PACE-HAR: Partitioned Adapter-Based Capacity Extension for Interference-Free Cross-Dataset Wearable Models

Vishal Banwari[1,2], Daniel Geißler[1,2], Paul Lukowicz[1,2], and Bo Zhou[1,2]

[1] German Research Center for Artificial Intelligence (DFKI), Kaiserslautern, Germany
[2] RPTU Kaiserslautern, Kaiserslautern, Germany
{vishal.banwari,daniel.geissler,bo.zhou,paul.lukowicz}@dfki.de

Abstract. Wearable sensing models are rarely static. A model trained for one device, placement, or sensor layout may later need to absorb new sensing configurations, from wrist-only accelerometry to multi-inertial-measurement-unit (IMU) or physiological setups. This creates a capacity-extension problem: new sensor domains must be added without disturbing what the model already represents. Fine-tuning risks interference, while training one model per domain duplicates capacity and complicates deployment. We introduce PACE-HAR (**P**artitioned **A**dapter-based **C**apacity **E**xtension), a partitioned adapter-based Vector-Quantized Variational Autoencoder (VQ-VAE) for interference-free cross-dataset capacity extension. PACE-HAR freezes the shared encoder–decoder path and extends it with domain-specific codebook partitions and lightweight input adapters, isolating discrete latent capacity while aligning heterogeneous sensor inputs to the frozen representation space. Across five heterogeneous HAR datasets, a shared codebook shows clear cross-domain interference, while partitioning alone preserves earlier domains exactly but remains sensitive to input shift. PACE-HAR combines both mechanisms: previous-domain reconstruction is preserved exactly (Domain Preservation Ratio, DPR = 1.00) at every extension step, and mean reconstruction error (MSE) drops by 8.28× relative to partitioning alone, at a cost of only 166 trainable non-codebook parameters per added domain. This preservation guarantee, not classifier accuracy, is the central claim of the paper. Downstream classification probes indicate that the representation retains some activity-relevant information, though probe accuracy remains modest; we report this as a narrow, exploratory check rather than a competitive recognition result, and discuss what limits it directly in the paper.

Keywords: Human Activity Recognition · Cross-Dataset Representation Learning · Vector-Quantized Autoencoders · Capacity Extension.

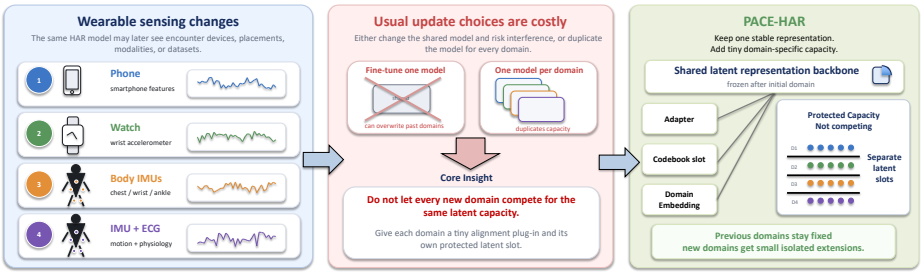

Fig. 1. PACE-HAR extends HAR representation models across sensor domains by freezing a shared VQ-VAE backbone and adding lightweight domain-specific adapters and codebook partitions.

1 Introduction

In deployed wearable sensing pipelines, the sensing configuration often becomes a moving design constraint rather than a fixed training assumption [7, 15]. A phone-based model may be extended to a smartwatch; a wrist-only setup may add ankle or chest IMUs; a motion-only system may incorporate ECG or other physiological channels [24, 25, 3]. Such changes alter the input geometry of the underlying model, including channel structure, signal statistics, and temporal patterns. Long-lived wearable systems therefore need more than a single well-trained model: they need a controlled way to grow capacity across changing sensing setups without the cost of retraining or the overhead of maintaining separate models per configuration.

Common update strategies handle this growth poorly. Fine-tuning a shared model gives the new domain full access to the representation, but can overwrite structure used by previous domains [11, 14]. Maintaining one model per dataset avoids this interference entirely, but duplicates capacity and complicates deployment, especially for compact wearable systems [13, 15]. We study an alternative that targets this growth cost directly, rather than aiming to replace existing recognition pipelines outright: *domain-aware cross-dataset capacity extension*, where a known new sensor domain is added through isolated components while the shared representation path and previously allocated domain capacity remain exactly fixed. Our primary evidence for this is reconstruction preservation; we treat downstream activity recognition as a secondary probe of representation utility rather than as the paper's main performance claim, for reasons we return to in Section 5.4 and Section 7.

We introduce PACE-HAR, a partitioned adapter-based capacity extension method for wearable sensor representation learning. PACE-HAR builds on a VQ-VAE, whose discrete codebook provides an explicit unit of latent capacity [20]. Each sensor domain receives a disjoint codebook partition and a lightweight domain embedding, while the shared encoder–decoder backbone is frozen after the initial domain. Because isolated latent capacity alone cannot compensate for input-side mismatch, PACE-HAR also adds a lightweight do-

main adapter that maps each sensor stream into the representation space expected by the frozen backbone.

We evaluate PACE-HAR across five heterogeneous HAR domains: UCI-HAR, WISDM, PAMAP2, OPPORTUNITY, and MHEALTH [2, 12, 24, 25, 3]. The evaluation separates three questions: whether shared codebooks cause cross-domain interference, whether partitioning alone preserves earlier domains but fails under sensor-input mismatch, and whether adapters with partitions can preserve previous domains while reconstructing newly added ones. Downstream probes further test whether the learned representations retain activity-relevant information. This paper makes four contributions:

- We formulate *cross-dataset capacity extension* for domain-aware, interference-free growth across heterogeneous wearable-sensor datasets.
- We introduce PACE-HAR, a partitioned adapter-based VQ-VAE with frozen shared reconstruction, domain-specific codebook partitions, and lightweight input adapters, giving an exact (DPR = 1.00) preservation guarantee at only 166 trainable non-codebook parameters per added domain.
- We show that partitions and adapters target complementary failures: latent interference and sensor mismatch, and that both are necessary for low reconstruction error under a frozen backbone.
- We evaluate preservation, reconstruction quality, domain-order sensitivity, and parameter efficiency across five HAR domains, and report a narrow downstream classification check as an exploratory, secondary signal of representation utility, together with an explicit discussion of what the common input canvas does and does not let us conclude from it.

2 Related Work

2.1 Cross-Domain and Long-Lived HAR Models

Sensor-based HAR is strongly affected by domain mismatch: models trained on one sensing setup often degrade when users, devices, placements, sampling regimes, or modalities change. Cross-dataset HAR has addressed this largely through transfer and adaptation to a target domain. Qin et al. [21] adapt spatial-temporal representations across domains, while Lu et al. [16] use semantic-discriminative mixup for domain generalization. Adversarial alignment has been explored for cross-domain activity recognition [27], and transformer-based adversarial learning with self-knowledge distillation targets similar cross-device shifts [28]. Dhekane and Ploetz [5] survey transfer learning approaches broadly across HAR, and Chakma et al. [4] study domain adaptation specifically for sensor-based recognition. These works establish heterogeneous HAR as a central deployment challenge, but usually focus on adapting or generalizing to a target domain rather than preserving a previously allocated representation model.

Long-lived HAR systems have also been studied through incremental and continual learning. Gudur et al. [6] propose on-device active learning for HAR, Jha et al. [9] benchmark continual learning methods empirically on activity data,

Ye et al. [32] use generative adversarial replay for continual activity recognition, and Adaimi and Thomaz [1] propose prototypical networks for lifelong adaptation. Their primary objective is typically recognition performance under changing activity data. In contrast, PACE-HAR treats extension as a reconstruction-compatible representation problem: new sensor domains are added while the shared encoder–decoder path and earlier domain capacity remain fixed.

2.2 Capacity Allocation and Discrete Latent Representations

Avoiding interference in sequential learning often requires capacity separation. Rusu et al. [26] introduce progressive networks that add a new column per task, and Yoon et al. [33] dynamically expand network capacity as needed. Residual adapters [23] and the parameter-efficient adapters of Houlsby et al. [8] instead insert small trainable modules into an otherwise fixed backbone. PackNet [19] and Piggyback [18] isolate capacity through iterative pruning and learned binary masks, respectively, while L2P [31] and DualPrompt [30] isolate capacity in a learned pool of prompts. These approaches isolate capacity in weights, heads, masks, or prompt pools, but they do not directly assign explicit latent capacity to sensor domains.

Vector-quantized representation learning provides such explicit latent units through discrete codebook entries [20], with hierarchical and higher-fidelity variants proposed since [22, 29]. Closer to continual learning, Malepathirana et al. [17] use vector quantization for neighborhood-aware prototype preservation, and Jiao et al. [10] use discrete VQ-based prompt selection, but both target discriminative or class-incremental settings rather than reconstruction. PACE-HAR instead couples a partitioned VQ codebook with lightweight input adapters for heterogeneous HAR reconstruction: adapters handle sensor-input mismatch, while private codebook partitions prevent latent-capacity interference.

3 Cross-Domain Capacity Extension

We consider a HAR representation model that must grow beyond the sensor domain it was first built for. A *domain* denotes a dataset-level sensing configuration: device type, body placement, sensor channels, preprocessing, sampling regime, and activity taxonomy. We assume that domain identities are known and can be used for routing, matching deliberate integration scenarios where a new device, placement, or dataset is added to an existing HAR system [15, 13]. The goal is therefore not domain-agnostic inference, but domain-aware extension without retraining the shared backbone or modifying components assigned to earlier domains.

Let $\mathcal{D}_1, \ldots, \mathcal{D}_T$ denote a sequence of HAR domains, where $\mathcal{D}_t = \{(x_i^{(t)}, y_i^{(t)})\}_{i=1}^{N_t}$ contains sensor windows and activity labels. The extension model is trained through reconstruction; labels are used only for downstream evaluation. After adding domain $\mathcal{D}_t$, the model is denoted by Θ_t.

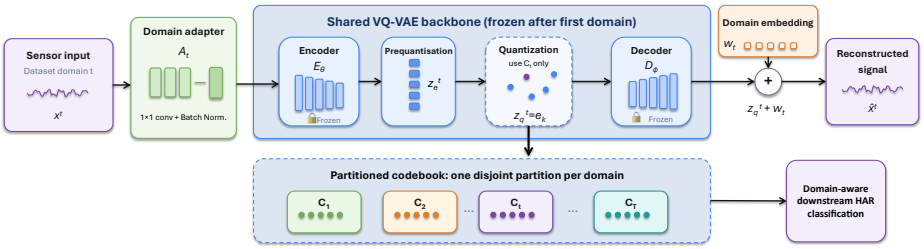

Fig. 2. PACE-HAR architecture. For a new sensor domain $\mathcal{D}_t$, only the domain adapter A_t, codebook partition C_t, and domain embedding w_t are trained. The shared VQ-VAE backbone remains frozen, while quantization is restricted to the active partition to isolate latent capacity across domains.

For domain $\mathcal{D}_i$ after extension step t, reconstruction error is

$$\mathrm{MSE}_i(\Theta_t) = \frac{1}{N_i} \sum_{x \in \mathcal{D}_i} \|x - \hat{x}_{\Theta_t}\|_2^2 . \tag{1}$$

We quantify preservation with the *Domain Preservation Ratio*:

$$\mathrm{DPR}_i^{(t)} = \frac{\mathrm{MSE}_i(\Theta_t)}{\mathrm{MSE}_i(\Theta_i)}, \quad t > i. \tag{2}$$

A value of DPR = 1.00 means that adding later domains leaves the reconstruction error of domain $\mathcal{D}_i$ unchanged; values above 1.00 indicate degradation. Downstream HAR classifiers are trained on the learned representations only to assess recognition utility.

4 PACE-HAR

We introduce PACE-HAR, a capacity-extension architecture for cross-dataset HAR representation learning, shown in Fig. 2. PACE-HAR freezes a shared VQ-VAE backbone after the first domain and assigns each later domain three trainable resources: a lightweight input adapter A_t, a private codebook partition C_t, and a domain embedding w_t. The adapter controls how a new sensor stream enters the frozen model, the partition controls where its latent patterns are stored, and the embedding provides a small domain-specific decoder shift.

4.1 VQ-VAE Backbone for Sensor Reconstruction

The shared backbone of PACE-HAR is a Vector-Quantized Variational Autoencoder (VQ-VAE) trained to reconstruct sensor windows [20]. Because later HAR domains may differ in channel structure, placement, and signal statistics, PACE-HAR does not feed each domain directly into the frozen encoder. Instead, each domain first passes through a lightweight domain adapter A_t, which

maps the input window into the representation space expected by the shared backbone.

For an input window $x^{(t)}$ from domain $\mathcal{D}_t$, the adapter produces an aligned input

$$\tilde{x}^{(t)} = A_t(x^{(t)}), \tag{3}$$

and the shared encoder maps this adapted signal to a continuous latent representation,

$$z_e^{(t)} = E_\theta(\tilde{x}^{(t)}). \tag{4}$$

A pre-quantisation layer maps this representation into the codebook space, where it is replaced by its nearest discrete codeword:

$$z_q^{(t)} = e_k, \quad k = \arg\min_j \left\| z_e^{(t)} - e_j \right\|_2. \tag{5}$$

The decoder reconstructs the sensor window from the quantised latent:

$$\hat{x}^{(t)} = D_\phi(z_q^{(t)}). \tag{6}$$

The reconstruction objective combines mean squared error with a standard VQ commitment term:

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N} \|x_i - \hat{x}_i\|_2^2 + \beta \frac{1}{N}\sum_{i=1}^{N} \|z_{e,i} - \mathrm{sg}[z_{q,i}]\|_2^2, \tag{7}$$

where $\mathrm{sg}[\cdot]$ denotes stop-gradient. We use $\beta = 0.25$ and update the codebook with exponential moving average, following common VQ-VAE practice [20, 22]. The key design choice in PACE-HAR is not the VQ-VAE backbone itself, but how its discrete codebook is allocated during extension.

4.2 Codebook Partitions as Isolated Latent Capacity

A shared codebook makes all domains compete for the same discrete latent capacity. With a new sensor domain, its updates can move codewords that earlier domains already depend on. PACE-HAR avoids this by partitioning the codebook into disjoint domain-specific subsets:

$$C = \bigcup_{t=1}^{T} C_t, \quad C_i \cap C_j = \emptyset \quad \text{for } i \neq j. \tag{8}$$

During training and inference for domain $\mathcal{D}_t$, quantisation is restricted to its partition C_t:

$$z_q^{(t)} = e_k, \quad k = \arg\min_{e_j \in C_t} \left\| z_e^{(t)} - e_j \right\|_2. \tag{9}$$

Thus, one domain never uses or updates codewords assigned to another domain. In our five-domain HAR setting, the codebook contains 192 entries with 32-dimensional code vectors, and each domain receives 32 codewords. The total codebook is pre-allocated, so extension does not require growing the shared backbone.

4.3 Freezing the Shared Representation Path

Partitioning the codebook protects earlier domains only if the surrounding representation path remains stable. If the encoder changes, earlier inputs may no longer map near their assigned codewords; if the decoder changes, the same codeword may reconstruct differently. PACE-HAR therefore freezes the encoder, pre-quantisation layer, and decoder after the first domain.

For the initial domain $\mathcal{D}_1$, the full VQ-VAE is trained. For each later domain $\mathcal{D}_t$, $t > 1$, the shared backbone remains fixed and training is restricted to the current domain's adapter A_t, codebook partition C_t, and embedding w_t. The embedding is added before decoding:

$$\hat{x}^{(t)} = D_\phi \left(z_q^{(t)} + w_t \right), \quad w_t \in \mathbb{R}^{32}. \tag{10}$$

This gives each domain a minimal amount of decoding flexibility without updating the decoder itself. After a domain has been added, its adapter, partition, and embedding are frozen as well.

4.4 Domain Adapters for Sensor Alignment

HAR domains differ not only in labels, but in the structure of the input signal itself: channel count, placement, preprocessing, sampling behavior, and sensor statistics. A frozen backbone trained on one domain may therefore be poorly aligned with later domains, even when those domains receive private codebook capacity. In PACE-HAR, the domain adapter is the interface between heterogeneous sensor inputs and the frozen representation path.

For each domain, a lightweight adapter transforms the input before it reaches the frozen encoder:

$$\tilde{x}^{(t)} = A_t(x^{(t)}), \quad z_e^{(t)} = E_\theta(\tilde{x}^{(t)}). \tag{11}$$

We implement A_t as a lightweight 1×1 convolution followed by batch normalisation, following the broader idea of parameter-efficient domain adaptation with lightweight trainable modules [23, 8]. It performs a channel-wise transformation that maps the current sensor stream into the representation space expected by the frozen backbone. The adapter is trained only together with the current partition C_t and embedding w_t, then frozen. Each adapter adds 134 parameters; together with the 32-dimensional domain embedding, each new domain adds only 166 trainable non-codebook parameters.

4.5 Extension Procedure

The first domain trains the full VQ-VAE. For each later domain $\mathcal{D}_t$, PACE-HAR allocates a fresh adapter A_t, codebook partition C_t, and embedding w_t, trains only these components, and then freezes them. Thus, extension requires neither replay, full retraining, nor a separate model per dataset.

Table 1. HAR domains used for capacity extension, spanning feature vectors, single-device IMU streams, multi-IMU setups, and physiological sensing.

Dataset	Act.	Subj.	Input	Sensor setup
UCI-HAR [2]	6	30	561	Smartphone IMU features
WISDM [12]	6	36	3×128	Phone/Watch IMU
PAMAP2 [24]	12	9	27×128	Chest/wrist/ankle IMU
OPPORTUNITY [25]	17	4	113×128	Full-body IMU
MHEALTH [3]	12	10	23×128	IMU + ECG

5 Experimental Setup

We evaluate PACE-HAR in two stages. First, we study reconstruction-based capacity extension, where the model must add heterogeneous HAR domains while preserving previously supported ones. Second, we run a narrow downstream classification check to see whether the learned representations retain any activity-relevant information at all.

5.1 Datasets and Domain Chains

We use five HAR datasets with different sensor configurations, input formats, and activity distributions, summarized in Table 1: UCI-HAR [2], WISDM [12], PAMAP2 [24], OPPORTUNITY [25], and MHEALTH [3]. Together, they span compact smartphone-derived features, wrist accelerometry, multi-sensor body-worn IMU streams, high-dimensional full-body sensing, and motion signals combined with physiological channels.

We evaluate reconstruction on three domain chains, labeled *easy*, *medium*, and *hard* according to their expected initialization and shift conditions. The easy chain starts from the rich multi-sensor PAMAP2 domain; the medium chain follows a canonical progression from compact smartphone features to richer sensor streams; the hard chain starts from narrow WISDM wrist accelerometry and adds larger shifts later.

Easy: PAMAP2 → MHEALTH → WISDM → OPPORTUNITY → UCI-HAR

Medium: UCI-HAR → WISDM → PAMAP2 → OPPORTUNITY → MHEALTH

Hard: WISDM → UCI-HAR → MHEALTH → PAMAP2 → OPPORTUNITY

5.2 Reconstruction Variants

For reconstruction, we compare three variants that isolate the main design choices, summarized in Table 2. *Shared VQ-VAE* uses one shared codebook that is updated across domains, testing whether shared latent capacity is sufficient under heterogeneous sensor shifts. *Partitioned VQ-VAE* assigns each domain a private codebook partition and domain embedding while freezing the shared

Table 2. Ablation variants separating shared-capacity interference, latent isolation, and input alignment. Parameter counts exclude codebook entries.

Variant	Backbone	Codebook	Input align.	Non-CB params
Shared VQ-VAE	updated	shared	no	all params
Partitioned VQ-VAE	frozen	private C_t	no	32
PACE-HAR	frozen	private C_t	yes	166

Table 3. The two classification variants compared in Section 6.3.

Variant	Uses VQ repr.	Retention strategy	Added capacity
Freeze + Probe	Yes	Frozen features	Small head
PACE-HAR+Probe	Yes	Partitioned features	Small head

backbone, testing whether isolated latent capacity alone preserves earlier domains. *PACE-HAR* adds a domain adapter before the frozen backbone, testing whether input alignment is necessary in addition to isolated latent capacity.

We report reconstruction error using MSE and preservation using DPR, as defined in Section 3. We also summarize adapter gains and order sensitivity across the easy, medium, and hard chains.

5.3 Classification Baseline

Beyond reconstruction, we run a lightweight check of whether the learned representation supports downstream activity recognition at all. This check is auxiliary: reconstruction remains the primary capacity-extension objective, and we keep the comparison deliberately narrow rather than benchmarking broadly against the continual-learning literature (see Section 7). We use a *probe* to mean a lightweight supervised classifier trained on top of a learned representation, while the reconstruction backbone is kept fixed. Thus, probe accuracy measures how much activity-relevant information is retained in the representation rather than how well the full model can be optimized for classification.

We compare two variants, summarized in Table 3. *Freeze + Probe* is the simplest reconstruction-compatible baseline: it trains a lightweight classifier on frozen reconstruction features from the shared (non-partitioned) backbone. *PACE-HAR + Probe* is our reconstruction-compatible variant: the probe is trained on the partitioned PACE-HAR representation while the reconstruction path remains preserved. We report final mean accuracy, parameter count, accuracy per million parameters, and old-domain retention.

5.4 Implementation Details

All datasets are mapped to a shared reconstruction format. UCI-HAR uses its standard 561-dimensional feature vectors; WISDM, PAMAP2, OPPORTUNITY, and MHEALTH are windowed with sizes/strides of 200/100, 256/128,

Table 4. Reconstruction and preservation across the canonical five-domain sequence. Lower MSE is better; DPR=1.00 indicates no reconstruction degradation.

	Shared		Partitioned		**PACE-HAR**	
Domain	MSE	DPR	MSE	DPR	MSE	DPR
UCI-HAR	0.0304	5.69	0.0021	1.00	**0.0019**	1.00
WISDM	0.0320	2.68	0.0224	1.00	**0.0014**	1.00
PAMAP2	0.0846	4.54	0.0745	1.00	**0.0013**	1.00
OPPORTUNITY	0.0747	3.24	0.0645	1.00	**0.0069**	1.00
MHEALTH	0.0669	–	0.0792	–	**0.0179**	–
Mean	0.0577	–	0.0485	–	**0.0059**	–

128/64, and 128/64 samples, respectively. Unlabeled samples are removed, missing values are imputed per dataset, and each sample is independently normalized, clipped to $[-3, 3]$, rescaled to $[-1, 1]$, padded or truncated to 3072 values, and reshaped to $(3, 32, 32)$. This common canvas enables controlled cross-domain reconstruction, while abstracting away some native temporal and channel structure.

PACE-HAR uses a convolutional VQ-VAE with 128 hidden channels, two downsampling layers, two residual layers, 32 residual hidden channels, and a 32-dimensional code embedding. The codebook is partitioned by domain; each domain also receives a 32-dimensional domain embedding. From the second domain onward, PACE-HAR adds a residual 1×1 input adapter with hidden width 16. Models are trained with Adam, learning rate 3×10^{-4}, batch size 128, and 50 epochs per domain, using reconstruction and commitment losses with EMA codebook updates and $\beta = 0.25$. UCI-HAR uses its official split; all other datasets use an 80/20 split after windowing with seed 42.

For classification probes, frozen reconstruction backbones are followed by linear probes or classifier heads trained for 30 epochs with Adam at 10^{-3}. Parameter counts include the backbone, codebook, domain embeddings, instantiated adapters, and downstream heads; each added PACE-HAR domain contributes 134 adapter parameters and 32 domain-embedding parameters, i.e., 166 trainable non-codebook parameters before probe parameters.

6 Results

We first evaluate reconstruction and preservation across the canonical five-domain sequence, then test sensitivity to domain order. Finally, we run a narrow downstream classification check against a single baseline, while treating reconstruction-compatible capacity extension as the primary objective.

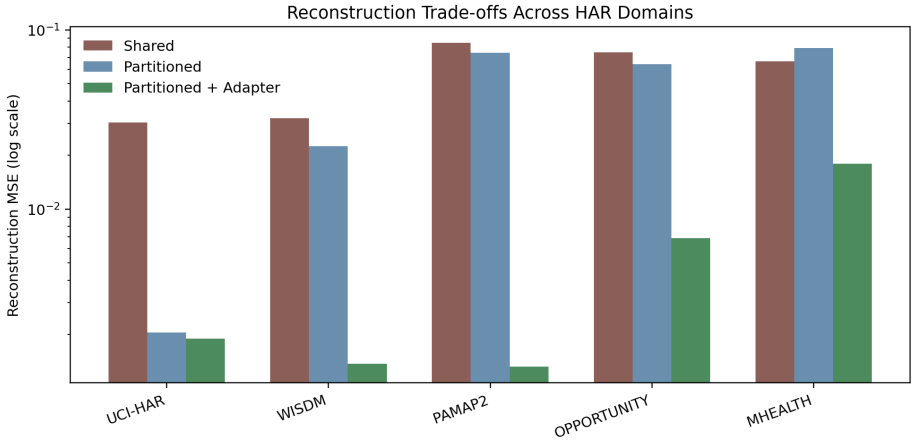

Fig. 3. Reconstruction MSE across domains. PACE-HAR stays in the low-MSE regime, unlike the shared and partitioned-only variants.

6.1 Reconstruction, Preservation, and Component Ablation

Table 4 tests the central trade-off: whether the model can stay stable without becoming too rigid to reconstruct new domains. The shared VQ-VAE suffers from cross-domain interference, with DPR values between 2.68 and 5.69 on earlier domains. Partitioning removes this interference by keeping DPR at 1.00, but reconstruction quality remains weak on several shifted domains. PACE-HAR combines both properties: it preserves all earlier domains with DPR 1.00 and reduces mean reconstruction MSE to 0.0059.

The main improvement comes from adding input alignment to latent isolation. Relative to the partitioned model without adapters, PACE-HAR reduces mean MSE from 0.0485 to 0.0059, an 8.28× improvement. Preservation itself is structurally expected because previously allocated adapters, codebook partitions, embeddings, and the shared backbone are frozen. The non-trivial result is that newly added heterogeneous domains can still be reconstructed well under this frozen-backbone constraint.

Figure 3 visualizes the same pattern across domains. PACE-HAR consistently stays below the shared and partitioned-only variants, showing that isolated codebook capacity alone is not sufficient when the frozen backbone receives heterogeneous sensor inputs. The adapter provides the missing input-side alignment, while the partitioned codebook preserves previously allocated latent capacity.

To characterize this input-side mismatch directly, we measure a per-dataset *shift score*, quantifying how far a domain's native input distribution lies from the domain used to initialize the frozen backbone, and report the mean and standard deviation of the resulting encoder-output gap alongside the adapter's multiplicative MSE improvement over the no-adapter partitioned model (Table 5).

Table 5. Domain shift relative to the initializing domain and the corresponding adapter improvement in reconstruction MSE over the no-adapter partitioned model. Higher shift scores correspond to larger adapter gains.

Dataset	Shift score	Mean gap	Std gap	Adapter improvement
UCI-HAR	0.000	0.000	0.000	1.09×
WISDM	0.211	0.141	0.071	16.37×
PAMAP2	0.401	0.262	0.139	56.60×
OPPORTUNITY	0.383	0.237	0.146	9.40×
MHEALTH	0.404	0.159	0.245	4.43×

Table 6. Order sensitivity of PACE-HAR across domain chains. *Preservation* reports mean / maximum DPR.

Order	Chain	Mean MSE	Preservation
Easy	PAMAP2 → MHEALTH → WISDM → OPPORTUNITY → UCI-HAR	0.006380	1.00 / 1.00
Medium	UCI-HAR → WISDM → PAMAP2 → OPPORTUNITY → MHEALTH	0.005665	1.00 / 1.00
Hard	WISDM → UCI-HAR → MHEALTH → PAMAP2 → OPPORTUNITY	0.006015	1.00 / 1.00

UCI-HAR is the domain used to initialize the frozen backbone in this ordering, so its shift score is zero by construction and the adapter contributes almost nothing (1.09×), which is the expected null case. Every other domain has a non-zero shift score, and the adapter's improvement is substantial across all of them, though the relationship is not strictly monotonic in the shift score alone: PAMAP2 has a similar shift score to MHEALTH and a lower score than OPPORTUNITY, yet shows by far the largest adapter improvement (56.60×). This suggests that raw distributional distance from the initializing domain is only part of the story; the adapter's benefit likely also depends on how the frozen encoder's learned features happen to align with each domain's specific channel structure, which the shift score does not fully capture. We view a more targeted characterization of this relationship as useful future work, but the overall pattern still supports the core claim of this subsection: input alignment matters most precisely where the frozen backbone would otherwise be poorly matched to the incoming domain.

6.2 Order Sensitivity Across Domain Chains

Table 6 tests whether PACE-HAR depends strongly on the domain used to initialize the frozen backbone. Across the three chains, mean MSE remains in a

Table 7. Aggregate classification probe results (mean across the five HAR datasets). This is a narrow, single-baseline check of representation utility, not a benchmark comparison; see Section 7. Retention is reported in %; MPar. denotes million parameters and Acc./M accuracy points per million parameters.

Method	Final Acc.	MPar.	Acc./M	Retention (Old)
Freeze + Probe	24.21	0.860	28.15	92.92
PACE-HAR+Probe	**41.12**	**0.710**	**57.96**	**110.73**

narrow range from 0.005665 to 0.006380. At the same time, preservation remains exact in every order, with mean and maximum DPR both equal to 1.00.

This result is important because the first domain defines the shared encoder–decoder path used by all later domains. Starting from a richer multi-sensor domain could provide a more flexible backbone, while starting from a narrower single-device domain could make later domains harder to align. The small differences across chains suggest that the domain adapter absorbs much of this input-side mismatch, so capacity extension is not overly dependent on a favorable domain initialization order.

6.3 Classification Probes as a Secondary Signal

We use classification here as a diagnostic probe of representation utility rather than as a recognition benchmark. PACE-HAR's primary claim remains the exact reconstruction-preservation guarantee established above; the probe results below are a first, exploratory look at how much activity-relevant information survives the reconstruction bottleneck, and we return to their limits in Section 7. Retention (Old) reports the ratio between final and best observed accuracy on previously learned domains.

Table 7 shows the result of this narrow check. PACE-HAR+Probe reaches 41.12% mean accuracy against 24.21% for the Freeze+Probe baseline, at fewer parameters (0.710M vs. 0.860M) and higher old-domain retention (110.73% vs. 92.92%). We read this as evidence that the partitioned representation retains more activity-relevant structure than an unpartitioned frozen backbone, not as evidence of competitive recognition performance: 41% mean accuracy is well below what a dedicated HAR classifier would achieve, and we do not present it as such. A fuller per-dataset breakdown and comparison against stronger classifier-oriented and continual-learning baselines is left to future work, as discussed in Section 7.

6.4 Qualitative Latent-Space Analysis

Figure 4 provides a qualitative view of the representation learned by PACE-HAR after all domains have been added. The projection shows visibly separated dataset-level regions, suggesting that the partitioned representation preserves

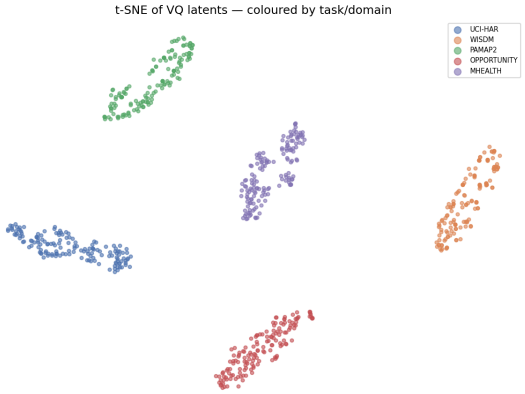

Fig. 4. t-SNE projection of PACE-HAR VQ latents after the full domain sequence, colored by dataset domain. The projection provides a qualitative view of domain-specific structure after sequential extension.

domain-specific structure rather than collapsing heterogeneous sensor inputs into a single mixed latent space. This supports the quantitative reconstruction and retention results, but should be interpreted as a qualitative diagnostic because t-SNE distances and cluster geometry can be affected by the projection.

7 Scope and Limitations

We discuss three limitations directly, since they bound what our results can and cannot support.

The shared reconstruction canvas trades native signal structure for cross-dataset comparability. To reconstruct five datasets with a single frozen backbone, every input is normalized, clipped, padded or truncated to a fixed 3072-value canvas, and reshaped to $(3, 32, 32)$ (Section 5.4). This is what makes a shared encoder–decoder path and a single MSE scale meaningful across UCI-HAR's 561-dimensional engineered features, WISDM's compact 3×128 streams, and OPPORTUNITY's 113×128 full-body IMU signal. It also means that a sizeable fraction of the canvas is padding for low-dimensional datasets and that high-dimensional datasets lose native content to truncation. Low reconstruction MSE should therefore be read as evidence of preservation and low interference *on this shared canvas*, not as direct evidence that all native sensor detail is retained. We view fixing this, for example through per-domain canvas sizes or adapters that project to a shared latent budget without a fixed spatial layout, as the most important direction for follow-up work.

Downstream classification accuracy is not yet at a level that would support practical deployment. Aggregate probe accuracy after five domains is 41.12%, well below dedicated HAR classifiers, and we report it only as a narrow check of representation utility against a single frozen-backbone baseline (Section 6.3),

not as a competitive recognition result or a claim that PACE-HAR is ready to replace a dedicated recognition pipeline. We have not yet broken this down per dataset or diagnosed how much of the gap is attributable to the shared canvas (limitation above) versus the probe architecture itself; both are natural next steps.

Baselines are chosen to isolate mechanism, not to exhaustively benchmark the design space. Our reconstruction comparison (Shared / Partitioned / PACE-HAR) is a controlled ablation intended to separate the contribution of latent isolation from input alignment. Our classification check is intentionally narrow: a single frozen-backbone baseline (Freeze + Probe), meant only to establish that partitioning helps relative to an unpartitioned representation, not to benchmark against the continual-learning or HAR classification literature. We have not compared against, for instance, regularization-based continual learning (e.g., EWC), larger classifier-oriented expansion methods, adapter-only extension without codebook partitioning, domain-specific normalization, or replay-based continual learning at matched parameter budgets; such comparisons would clarify both how much of PACE-HAR's classification benefit is attributable to codebook partitioning specifically, and how the trade-off against dedicated recognition-oriented methods actually looks. We leave a fuller benchmark to future work.

8 Conclusion

We introduced PACE-HAR, a partitioned adapter-based VQ-VAE for interference-free cross-dataset capacity extension. The method combines disjoint codebook partitions with lightweight domain adapters to address latent-capacity interference and sensor-input mismatch while keeping the shared encoder–decoder path exactly fixed.

Across five heterogeneous HAR datasets, PACE-HAR preserves previously supported domains exactly (DPR = 1.00) at every extension step, improves reconstruction quality under domain shift by 8.28× over partitioning alone, and does so at only 166 trainable non-codebook parameters per added domain, regardless of domain order. A narrow downstream classification check indicates that some activity-relevant information survives this reconstruction bottleneck relative to an unpartitioned frozen baseline, but aggregate accuracy remains modest; we report this as an exploratory signal rather than a resolved result, and discuss its limits explicitly in Section 7. The resulting system is best viewed as a controlled, parameter-efficient mechanism for growing a shared generative representation across sensing configurations, not as a replacement for dedicated activity-recognition pipelines in its current form. Future work should relax the shared canvas constraint, for example through per-domain adapters that project native input shapes into a shared latent budget without a fixed spatial layout, and should study automatic domain routing, streaming deployments, and matched-budget comparisons against adapter-only and replay-based extension methods.

Acknowledgments.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Adaimi, R., Thomaz, E.: Lifelong adaptive machine learning for sensor-based human activity recognition using prototypical networks. In: Sensors (2022)
2. Anguita, D., Ghio, A., Oneto, L., Parra, X., Reyes-Ortiz, J.L.: A public domain dataset for human activity recognition using smartphones. In: European Symposium on Artificial Neural Networks (ESANN) (2013)
3. Banos, O., Garcia, R., Holgado-Terriza, J.A., et al.: mHealthDroid: A novel framework for agile development of mobile health applications. Ambient Assisted Living and Daily Activities pp. 91–98 (2014)
4. Chakma, A., et al.: Domain adaptation for sensor-based human activity recognition. In: IEEE International Conference on Pervasive Computing and Communications (PerCom) (2023)
5. Dhekane, S.G., Ploetz, T.: Transfer learning in human activity recognition: A survey. ACM Computing Surveys (2025)
6. Gudur, G.K., Sundaramoorthy, P., Umaashankar, V.: ActiveHARNet: Towards on-device deep bayesian active learning for human activity recognition. In: International Workshop on Embedded and Mobile Deep Learning (EMDL) (2019)
7. Hong, Z., et al.: CrossHAR: Generalizing cross-dataset human activity recognition via hierarchical self-supervised pretraining. In: Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT) (2024)
8. Houlsby, N., Giurgiu, A., Jastrzebski, S., et al.: Parameter-efficient transfer learning for NLP. In: International Conference on Machine Learning (ICML) (2019)
9. Jha, S., Schiemer, M., Zambonelli, F., Ye, J.: Continual learning in sensor-based human activity recognition: An empirical benchmark analysis. In: Information Sciences (2021)
10. Jiao, L., et al.: Vector quantization prompting for continual learning. In: Advances in Neural Information Processing Systems (NeurIPS) (2024)
11. Kirkpatrick, J., Pascanu, R., Rabinowitz, N., et al.: Overcoming catastrophic forgetting in neural networks. Proceedings of the National Academy of Sciences **114**(13), 3521–3526 (2017)
12. Kwapisz, J.R., Weiss, G.M., Moore, S.A.: Activity recognition using cell phone accelerometers. ACM SIGKDD Explorations Newsletter **12**(2), 74–82 (2011)
13. Leite, C., Xiao, Y.: Resource-constrained continual learning for human activity recognition. In: IEEE Transactions on Mobile Computing (2022)
14. Li, Z., Hoiem, D.: Learning without forgetting. In: IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) (2018)
15. Liu, Y., et al.: Cross-dataset human activity recognition with lightweight models. In: IEEE International Conference on Pervasive Computing and Communications (PerCom) (2024)
16. Lu, W., et al.: Semantic-discriminative mixup for generalizable sensor-based cross-domain activity recognition. In: Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT) (2022)

17. Malepathirana, T., Senanayake, D., Halgamuge, S.: NAPA-VQ: Neighborhood-aware prototype augmentation with vector quantization for continual learning. In: IEEE/CVF International Conference on Computer Vision (ICCV) (2023)
18. Mallya, A., Davis, D., Lazebnik, S.: Piggyback: Adapting a single network to multiple tasks by learning to mask weights. In: European Conference on Computer Vision (ECCV) (2018)
19. Mallya, A., Lazebnik, S.: PackNet: Adding multiple tasks to a single network by iterative pruning. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2018)
20. van den Oord, A., Vinyals, O., Kavukcuoglu, K.: Neural discrete representation learning. In: Advances in Neural Information Processing Systems (NeurIPS) (2017)
21. Qin, X., Chen, Y., Wang, J., Yu, C.: Adaptive spatial-temporal transfer learning for cross-domain activity recognition. In: International Joint Conference on Artificial Intelligence (IJCAI) (2019)
22. Razavi, A., van den Oord, A., Vinyals, O.: Generating diverse high-fidelity images with VQ-VAE-2. In: Advances in Neural Information Processing Systems (NeurIPS) (2019)
23. Rebuffi, S.A., Bilen, H., Vedaldi, A.: Learning multiple visual domains with residual adapters. In: Advances in Neural Information Processing Systems (NeurIPS) (2017)
24. Reiss, A., Stricker, D.: Introducing a new benchmarked dataset for activity monitoring. In: International Symposium on Wearable Computers (ISWC) (2012)
25. Roggen, D., Calatroni, A., Rossi, M., et al.: Collecting complex activity datasets in highly rich networked sensor environments. In: International Conference on Networked Sensing Systems (INSS) (2010)
26. Rusu, A.A., Rabinowitz, N.C., Desjardins, G., et al.: Progressive neural networks. In: arXiv preprint arXiv:1606.04671 (2016)
27. Sanabria, A.R., Zambonelli, F., Ye, J.: Contrastive adversarial domain adaptation for human activity recognition. In: IEEE Transactions on Neural Networks and Learning Systems (TNNLS) (2021)
28. Suh, S., et al.: TASKED: Transformer-based adversarial learning for human activity recognition using wearable sensors via self-knowledge distillation. In: Knowledge-Based Systems (2023)
29. Takida, Y., et al.: HQ-VAE: Hierarchical discrete representation learning with variational bayes. In: Transactions on Machine Learning Research (TMLR) (2024)
30. Wang, Z., Zhang, Z., Ebrahimi, S., et al.: DualPrompt: Complementary prompting for rehearsal-free continual learning. In: European Conference on Computer Vision (ECCV) (2022)
31. Wang, Z., Zhang, Z., Lee, C.Y., et al.: Learning to prompt for continual learning. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2022)
32. Ye, J., Nakwijit, P., Schiemer, M., et al.: Continual activity recognition with generative adversarial networks. In: ACM Transactions on Intelligent Systems and Technology (TIST) (2021)
33. Yoon, J., Yang, E., Lee, J., Hwang, S.J.: Lifelong learning with dynamically expandable networks. In: International Conference on Learning Representations (ICLR) (2018)