**75.12%**	**73.55%**	**66.03%**	**71.89%**
ObjectFolder	Uncompressed	24 (1×)	**44.11%**	**40.68%**	**37.14%**	**42.92%**
	VTM-Intra	0.384	42.23%	39.48%	36.00%	40.71%
	JPEG-XL	0.499	40.27%	37.63%	34.36%	38.74%
	LALIC	0.477	41.00%	38.28%	35.44%	39.91%
	TaCo-L (ours)	0.453 (53×)	**43.08%**	**39.85%**	**36.27%**	**41.02%**
YCB-Slide	Uncompressed	24 (1×)	**98.75%**	**98.72%**	**98.58%**	**99.18%**
	VTM-Intra	0.118	97.36%	96.41%	97.08%	97.24%
	JPEG-XL	0.121	94.08%	93.11%	93.67%	93.97%
	LALIC	0.130	95.67%	95.22%	95.76%	96.23%
	TaCo-L (ours)	0.126 (190×)	**98.01%**	**97.35%**	**97.88%**	**98.20%**

adapted from the latest LALIC, achieves the best compression performance at the cost of incremental complexity, and the encoding/decoding FPS ranges from 4 FPS to 48 FPS at different resolutions.

Fig. 4 presents a representative rate-distortion (RD) curve comparison on the TouchandGo dataset when using image compressors. Additional RD curves are provided in the appendix. We further compare the subjective reconstruction quality by visualizing representative examples from the YCB-Slide dataset, as shown in Fig. 8 in the appendix.

4.4 Lossy Compression for Classification

Evaluation Metrics. We evaluate the semantic fidelity of lossy compression using two tactile understanding tasks: material classification (on TouchandGo and ObjectFolder-1.0) and object classification (on YCB-Slide). For each dataset, we use four standard classifiers, SVM (Burges, 1998), Random Forest (Rigatti, 2017), K-NN (Peterson, 2009), and Linear Regression (Seber & Lee, 2012), with a fixed 60%/40% train-test data split. The top-1 accuracy is used as evaluation metric. Four representative lossy codecs, VTM-Intra, JPEG-XL, LALIC, and TaCo-L, are used for comparison.

Results. Table. 6 As shown in Table. 6, all methods achieve classification performance close to the uncompressed data, despite substantial bitrate savings (e.g., from 24 bpp to as low as 0.118 bpp). On TouchandGo, TaCo-L achieves 75.12% (SVM) and 71.89% (Linear Regression), similar to 76.63% and 73.51% when using uncompressed data. On ObjectFolder, where the task is more challenging, the top-1 accuracy under SVM drops slightly from 44.11% to 43.08% after compression with TaCo-L. On YCB-Slide, TaCo-L also preserves superior classification accuracy (98.01% and 98.20% when using SVM and Linear Regression, respectively), while reducing the bitrate by 190×.

4.5 Lossy Compression for Dexterous Grasping

Table 7: Evaluation results on the dexterous grasping. Best results are shown in **blue bold**, the second-best results are denoted in bold, and the third-best in underline. We also list the accuracy loss relative to the uncompressed data (8 bits/Byte) in parentheses.

	Compressor	BPP	Small Obj.	Medium Obj.	Large Obj.	Deform. Obj.	Avg
s_{lift}	Uncompressed	24	**54.7%**	**67.4%**	**69.2%**	**63.9%**	**63.8%** (-0.0%)
	JPEG-XL	0.0505	47.2%	58.0%	59.7%	55.1%	55.0% (-8.8%)
	VTM-Intra	0.0498	**54.1%**	**66.6%**	**68.4%**	**63.1%**	**63.1%** (-0.7%)
	LALIC	0.0397	51.5%	63.4%	65.0%	60.1%	60.0% (-3.8%)
	TaCo-L (ours)	**0.0251**	53.1%	65.3%	**68.4%**	61.9%	62.2% (-1.6%)
s_{disturb}	Uncompressed	24	**51.8%**	**65.8%**	**67.3%**	**61.8%**	**61.7%** (-0.0%)
	JPEG-XL	0.0505	46.4%	56.1%	57.4%	52.8%	53.2% (-8.5%)
	VTM-Intra	0.0498	**52.5%**	**65.0%**	**66.7%**	**61.1%**	**61.3%** (-0.4%)
	LALIC	0.0397	49.9%	61.8%	63.1%	58.0%	58.2% (-3.5%)
	TaCo-L (ours)	**0.0251**	51.3%	63.6%	65.0%	59.7%	59.9% (-1.8%)

Task Definition. Many contact-rich manipulation algorithms for dexterous hands rely heavily on high-fidelity tactile signals, which motivates us to conduct dexterous grasping experiment. In real-world deployment scenarios, however, tactile data compression may affect the downstream performance of such algorithms. Therefore, we introduce this experiment to evaluate the impact of tactile compression quality on a realistic, task-driven benchmark. The goal of this task is to reach for an object, grasp and lift it. We build the simulation using Nvidia IssacSim Makoviychuk et al. (2021), and use a simple DexHand13 module Paxini. (2024) equipped with eleven tactile sensors. We modify the input tactile signals by compressing them first and then feed it into a tactile-aware reinforcement learning algorithm. In total we use 100 objects to evaluate the grasping performance, consisting of 29 small objects, 41 medium objects and 22 large objects, 8 deformable objects. For the following section, we list the performance for each category.

Evaluation Metrics. To ensure robust interference capabilities during grasping, we evaluate the performance using two evaluation metrics in the simulation. One is the success rate of lifting s_{lift}, recorded when objects maintain stability lifted to the height of 0.1m. The other is the success rate with disturbance resistance s_{disturb}, measured by applying 2.5N external forces along six axes for 2 s after lifting and the object moves below 0.02 m.

Results. Table. 7 benchmarks the grasping performance across different objects. Since the tactile signal simulated by Isaac Sim is relatively sparse, the achieved compression ratio is higher (up to 1000×) than what is typically attainable in the physical world (i.e. physical data achieve at most 22× (ObjTac) compression ratio). All compression methods successfully reduced the raw 24 bpp tactile signal to substantially smaller sizes, ranging from 0.025 bpp to 0.5 bpp, with only a moderate decrement in grasping success rate. Among them, TaCo-L outperforms JPEG-XL and LALIC, achieving a higher compression ratio while maintaining competitive lifting success rate of 62.2% compared to the baseline 63.8% and disturb-resistant grasping success rate of 59.9% compared to the baseline 61.7%. While our method TaCo-L is only approximately 1% of VTM's performance (62.2% vs 63.1% and 59.9% vs 61.3%) in terms of task success rate, it achieves significantly higher compression efficiency by operating at nearly half the bitrate (0.0251bpp vs. 0.0498 bpp).

5 Conclusion

This paper introduced the **TaCo** benchmark, the first comprehensive framework for evaluating tactile data codecs. This is a suite of 30 codecs, 5 datasets and 4 types of tasks to advance the research on tactile sensing and tactile data compression. Further, we presented **TaCo-LL** and **TaCo-L**, data-driven codecs that learn the latent distribution of tactile data end-to-end. Extensive experiments demonstrate that our proposed models establish a new state-of-the-art result, outperforming existing methods across lossless/lossy compression, classification, and grasping tasks. Our work provides a critical foundation and a baseline for future research in efficient tactile perception and transmission.

ETHICS STATEMENT

This work does not involve any sensitive personal data, private information, or high-risk deployment scenarios. Our evaluation relies solely on publicly available tactile datasets. No human subjects, practices to data set releases, potentially harmful insights, methodologies and applications, potential conflicts of interest and sponsorship, discrimination/bias/fairness concerns, privacy and security issues, legal compliance, and research integrity issues were involve.

REPRODUCIBILITY STATEMENT

Our experimental results rely on already published, publicly available datasets and compression models. For novel models or algorithms, a link to an anonymous downloadable source code can be submitted as supplementary materials; for theoretical results, clear explanations of any assumptions and a complete proof of the claims can be included in the appendix; for any datasets used in the experiments, a complete description of the data processing steps can be provided in the supplementary materials. More implementation and training details are explained in the appendix. We will release the code base used for experiments in Sec. 4 along with the code for evaluating our benchmark.

REFERENCES

AI@Meta. Llama 3 model card. 2024. URL https://github.com/meta-llama/llama3/blob/main/MODEL_CARD.md.

Yuanchao Bai, Xianming Liu, Kai Wang, Xiangyang Ji, Xiaolin Wu, and Wen Gao. Deep lossy plus residual coding for lossless and near-lossless image compression. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(5):3577–3594, 2024.

Chiara Bartolozzi, Paolo Motto Ros, Francesco Diotalevi, Nawid Jamali, Lorenzo Natale, Marco Crepaldi, and Danilo Demarchi. Event-driven encoding of off-the-shelf tactile sensors for compression and latency optimisation for robotic skin. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 166–173. IEEE, 2017.

Fabrice Bellard. Bpg image format. https://bellard.org/bpg/, 2014. Accessed: 2025-02-28.

Gisle Bjontegaard. Calculation of average psnr differences between rd-curves. *ITU SG16 Doc. VCEG-M33*, 2001.

Thomas Boutell. *PNG (Portable Network Graphics) Specification Version 1.0*. W3C, 1997. URL https://www.w3.org/TR/PNG/.

Benjamin Bross, Ye-Kui Wang, Yan Ye, Shan Liu, Jianle Chen, Gary J. Sullivan, and Jens-Rainer Ohm. Overview of the versatile video coding (vvc) standard and its applications. *IEEE Transactions on Circuits and Systems for Video Technology*, 31:3736–3764, 2021.

Christopher JC Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167, 1998.

Kecheng Chen, Pingping Zhang, Hui Liu, Jie Liu, Yibing Liu, Jiaxin Huang, Shiqi Wang, Hong Yan, and Haoliang Li. Large language models for lossless image compression: Next-pixel prediction in language space is all you need. *arXiv preprint arXiv:2411.12448*, 2024.

Ning Cheng, Jinan Xu, Changhao Guan, Jing Gao, Weihao Wang, You Li, Fandong Meng, Jie Zhou, Bin Fang, and Wenjuan Han. Touch100k: A large-scale touch-language-vision dataset for touch-centric multimodal representation. *Information Fusion*, pp. 103305, 2025a.

Zhengxue Cheng, Yiqian Zhang, Wenkang Zhang, Haoyu Li, Keyu Wang, Li Song, and Hengdi Zhang. Omnivtla: Vision-tactile-language-action model with semantic-aligned tactile sensing. *arXiv preprint arXiv:2508.08706*, 2025b.

Gregoire Deletang, Anian Ruoss, Paul-Ambroise Duquenne, Elliot Catt, Tim Genewein, Christopher Mattern, Jordi Grau-Moya, Li Kevin Wenliang, Matthew Aitchison, Laurent Orseau, Marcus Hutter, and Joel Veness. Language modeling is compression. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=jznbgiynus.

Carl Doersch. Tutorial on variational autoencoders. *arXiv preprint arXiv:1606.05908*, 2016.

Donghui Feng, Zhengxue Cheng, Shen Wang, Ronghua Wu, Hongwei Hu, Guo Lu, and Li Song. Linear attention modeling for learned image compression. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 7623–7632, 2025a.

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. *arXiv preprint arXiv:2502.12191*, 2025b.

Ruohan Gao, Yen-Yu Chang, Shivani Mall, Li Fei-Fei, and Jiajun Wu. Objectfolder: A dataset of objects with implicit visual, auditory, and tactile representations. *arXiv preprint arXiv:2109.07991*, 2021.

Ruohan Gao, Zilin Si, Yen-Yu Chang, Samuel Clarke, Jeannette Bohg, Li Fei-Fei, Wenzhen Yuan, and Jiajun Wu. Objectfolder 2.0: A multisensory object dataset for sim2real transfer. In *CVPR*, 2022.

Google. Webp image format. `https://developers.google.com/speed/webp`, 2010. Accessed: 2025-02-28.

Jingning Han, Bohan Li, Debargha Mukherjee, Ching-Han Chiang, Adrian Grange, Cheng Chen, Hui Su, Sarah Parker, Sai Deng, Urvang Joshi, et al. A technical overview of av1. *Proceedings of the IEEE*, 109(9): 1435–1462, 2021.

Rania Hassen, Basak Gülecyüz, and Eckehard Steinbach. Pvc-slp: Perceptual vibrotactile-signal compression based-on sparse linear prediction. *IEEE Transactions on Multimedia*, 23:4455–4468, 2020.

Dailan He, Zi Yang, Weikun Peng, Rui Ma, Hongwei Qin, and Yan Wang. Elic: Efficient learned image compression with unevenly grouped space-channel contextual adaptive coding. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5708–5717, 2022. URL `https://api.semanticscholar.org/CorpusID:247594672`.

Carolina Higuera, Akash Sharma, Chaithanya Krishna Bodduluri, Taosha Fan, Patrick Lancaster, Mrinal Kalakrishnan, Michael Kaess, Byron Boots, Mike Lambeta, Tingfan Wu, et al. Sparsh: Self-supervised touch representations for vision-based tactile sensing. *arXiv preprint arXiv:2410.24090*, 2024.

Brayden Hollis, Stacy Patterson, and Jeff Trinkle. Compressed sensing for tactile skins. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 150–157. IEEE, 2016.

Brayden Hollis, Stacy Patterson, and Jeff Trinkle. Compressed learning for tactile object recognition. *IEEE Robotics and Automation Letters*, 3(3):1616–1623, 2018.

ISO/IEC. Jpeg 2000 image coding system. `https://www.jpeg.org/jpeg2000/`, 2000. Accessed: 2025-02-28.

Zhaoyang Jia, Bin Li, Jiahao Li, Wenxuan Xie, Linfeng Qi, Houqiang Li, and Yan Lu. Towards practical real-time neural video compression. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 12543–12552, 2025.

Justin Kerr, Huang Huang, Albert Wilcox, Ryan Hoque, Jeffrey Ichnowski, Roberto Calandra, and Ken Goldberg. Self-supervised visuo-tactile pretraining to locate and follow garment features. *arXiv preprint arXiv:2209.13042*, 2022.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Mike Lambeta, Po-Wei Chou, Stephen Tian, Brian Yang, Benjamin Maloon, Victoria Rose Most, Dave Stroud, Raymond Santos, Ahmad Byagowi, Gregg Kammerer, et al. Digit: A novel design for a low-cost compact high-resolution tactile sensor with application to in-hand manipulation. *IEEE Robotics and Automation Letters*, 5(3):3838–3845, 2020.

Jiahao Li, Bin Li, and Yan Lu. Neural video compression with diverse contexts. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 22616–22626, 2023.

Jiahao Li, Bin Li, and Yan Lu. Neural video compression with feature modulation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 26099–26108, 2024.

Yang Li, Yan Zhao, Zhengxue Cheng, and Hengdi Zhang. Taccompress: A benchmark for multi-point tactile data compression in dexterous manipulation. *arXiv preprint arXiv:2505.16289*, 2025.

Yunzhu Li, Jun-Yan Zhu, Russ Tedrake, and Antonio Torralba. Connecting touch and vision via cross-modal prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10609–10618, 2019.

Guohong Liu, Xiaomeng Li, Cong Wang, and Shuai Lv. Online compression and reconstruction for communication of force-tactile signals. *IEEE Communications Letters*, 27(3):981–985, 2023a.

Jinming Liu, Heming Sun, and Jiro Katto. Learned image compression with mixed transformer-cnn architectures. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 14388–14397, 2023b.

Qingtao Liu, Yu Cui, Zhengnan Sun, Gaofeng Li, Jiming Chen, and Qi Ye. Vtdexmanip: A dataset and benchmark for visual-tactile pretraining and dexterous manipulation with reinforcement learning. In *The Thirteenth International Conference on Learning Representations*.

Tianyi Liu and Benjamin Ward-Cherrier. The tip benchmark: A tactile image-based psychophysics-inspired benchmark for artificial tactile sensors. In *International Conference on Human Haptic Sensing and Touch Enabled Computer Applications*, pp. 94–106. Springer, 2024.

Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*, 2016.

Hang Lu, Xinmeng Tan, Mingkai Chen, Zhe Zhang, Xuguang Zhang, Jianxin Chen, Xin Wei, and Tiesong Zhao. Cross-modal haptic compression inspired by embodied ai for haptic communications. *IEEE Transactions on Multimedia*, 2025.

Siwei Ma, Xinfeng Zhang, Chuanmin Jia, Zhenghui Zhao, Shiqi Wang, and Shanshe Wang. Image and video compression with neural networks: A review. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(6):1683–1698, 2019.

Viktor Makoviychuk, Lukasz Wawrzyniak, Yunrong Guo, Michelle Lu, Kier Storey, Miles Macklin, David Hoeller, Nikita Rudin, Arthur Allshire, Ankur Handa, and Gavriel State. Isaac gym: High performance gpu-based physics simulation for robot learning, 2021.

Fabian Mentzer, Eirikur Agustsson, Michael Tschannen, Radu Timofte, and Luc Van Gool. Practical full resolution learned lossless image compression. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10629–10638, 2019. doi: 10.1109/CVPR.2019.01089.

Meta. ZSTD, Zstandard - Fast real-time compression algorithm. `https://github.com/facebook/zstd`, 2015. Accessed: 2024-08-10.

NVIDIA. Apex (a pytorch extension). `https://nvidia.github.io/apex/optimizers.html`, 2018. URL `https://nvidia.github.io/apex/`. API Documentation for NVidia's Apex optimizers.

Abby O'Neill and et al. Rehman. Open x-embodiment: Robotic learning datasets and rt-x models : Open x-embodiment collaboration0. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 6892–6903, 2024. doi: 10.1109/ICRA57147.2024.10611477.

Richard C. Pasco. GZIP file format specification version 4.3. *RFC 1952*, 1996.

Paxini. Paxini DexH13 Gen2. `https://paxini.com/dex/gen2`, 2024. Accessed: 2025-09-23.

Paxini. Px-6ax: Itpu tactile processing unit. `https://paxini.com/ax/gen2`, 2025. Accessed: 2025-09-16.

Bo Peng, Ruichong Zhang, Daniel Goldstein, Eric Alcaide, Haowen Hou, Janna Lu, William Merrill, Guangyu Song, Kaifeng Tan, Saiteja Utpala, et al. Rwkv-7" goose" with expressive dynamic state evolution. *arXiv preprint arXiv:2503.14456*, 2025.

Leif E Peterson. K-nearest neighbor. *Scholarpedia*, 4(2):1883, 2009.

Steven J Rigatti. Random forest. *Journal of Insurance Medicine*, 47(1):31–39, 2017.

A. Rosenfeld and A. C. Kak. Digital picture processing. *Academic Press*, 1982.

Tim Schneider, Guillaume Duret, Cristiana de Farias, Roberto Calandra, Liming Chen, and Jan Peters. Tactile mnist: Benchmarking active tactile perception. *arXiv preprint arXiv:2506.06361*, 2025.

George AF Seber and Alan J Lee. *Linear regression analysis*. John Wiley & Sons, 2012.

Patrick Seeling, Martin Reisslein, and Frank HP Fitzek. Real-time compression for tactile internet data streams. *Sensors*, 21(5):1924, 2021.

Julian Seward. On the Performance of BWT Sorting Algorithms. *Proceedings of the IEEE Data Compression Conference 2000*, 2000.

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

Ariel Slepyan, Michael Zakariaie, Trac Tran, and Nitish Thakor. Wavelet transforms significantly sparsify and compress tactile interactions. *Sensors*, 24(13):4243, 2024.

Jon Sneyers. Flif - free lossless image format. `https://flif.info/`, 2015. Accessed: 2023-10-01.

Gary J. Sullivan, Jens-Rainer Ohm, Woojin Han, and Thomas Wiegand. Overview of the high efficiency video coding (hevc) standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 22:1649–1668, 2012.

Sudharshan Suresh, Zilin Si, Stuart Anderson, Michael Kaess, and Mustafa Mukadam. Midastouch: Monte-carlo inference over distributions across sliding touch. In *Conference on Robot Learning*, pp. 319–331. PMLR, 2023.

Sudharshan Suresh, Haozhi Qi, Tingfan Wu, Taosha Fan, Luis Pineda, Mike Lambeta, Jitendra Malik, Mrinal Kalakrishnan, Roberto Calandra, Michael Kaess, et al. Neuralfeels with neural fields: Visuotactile perception for in-hand manipulation. *Science Robotics*, 9(96):eadl0628, 2024.

JPEG XL Team. Jpeg xl image coding system. `https://jpeg.org/jpegxl/`, 2021. Accessed: 2025-02-28.

Keyu Wang, Bingcong Lu, Zhengxue Cheng, Hengdi Zhang, and Li Song. D3grasp: Diverse and deformable dexterous grasping for general objects, 2025. URL `https://arxiv.org/abs/2509.19892`.

Fengyu Yang, Chenyang Ma, Jiacheng Zhang, Jing Zhu, Wenzhen Yuan, and Andrew Owens. Touch and go: Learning from human-collected vision and touch. *arXiv preprint arXiv:2211.12498*, 2022.

Samson Yu, Kelvin Lin, Anxing Xiao, Jiafei Duan, and Harold Soh. Octopi: Object property reasoning with large tactile-language models. *arXiv preprint arXiv:2405.02794*, 2024.

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

Wenzhen Yuan, Shaoxiong Wang, Siyuan Dong, and Edward Adelson. Connecting look and feel: Associating the visual and tactile properties of physical materials. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5580–5588, 2017b.

Wenzhen Yuan, Yuchen Mo, Shaoxiong Wang, and Edward Adelson. Active clothing material perception using tactile sensing and deep learning, 2018a. URL `https://arxiv.org/abs/1711.00574`.

Wenzhen Yuan, Yuchen Mo, Shaoxiong Wang, and Edward H Adelson. Active clothing material perception using tactile sensing and deep learning. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4842–4849. IEEE, 2018b.

Jialiang Zhao, Yuxiang Ma, Lirui Wang, and Edward H Adelson. Transferable tactile transformers for representation learning across diverse sensors and tasks. *arXiv preprint arXiv:2406.13640*, 2024.

Yan Zhao, Zhengxue Cheng, Junxuan Zhang, Qunshan Gu, Qi Wang, and Li Song. Dualcomp: End-to-end learning of a unified dual-modality lossless compressor. *arXiv preprint arXiv:2505.16256*, 2025.

A APPENDIX

A.1 BASELINE AND IMPLEMENTATION DETAILS

We benchmark the performance of tactile compression on five representative tactile datasets: Touch and Go[2], ObjectFolder 1.0[3], SSVTP[4], YCB-Slide[5], and ObjTac[6], as detailed in Section 3.1. Specifically, 70% of the samples from Touch and Go and ObjectFolder are used for training, while the remaining 30%, along with the full SSVTP and ObjTac datasets, are used for evaluation. For the Touch and Go dataset, while the official guideline recommends splitting by collection trajectories, it does not specify an exact train/test ratio or content. We followed this recommendation by grouping data at the trajectory level and applied a common 70% and 30% split for training and testing. Each trajectory was then decomposed into individual frames, ensuring that all frames from the same trajectory are contained entirely within either the training or the testing set, avoiding any data leakage.

For **TaCo-LL**, adapted from DualComp-I (Zhao et al., 2025), we train it using the FusedAdam optimizer (NVIDIA, 2018) with a cosine annealing learning rate schedule. (Loshchilov & Hutter, 2016), starting from 1×10^{-4} and decaying to 2×10^{-5} over 20 epochs. The batch size is set to 64. The model is trained with a standard cross-entropy loss:

$$\mathcal{L}_{\text{TaCo-LL}} = -\sum q \log p \tag{3}$$

where q and p are the target and predicted distributions, respectively.

For **TaCo-L**, We train our models using the Adam optimizer (Kingma & Ba, 2014) with a batch size of 8. The model is optimized with a rate-distortion loss:

$$\mathcal{L}_{\text{TaCo-L}} = R + \lambda \cdot \text{MSE} \tag{4}$$

where R denotes the estimated bitrate and λ controls the trade-off between rate and distortion. For MSE-optimized models, λ is set to $\{0.0018, 0.0067, 0.025, 0.0483\}$ to achieve different bitrates. The learning rate is set to 1×10^{-4} for 40 epochs, and then decayed to 1×10^{-5} for another 4 epochs. During training, input tactile images are randomly cropped or padded to 256×256 resolution.

These two models are trained using two NVIDIA A100 GPUs with mixed precision enabled.

A.2 CROSS-DATASET COMPRESSION PERFORMANCE

Furthermore, the above training datasets are mainly collected on rigid and lambertian objects, and we also validate compression performance on two new test datasets: Active Cloth (Yuan et al., 2018a) covering soft and textured objects, and ObjectFolder-2.0 comprising a wide variety of everyday 3D objects, as shown in Table. 8. Due to the large scale of both datasets, we conduct quick validation on approximately the first 10% of the data from each: 10 objects from Active Cloth and 100 objects from ObjectFolder-2.0. Lossless and lossy compression are performed, as summarized in Table 9 and Table 10.

Table 8: Introduction of two additional tactile datasets, where ActiveCloth (Yuan et al., 2018a) consists of 153 varied pieces of clothes and ObjectFolder-2.0 (Gao et al., 2022) mainly extends ObjectFolder-1.0 (Gao et al., 2021) with 100 virtualized objects to 1000 common household real objects.

Dataset	#Objects	#Frames	Resolution	Sensor
ActiveCloth (Yuan et al., 2018a)	153	494655	$640 \times 480 \times 30$Hz	GelSight (Yuan et al., 2017a)
ObjectFolder 2.0 (Gao et al., 2022)	1000	76000	$120 \times 160 \times 30$Hz	GelSight (Yuan et al., 2017a)

When comparing Active Cloth and Touch and Go at the same resolution of 640×480, our TaCo-LL model, with 96M parameters, achieves the best performance on both datasets. It achieves 0.723

[2] https://touch-and-go.github.io/
[3] https://objectfolder.stanford.edu/
[4] https://sites.google.com/berkeley.edu/ssvtp
[5] https://github.com/rpl-cmu/YCB-Slide
[6] https://readerek.github.io/Objtac.github.io/

Table 9: Comparison of lossless compression performance (bits/Byte) on two additional tactile datasets to validate the cross-dataset performance. The best results are highlighted in **bold blue**, second-best in **bold**, and third in underline. For TaCo-LL, 12M/48M/96M denotes the model parameter. To show the compression performance more clearly, we also list the compression ratios relative to the uncompressed data (8 bits/Byte) in parentheses only for top three results.

	Compressor	bits/Byte↓	
		ActiveCloth	ObjectFolder-2.0
	uncompressed	8 (1×)	8 (1×)
Off-the-Shelf	gzip (Pasco., 1996)	2.762	4.040
	zstd (Meta., 2015)	2.771	4.037
	bzip2 (Seward, 2000)	2.771	4.103
	FLIF (Sneyers, 2015)	0.882	3.831
	BPG (Bellard, 2014)	1.645	3.792
	WebP (Google, 2010)	1.063	3.676
	JPEG-XL (Team, 2021)	0.841 (9.5×)	3.659
	JPEG2000 (ISO/IEC, 2000)	1.804	4.061
	PNG (Boutell, 1997)	2.667	4.035
Neural	DLPR (Bai et al., 2024)	1.453	3.852
	P2LLM (Chen et al., 2024)	2.193	3.470
	Llama3* (Deletang et al., 2024)	2.620	3.659
	RWKV* (Deletang et al., 2024)	2.640	3.800
	DualComp-I (Zhao et al., 2025)	1.158	3.308
	TaCo-LL-12M (ours)	1.059	3.179 (2.5×)
	TaCo-LL-48M (ours)	**0.816** (10×)	**3.002** (2.7×)
	TaCo-LL-96M (ours)	**0.723** (11×)	**2.855** (2.8×)

Table 10: Lossy compression performance (BD-Rate) on two additional tactile datasets to validate the cross-dataset performance. The best results are shown in **blue bold**, the second-best in **bold**, and the third-best in underline. For TaCo, 12M/48M/96M denotes the model parameter.

	Compressor	BD-Rate (%)↓	
		ActiveCloth	ObjectFolder-2.0
Off-the-Shelf	HM-Intra (Sullivan et al., 2012)	0%	0%
	HM-SCC (Sullivan et al., 2012)	-12.9%	2.2%
	VTM-Intra (Bross et al., 2021)	-28.0%	**-21.0%**
	VTM-SCC (Bross et al., 2021)	-26.1%	-19.3%
	JPEG-XL (Team, 2021)	46.9%	80.7%
	JPEG2000 (ISO/IEC, 2000)	86.5%	79.0%
Neural	ELIC (He et al., 2022)	**-57.1%**	3.2%
	LALIC (Feng et al., 2025a)	-54.8%	2.8%
	TCM (Liu et al., 2023b)	-49.8%	23.7%
	TaCo-L (Ours)	**-65.4%**	**-26.4%**

bit/Byte (in Table 9) and 0.447 bit/Byte (in Table. 2), corresponding to compression ratios of 11× and 18×, respectively. The results also suggest that soft objects in Active Cloth are more difficult to compress than rigid objects, as deformable surfaces tend to generate more complex tactile data. When comparing ObjectFolder-1.0 and ObjectFolder-2.0 at the same resolution of 120 × 160, all the compression methods basically achieve consistent results and our TaCo-LL also achieve the best performance with 2.855 bit/Byte, corresponding to compression ratios of 2.8×.

These findings are further supported by the BD-Rate comparisons in Table 10, where TaCo-L consistently achieves the lowest BD-Rate across both ActiveCloth and ObjectFolder-2.0 datasets, outperforming state-of-the-art neural compressors such as ELIC, LALIC, and TCM.

A.3 Cross-object Compression Performance

Table. 11 and Table. 12 present a evaluation of cross-object lossless compression performance (in bits/Byte). Table. 11 focuses on rigid objects from TouchandGo and ObjTac datasets, while Table. 12 evaluates soft objects from ActiveCloth and ObjTac datasets. A key observation is that compression performance is influenced primarily by the type of sensor modality, as evidenced by consistent trends

Table 11: Cross-object lossless compression performance (bits/Byte) on RIGID objects. The best results are highlighted in **bold blue**, the second-best in **bold**, and the third in underline. In TouchandGo, Tree and Wood are training objects, while Concrete is unseen. The three objects in ObjTac are all unseen objects (*).

	Compressor	bits/Byte↓					
		Touch and Go			ObjTac		
		Tree	Wood	Concrete*	Stone*	Pebble*	Tile*
	uncompressed	8 (1×)	8 (1×)	8 (1×)	8 (1×)	8 (1×)	8 (1×)
Off-the-Shelf	gzip (Pasco., 1996)	2.531	2.082	2.214	1.100	0.943	0.529
	zstd (Meta., 2015)	2.327	2.033	2.080	1.098	0.930	0.505
	bzip2 (Seward, 2000)	2.486	2.068	2.186	1.123	0.976	0.541
	FLIF (Sneyers, 2015)	0.865	0.737	0.782	0.697	0.648	0.303
	BPG (Bellard, 2014)	1.395	1.141	1.246	1.061	0.847	0.453
	WebP (Google, 2010)	1.000	0.855	0.924	0.848	0.720	0.387
	JPEG-XL (Team, 2021)	0.796	**0.670**	**0.730**	0.742	0.656	0.372
	JPEG2000 (ISO/IEC, 2000)	1.617	1.421	1.540	1.181	0.990	0.500
	PNG (Boutell, 1997)	2.765	2.249	2.390	1.097	0.939	0.527
Neural	DLPR (Bai et al., 2024)	1.127	0.935	1.062	0.906	0.917	0.551
	P2LLM (Chen et al., 2024)	1.946	1.475	1.832	0.933	0.911	0.534
	Llama3* (Deletang et al., 2024)	2.479	2.010	2.145	1.098	1.102	0.809
	RWKV* (Deletang et al., 2024)	2.558	2.120	2.396	1.175	1.110	0.832
	DualComp-I (Zhao et al., 2025)	0.840	0.726	0.857	0.810	0.685	0.339
	TaCo-LL-12M (ours)	0.810	0.704	0.815	0.796	0.680	0.336
	TaCo-LL-48M (ours)	**0.719**	**0.611**	**0.730**	**0.635**	**0.627**	**0.300**
	TaCo-LL-96M (ours)	**0.607**	**0.598**	**0.700**	**0.590**	**0.596**	**0.288**

Table 12: Cross-object lossless compression performance (bits/Byte) on SOFT objects. The best results are highlighted in **bold blue**, the second-best in **bold**, and the third in underline. All these objects are unseen (*).

	Compressor	bits/Byte↓					
		Active Cloth			ObjTac		
		Cloth-12*	Cloth-29*	Cloth-33*	Sponge*	Jeans*	Leather Bag*
	uncompressed	8 (1×)	8 (1×)	8 (1×)	8 (1×)	8 (1×)	8 (1×)
Off-the-Shelf	gzip (Pasco., 1996)	3.540	1.902	3.609	0.214	0.178	0.129
	zstd (Meta., 2015)	3.550	1.911	3.619	0.210	0.173	0.128
	bzip2 (Seward, 2000)	3.552	1.907	3.619	0.252	0.206	0.144
	FLIF (Sneyers, 2015)	1.097	**0.668**	**1.101**	0.106	0.079	**0.071**
	BPG (Bellard, 2014)	2.043	1.194	2.064	0.207	0.148	0.144
	WebP (Google, 2010)	1.319	0.802	1.323	0.178	0.092	**0.065**
	JPEG-XL (Team, 2021)	**1.055**	**0.621**	**1.057**	0.106	**0.076**	0.088
	JPEG2000 (ISO/IEC, 2000)	2.143	1.419	2.140	0.246	0.263	0.204
	PNG (Boutell, 1997)	3.549	1.586	3.619	0.213	0.197	0.165
Neural	DLPR (Bai et al., 2024)	1.877	1.590	1.985	0.148	0.094	0.150
	P2LLM (Chen et al., 2024)	2.033	1.724	2.082	0.170	0.142	0.143
	Llama3* (Deletang et al., 2024)	3.147	1.883	3.251	0.492	0.185	0.158
	RWKV* (Deletang et al., 2024)	3.219	1.890	3.238	0.510	0.179	0.163
	DualComp-I (Zhao et al., 2025)	1.696	1.125	2.000	0.105	0.109	0.110
	TaCo-LL-12M (ours)	1.710	1.147	1.991	0.106	0.100	0.113
	TaCo-LL-48M (ours)	1.332	0.877	1.930	**0.100**	0.078	0.095
	TaCo-LL-96M (ours)	**1.016**	0.725	1.255	**0.097**	**0.053**	0.079

within datasets from the same source (e.g., ActiveCloth vs. TouchandGo). However, the physical properties of the object (like rigidity or softness) also have an obvious impact.

A.4 Analysis of Tactile Data Characteristic

Fig. 5 includes samples from YCB-Slide (Digit sensor, rigid objects such as Sugar Box, Mug, and Mustard Bottle), SSVTP (Digit sensor, soft cloth objects like Cloth Corner and Interior), Active Cloth (GelSight sensor, soft fabrics including Cloth-12, 29, and 33), Touch and Go (GelSight sensor, rigid surfaces such as Tree, Wood, and Concrete), and ObjTac (Force sensor, both soft objects like Jeans, Leather Bag, Sponge, and rigid ones like Pebble, Stone, Tile). As depicted, the entropy maps and FFT spectra show that tactile images are dominated by low-frequency energy with highly repetitive, grid-like spatial structures. These signals exhibit strong directional patterns and

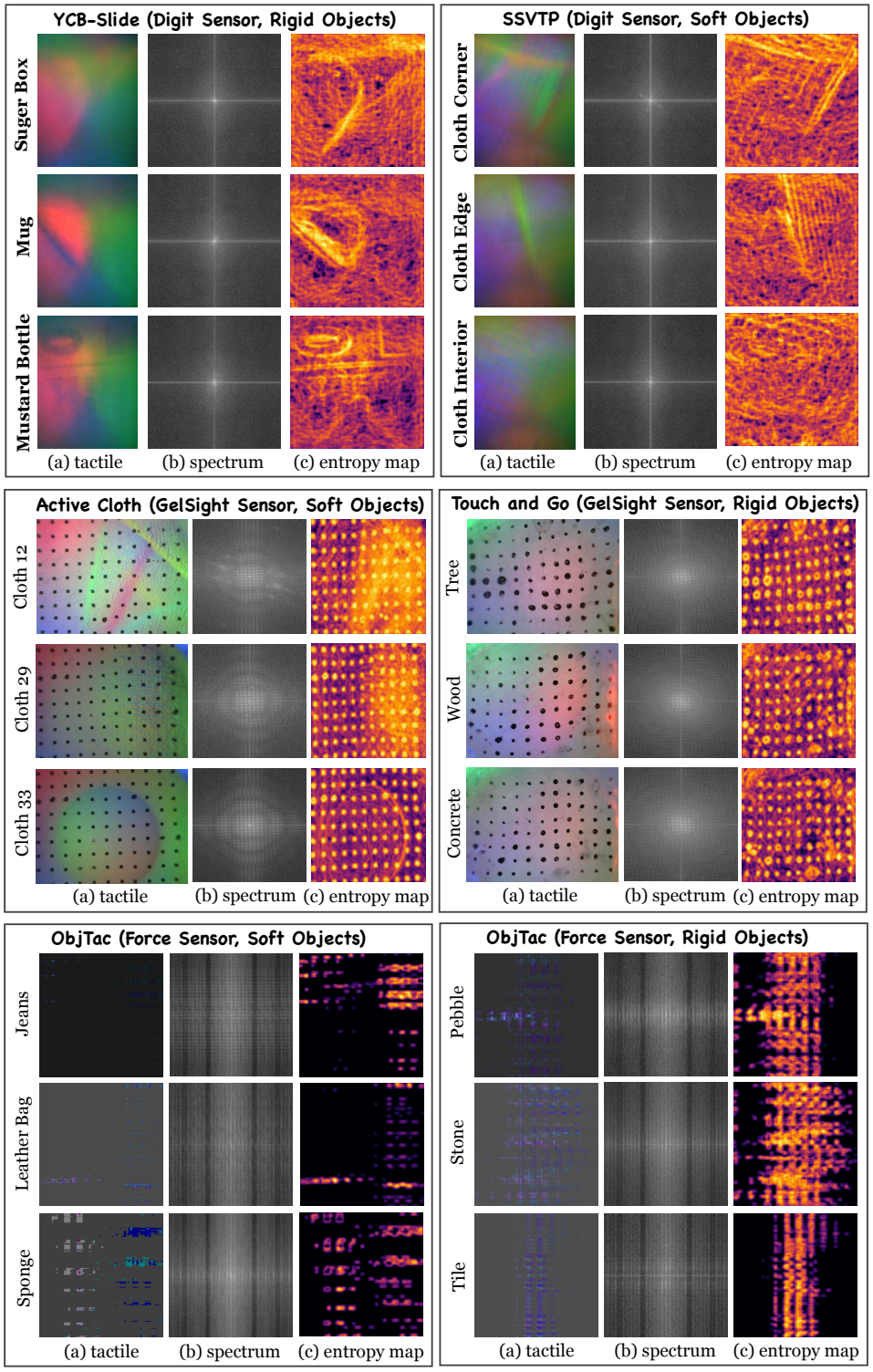


Figure 5: Visualization of tactile data characteristics across different datasets, sensors, and object types. Each subfigure displays the raw tactile image, its frequency spectrum, and the corresponding entropy map.

locally predictable regions, leading to sparse residuals after prediction. As a result, block-based lossy codecs such as SCC perform especially well, since their intra prediction, palette modes, and transform coding are optimized for smooth, structured, and repetitive content. The same properties

also explain the behavior of lossless codecs: low entropy regions compress extremely well, while periodic patterns favor context-based or LZ-type entropy models.

A.5 More Lossy Compression Performance for Human Vision

In addition to intra-frame compression methdos, Table. 13 benchmarks the use of video codecs for compressing tactile data, focusing on their ability to exploit inter-frame redundancy. It can be seen that DCVC-RT achieves the best performance on most dataset, followed by DCVC-FM and VVenC. Since each row of an ObjTac image corresponds to a distinct timestamp, the dataset does not have video format and video compression evaluation.

Table 13: Evaluation on lossy compression performance with regard to intra-frame and inter-frame correlations. The best results are denoted in **bold**, and the second-best in underline

	Compressor	BD-Rate (%)↓			
		TouchandGo	ObjectFolder	SSVTP	YCB-Slide
Off-the-Shelf	x265 (Sullivan et al., 2012)	0%	0%	0%	0%
	SVT-AV1 (Han et al., 2021)	-40.6%	-34.2%	-28.2%	-12.1%
	VVenC (Bross et al., 2021)	-67.6%	-16.4%	-33.6%	-52.2%
Neural	DCVC-DC (Li et al., 2023)	-75.6%	-12.2 %	-20.4%	-27.1%
	DCVC-FM (Li et al., 2024)	**-80.0%**	-43.8%	-45.2%	-58.1%
	DCVC-RT (Jia et al., 2025)	-78.1%	**-48.8%**	**-50.9%**	**-65.5%**

To complement the BD-Rate results in Table. 5 and Table. 13, we present full rate-distortion (RD) curves for each dataset in both image-based (Fig. 6) and video-based (Fig. 7) compression settings, using PSNR and MS-SSIM as distortion metrics. For the TouchandGo dataset, image-based RD curves are shown in Fig. 4 of the main text.

These curves provide a more detailed view of compression performance across bitrates. In the image-based setting, JPEG-XL and JPEG2000 consistently result to relatively poor performance. TaCo-L consistently achieves the best performance across all datasets except ObjTac, where screen-content-coding (SCC) modes in VTM and HM are particularly effective due to the screen-content-like patterns. In the video-based setting, neural codecs like DCVC variants outperform traditional video codecs like x265, SVT-AV1 and VVenC, especially in the low-bitrate region.

Aside from objective metrics, We also use the YCB-Slide dataset as an example and provide per-pixel RMSE error maps of the reconstructed tactile signals in Fig. 8. As discussed in Fig. 1 of the main paper, although tactile images carry meaningful physical information, their visual appearance is often unintuitive for human interpretation. Therefore, instead of relying on perceptual inspection, we quantify local reconstruction discrepancies using the pixel-wise RMSE,

$$\mathrm{RMSE}(x, \hat{x}) = \sqrt{(x - \hat{x})^2}. \tag{5}$$

As shown in Fig. 8, all methods produce relatively low reconstruction errors even at low bitrates (below 0.1 bpp, i.e., over $240\times$ compression), indicating that lossy compression at relatively high ratios is generally acceptable for human-viewing purposes.

A.6 More Lossy Compression Results for Classification

To complement the results in Table. 6 and illustrate the full-bitrate performance, we present the bitrate-accuracy curves on Touch and Go, ObjectFolder-1.0, and YCB-Slides dataset, as shown in Fig. 9. Each curve shows how classification accuracy changes as the bitrate varies, with dotted lines indicating the performance on uncompressed data (24 bpp). These plots provide a more comprehensive view of semantic preservation across different compression levels.

Specifically, the bitrate is varied by adjusting the quantization parameter (QP) for each compressor. For each classification task, we split the reconstructed data into 60% for training and 40% for testing, and apply four standard classifiers (SVM, Random Forest, K-NN, and Linear Regression) to perform material classification (TouchandGo and ObjectFolder) or object classification (YCB-Slide). Overall, even at over $200\times$ compression, the impact on classification accuracy remains minor for all methods, suggesting that lossy compression can be applied without substantially compromising downstream understanding tasks. Among them, TaCo-L consistently achieves the highest accuracy across the full bitrate range, and closely approaches the accuracy of raw data (24 bpp).

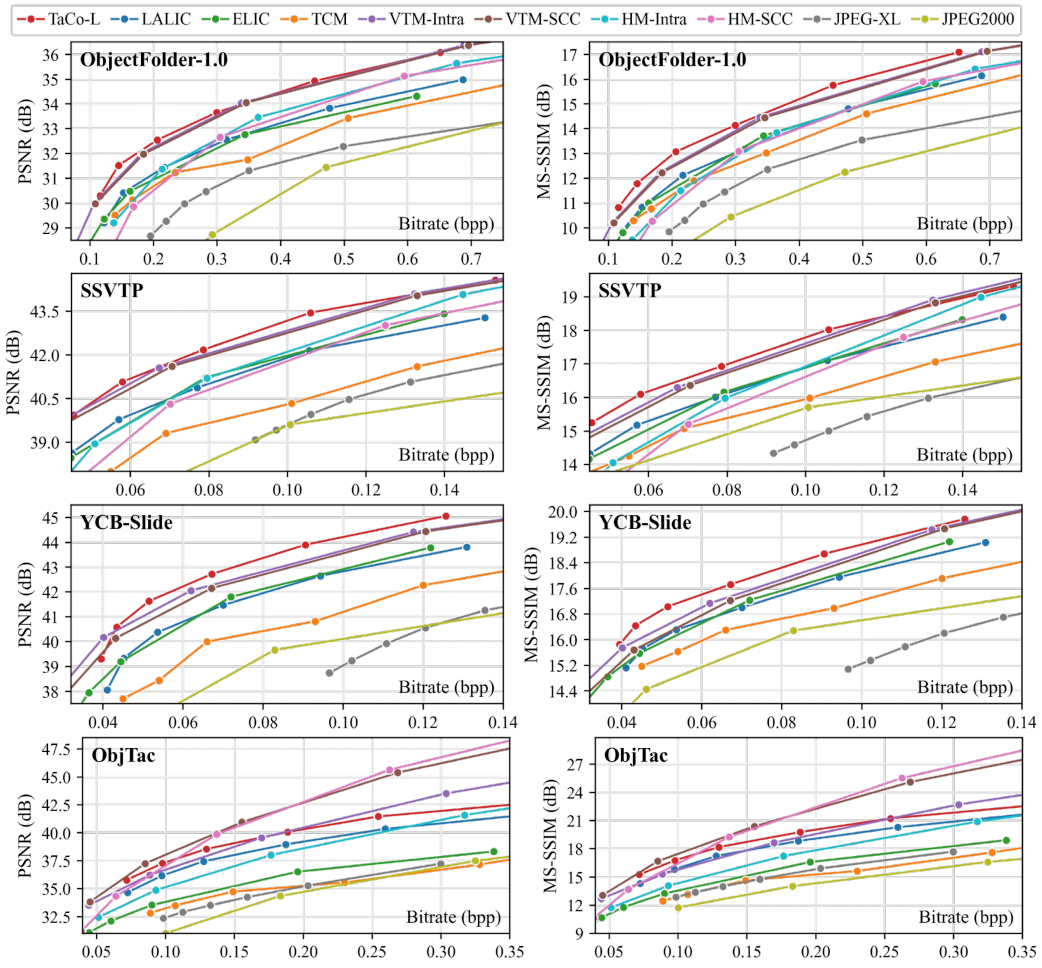


Figure 6: Rate-distortion performance across four tactile datasets when treating tactile data as images.

A.7 More Lossy Compression Results for Dexterous Grasping

We further visualize the simulation environment from the IssacSim and part of the assets, as shown in Fig. 10. The hand model is based on the Paxini DexHand13 Paxini. (2024), which has four fingers and a total of 16 DoF. Each finger except the thumb is equipped with three tactile sensors and the thumb finger is equipped with two tactile sensors, resulting to a total of 11 tactile sensors. We deploy a simple asymmetric actor-critic (AAC) network with the tactile data as the input, to learn the dexterous grasping for general objects (Wang et al., 2025). Although the grasping success rate of our baseline model is not very high, we focus on the impact of tactile compression.

We have conducted grasping experiments in real-world settings and employed four mature encoders (JPEG2000, JPEG XL, BPG, VTM) to compress tactile signals with varying quantization parameters (QP). Using four fingertip positions as primary observation metrics, we present the sensory force data along the x, y, and z axes across these four fingertips, with the results illustrated in Fig. 11.

Fig. 12, Fig. 13,Fig. 14 and Fig. 15 illustrate the visualization results of tactile signals from four fingertips using four different codecs in real-world experiments. Meanwhile, Fig. 16, Fig. 17,Fig. 18 and Fig. 19 illustrate the visualization results of tactile signals from four fingertips using four different codecs in the simulations. These figures demonstrate that the compression algorithm itself does not actually affect the main variation distribution of the tactile data, and therefore will not have a catastrophic impact on the accuracy of real-world tasks.

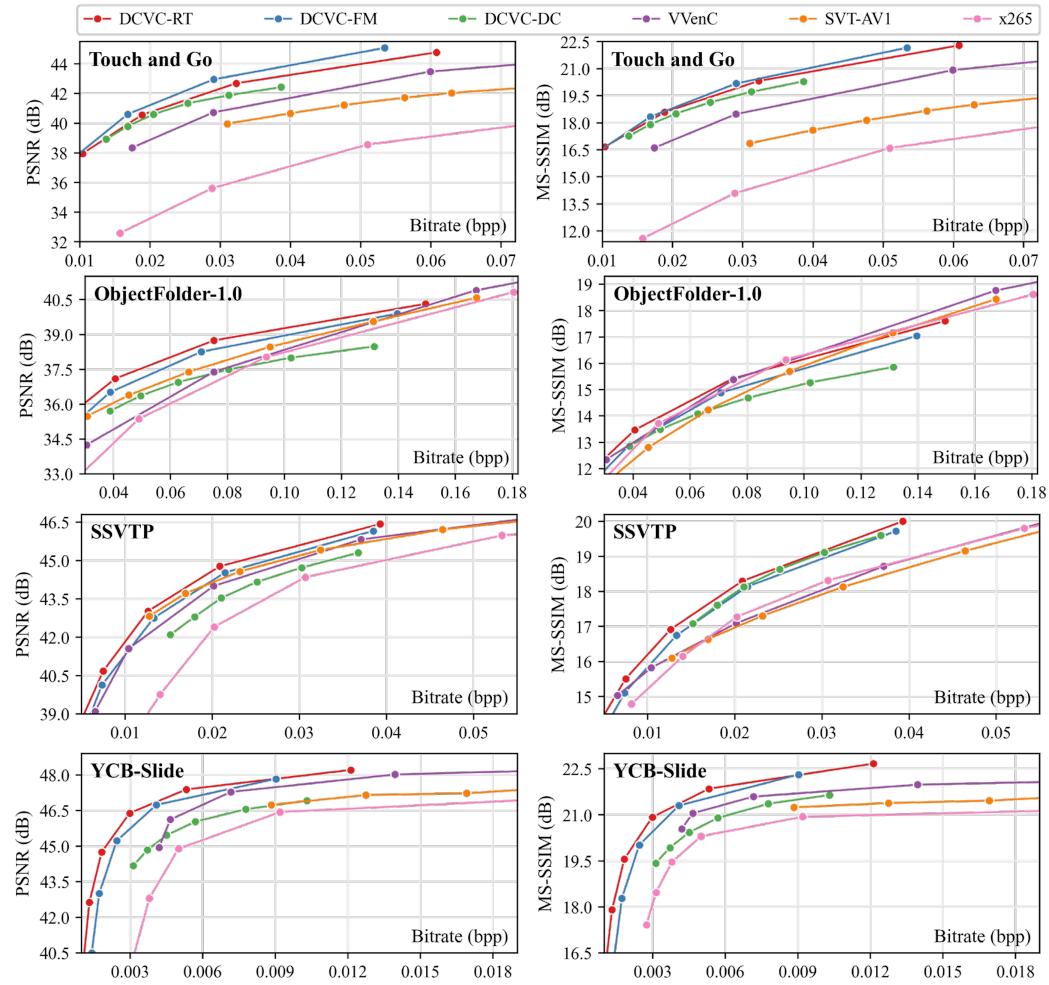


Figure 7: Rate-distortion performance across four tactile datasets when treating tactile data as videos.

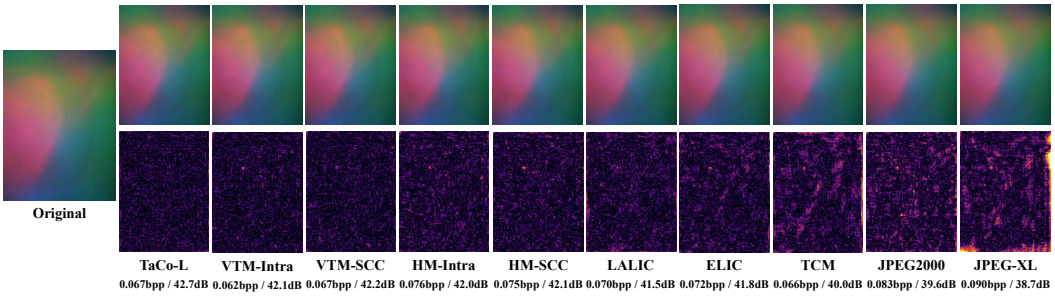


Figure 8: Visualization of reconstructed tactile images (top row) and their corresponding per-pixel root mean squared error (RMSE) maps (bottom row) on the YCB-Slide dataset. The RMSE maps highlight local reconstruction errors, with brighter regions indicating larger residuals.

Regarding the implementation details, the simulation environment for the reinforcement learning controller operates at a control frequency of 100 Hz, which is determined by the simulation time step of 0.01 seconds. Specifically, (1) the tactile sensors are updated at every simulation step, resulting in a tactile sampling rate of 100 Hz. (2) The overall latency of the control loop is approximately 0.01 seconds, plus the time required for policy inference. The policy inference is performed using an ONNX model, and the inference time is logged during execution. If the inference time exceeds

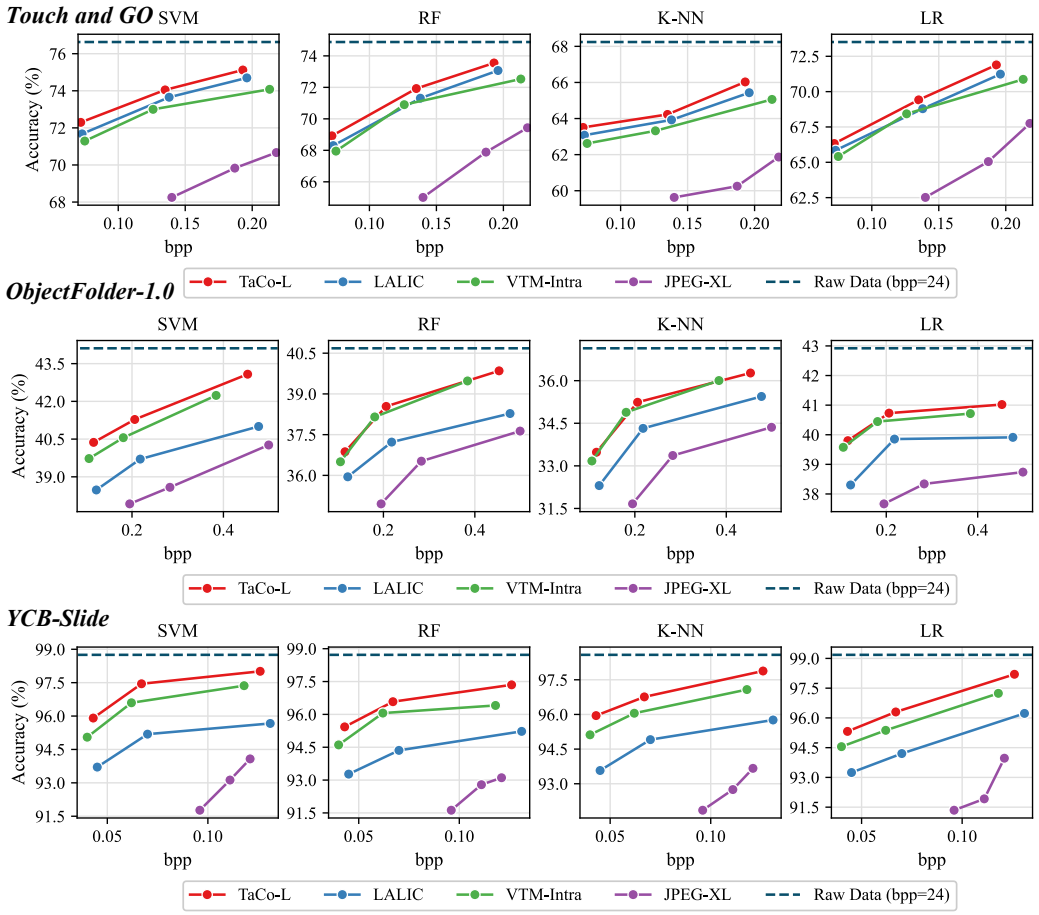


Figure 9: Bpp-accuracy curves for material classification task on the TouchandGo and ObjectFolder-1.0 datasets, and object classification task on the YCB-Slide dataset.

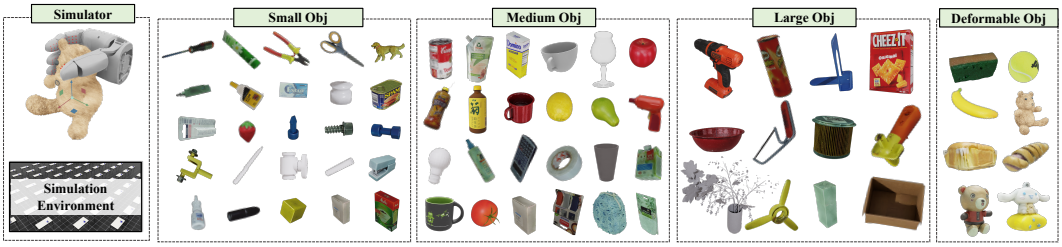
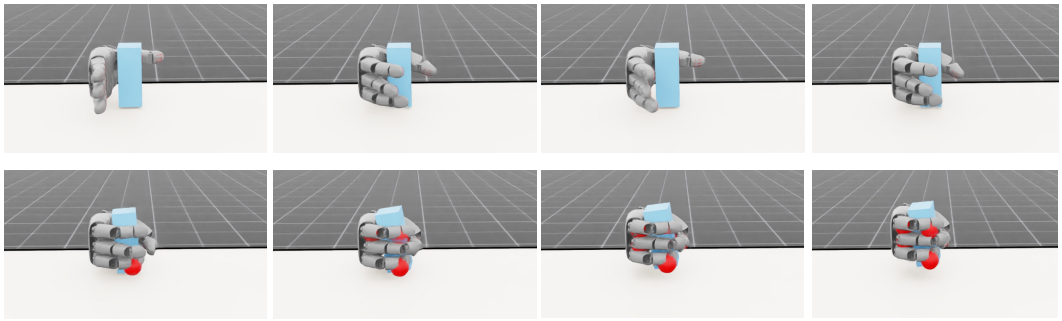


Figure 10: Simulation environment and part of object assets we use in the grasping exeriments.

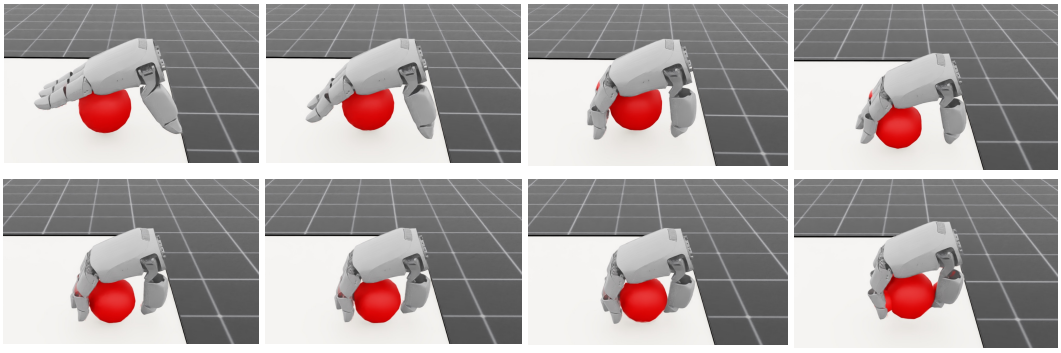
the simulation time step, the control frequency may decrease, and the latency would increase accordingly. (3) When the combined codec and inference latency approximately equals the simulation update interval, the additional delay introduced to the simulation environment becomes negligible, as it aligns with the natural timing cycle of the control loop. However, in the current implementation, the control command is applied in the same simulation step after inference, so the latency is primarily determined by the simulation step and the inference time.

A.8 Our Motivation and Future Work

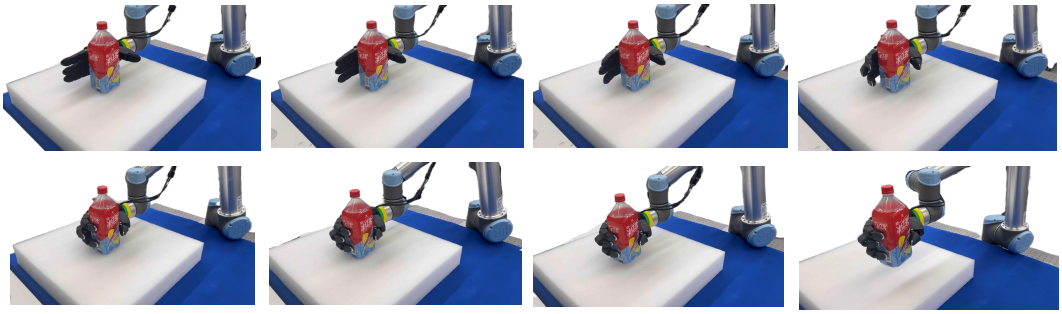
In this section, we simply present the need for advancing tactile compression. The development of this tactile codec benchmark is motivated by three critical challenges in practical robotics appli-



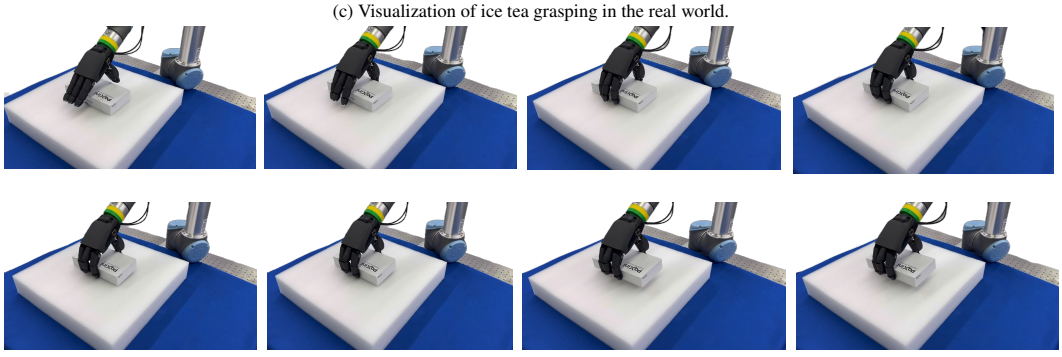
(a) Visualization of cube grasping in the simulation.



(b) Visualization of ball grasping in the simulation.



(c) Visualization of ice tea grasping in the real world.



(d) Visualization of box grasping in the real world.

Figure 11: Visualization of grasping sequences in the simulation and real-world experiments.

cations. First, for dexterous manipulation, tactile data from high-resolution sensor arrays on robot hands can consume a significant portion of the available bandwidth. This is analogous to the Cortical Homunculus, where the hands claim a disproportionately large share of neural resources. The limited bandwidth of low-cost microcontrollers (MCUs) embedded in such hands creates a fundamental bottleneck for real-time sensorimotor control.

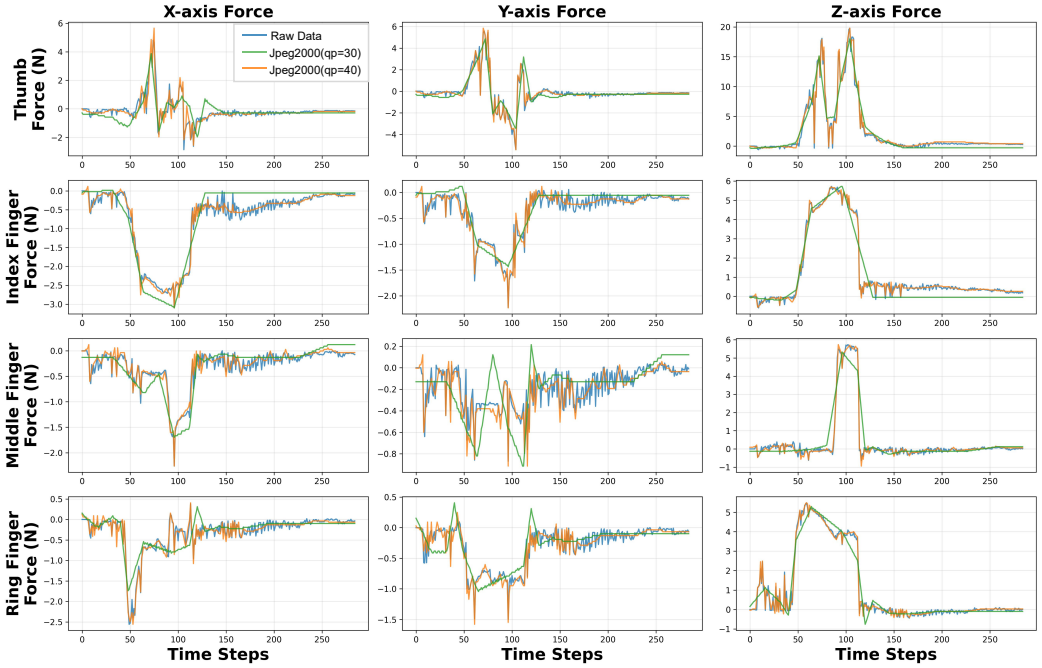


Figure 12: Visualization of tactile signals in the real-world experiments with JPEG2000 as the codec.

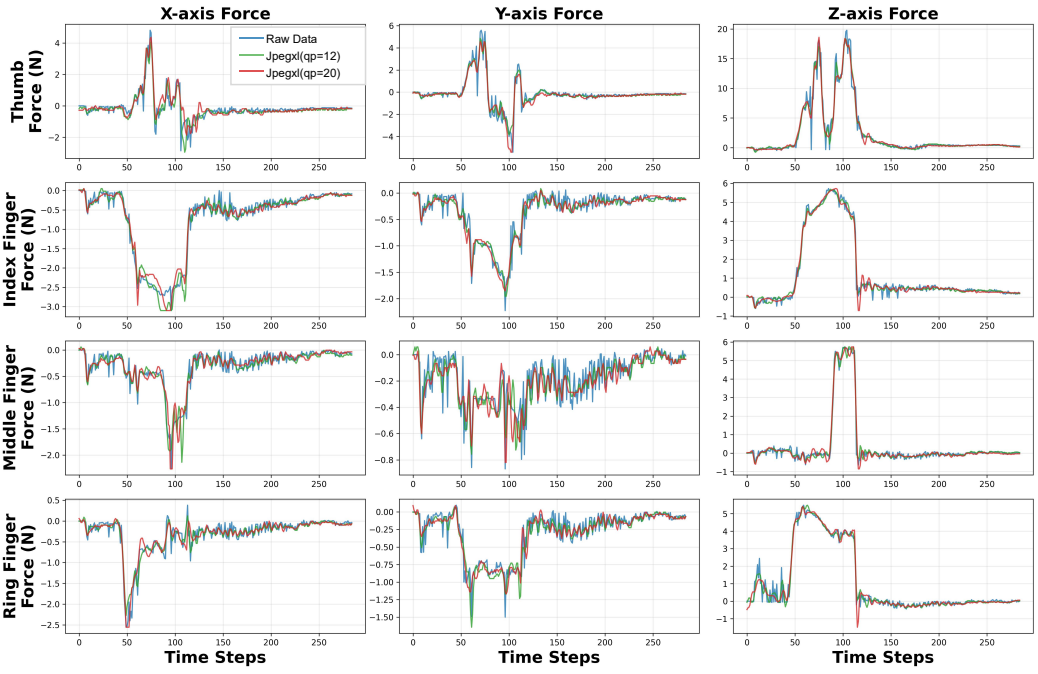


Figure 13: Visualization of tactile signals in the real-world experiments with JPEG-XL as the codec.

Second, in robotic tele-operation systems, achieving stable and transparent remote control requires low-latency, high-fidelity transmission of tactile signals. Effective compression is paramount to close the feedback loop for delicate tasks, enabling true physical understanding and interaction at a distance.

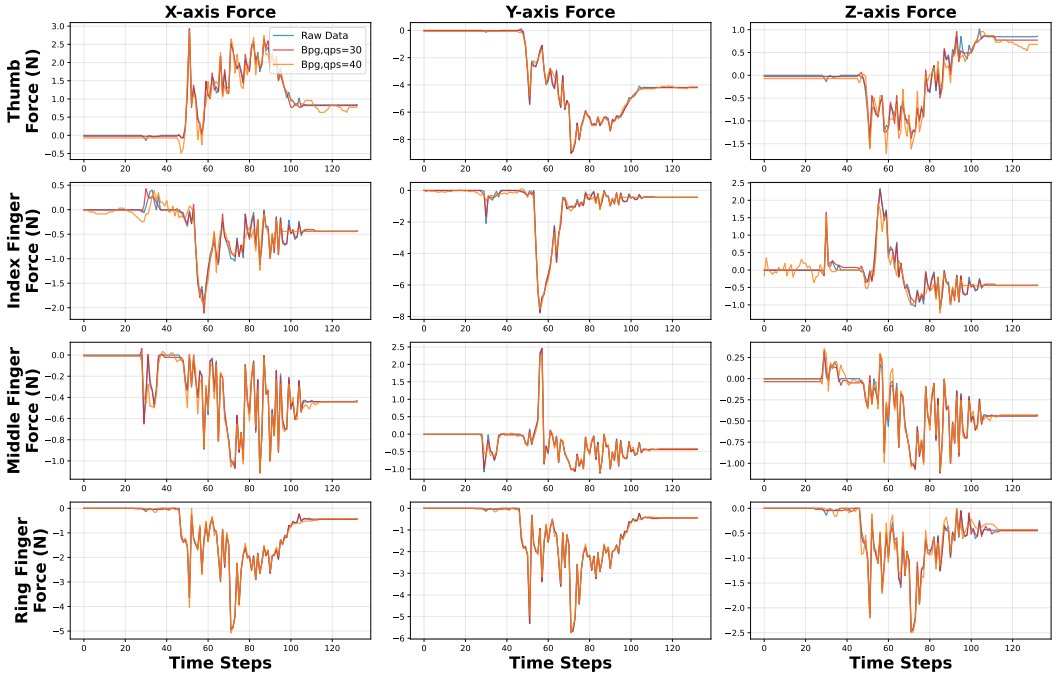


Figure 14: Visualization of tactile signals in the real-world experiments with BPG as the codec.

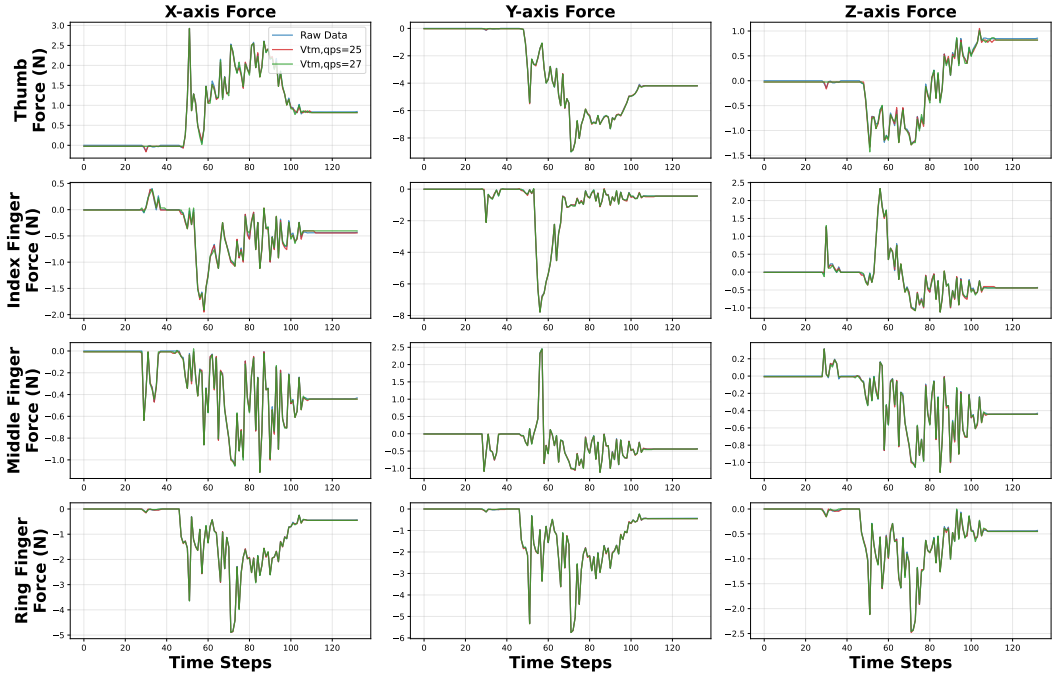


Figure 15: Visualization of tactile signals in the real-world experiments with VTM as the codec.

Third, to scale up in the field of embodied intelligence, the creation of large-scale training datasets necessitates efficient storage solutions. Specifically, Google introduced Open X-Embodiment Dataset, the largest open-source real robot dataset to date. It contains 1M+ real robot trajectories (download size is **8965 GB**) O'Neill & Rehman (2024). While video compression is mature, specialized algorithms for tactile data remain underdeveloped, hindering our ability to build and manage

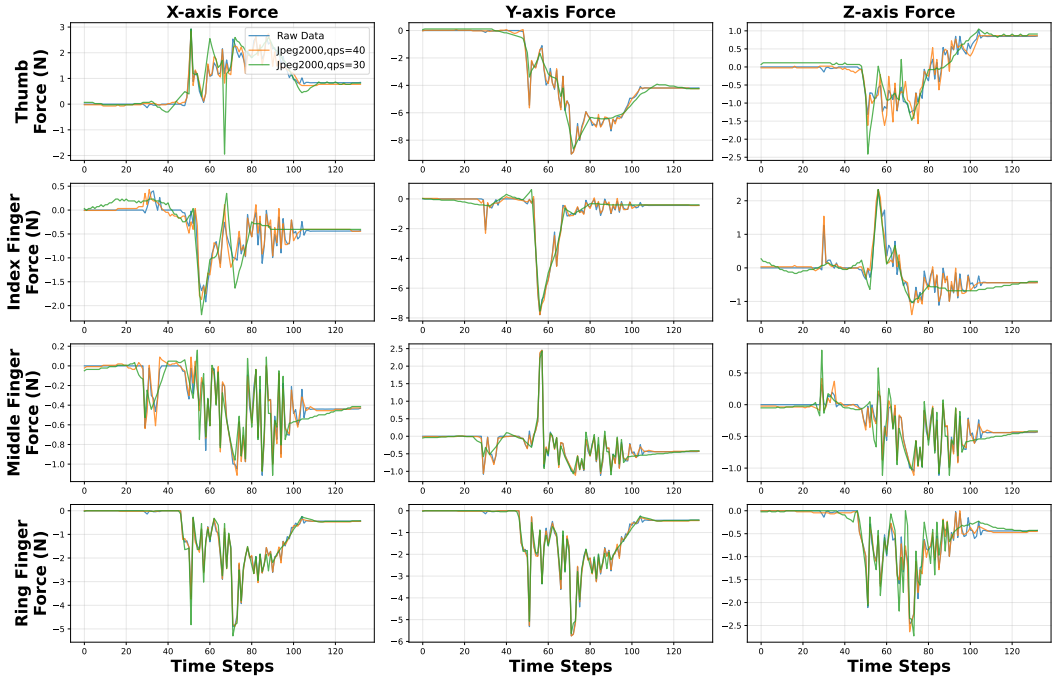


Figure 16: Visualization of tactile signals in the simulation experiments with JPEG2000 as the codec.

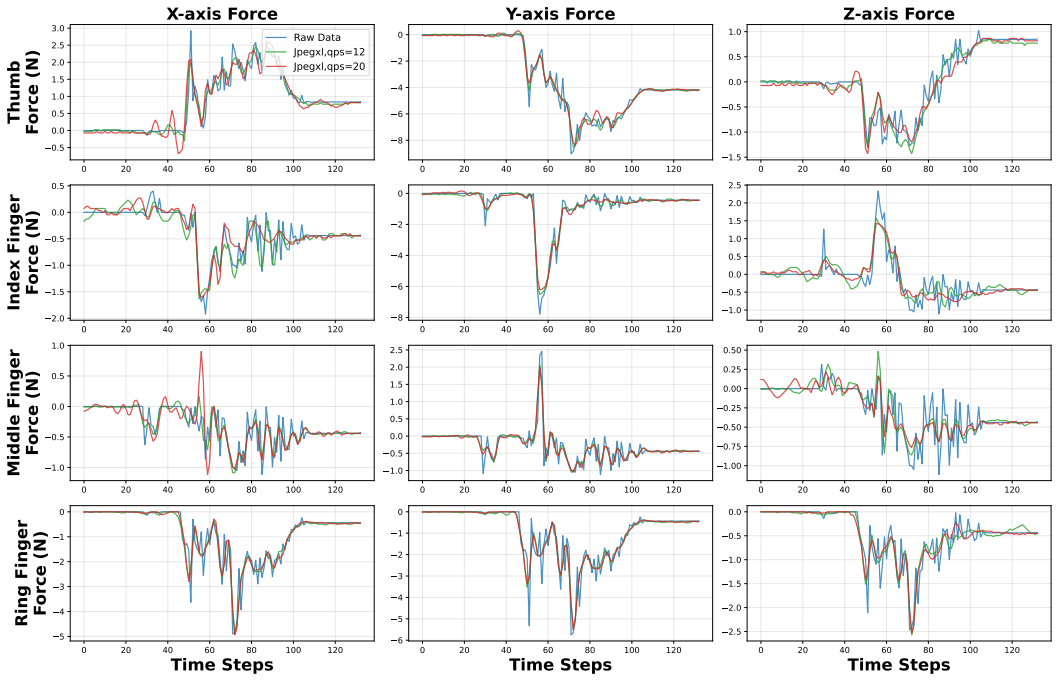


Figure 17: Visualization of tactile signals in the simulation experiments with JPEG-XL as the codec.

the vast datasets required for training generalist robotic models. These pressing needs collectively motivate the establishment of a rigorous benchmark to advance the field of tactile data compression.

For the future work, we will develop a video-like tactile codec by retraining the tactile dataset using the latest neural video compression models, like DCVC-serier models.

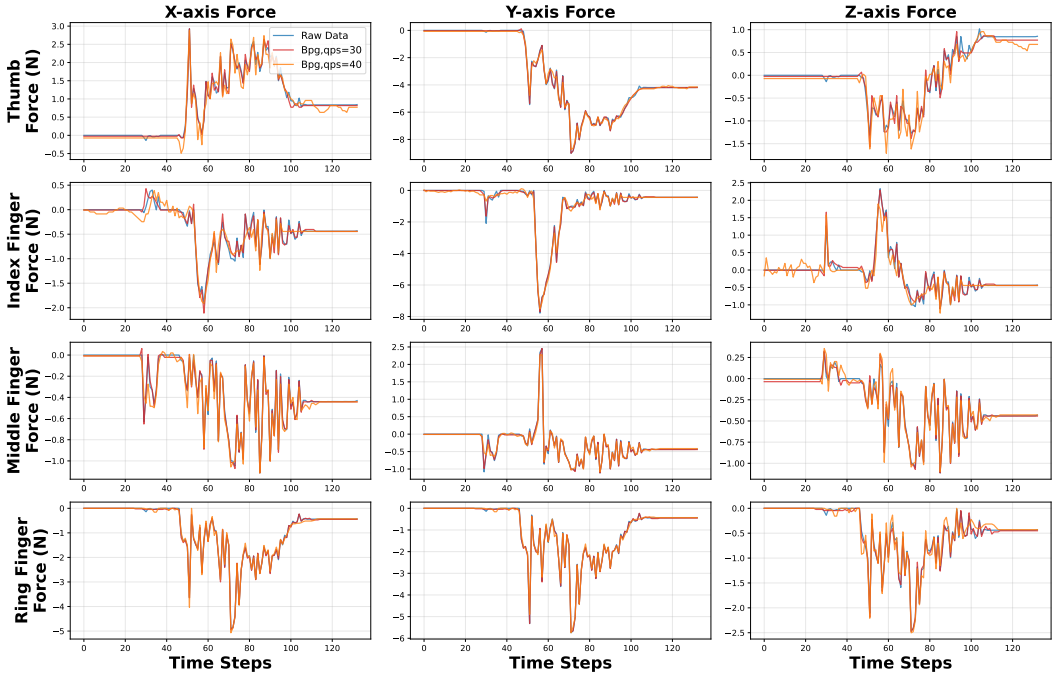


Figure 18: Visualization of tactile signals in the simulation experiments with BPG as the codec.

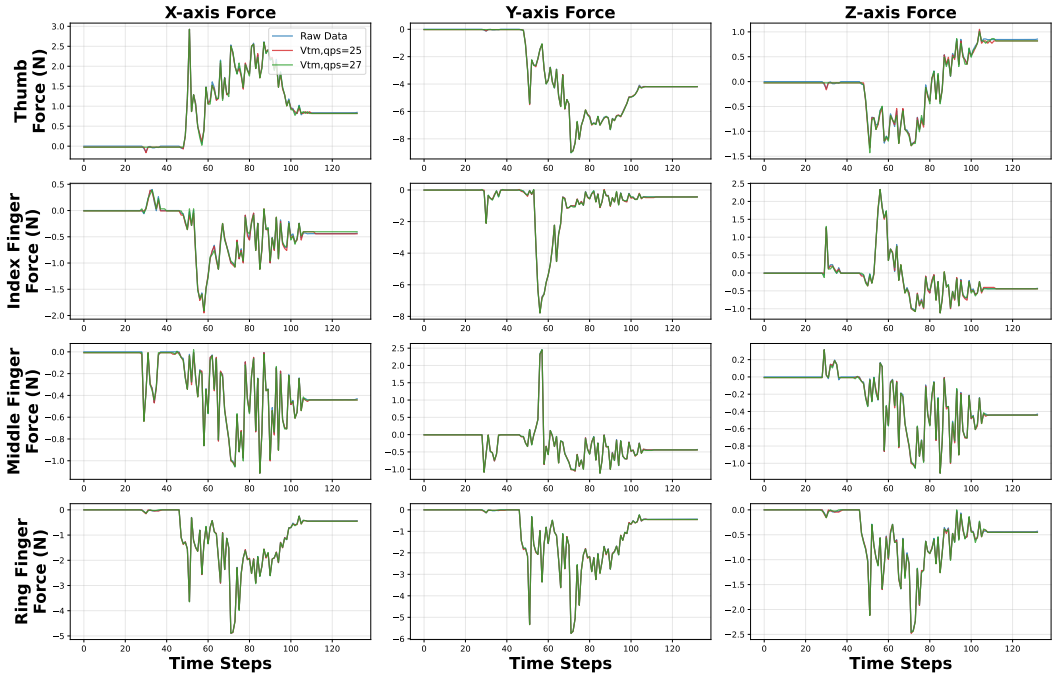


Figure 19: Visualization of tactile signals in the simulation experiments with VTM as the codec.

A.9 LLM USAGE STATEMENT

Large Language Models (LLMs) were not used during the research, experimentation, or analysis phases of this work. During the manuscript preparation, LLMs were used solely for minor grammar and language refinements. No content, ideas, or technical writing was generated by LLMs.