ConSense: Continually Sensing Human Activity with WiFi via Growing and Picking

Rong Li, Tao Deng*, Siwei Feng*, Mingjie Sun, Juncheng Jia

School of Computer Science and Technology, Soochow University, China
rli23@stu.suda.edu.cn, {dengtao, swfeng, mjsun, jiajuncheng}@suda.edu.cn

Abstract

WiFi-based human activity recognition (HAR) holds significant application potential across various fields. To handle dynamic environments where new activities are continuously introduced, WiFi-based HAR systems must adapt by learning new concepts without forgetting previously learned ones. Furthermore, retaining knowledge from old activities by storing historical exemplar is impractical for WiFi-based HAR due to privacy concerns and limited storage capacity of edge devices. In this work, we propose ConSense, a lightweight and fast-adapted exemplar-free class incremental learning framework for WiFi-based HAR. The framework leverages the transformer architecture and involves dynamic model expansion and selective retraining to preserve previously learned knowledge while integrating new information. Specifically, during incremental sessions, small-scale trainable parameters that are trained specifically on the data of each task are added in the multi-head self-attention layer. In addition, a selective retraining strategy that dynamically adjusts the weights in multilayer perceptron based on the performance stability of neurons across tasks is used. Rather than training the entire model, the proposed strategies of dynamic model expansion and selective retraining reduce the overall computational load while balancing stability on previous tasks and plasticity on new tasks. Evaluation results on three public WiFi datasets demonstrate that ConSense not only outperforms several competitive approaches but also requires fewer parameters, highlighting its practical utility in class-incremental scenarios for HAR.

Code — https://github.com/kikihub/consense

Introduction

Human activity recognition (HAR) technologies have broad use in scenarios such as medical monitoring (Ge et al. 2022), smart homes (Jobanputra, Bavishi, and Doshi 2019), and security detection (Lolla and Zhao 2019). Traditional video surveillance faces challenges related to privacy, field of view limitations, and lighting conditions. In contrast, wireless signal sensing enables non-invasive monitoring to safeguard privacy. WiFi stands out as an optimal choice for implementing HAR due to its widespread availability and small hardware requirements (Guo et al. 2019; Qian et al. 2017).

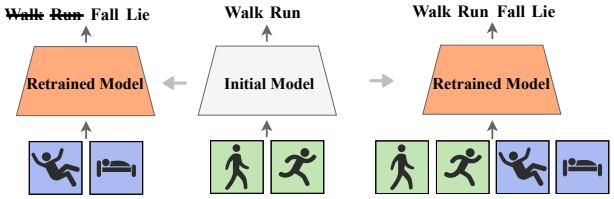


Figure 1: The initial model in the center is trained on two activities, walk and run. The retrained model on the left, fine-tuned with two new activities, fall and lie, loses the ability to recognize walk and run. In contrast, the retrained model on the right, which incorporates training data from both old and new activities, can recognize all four activities.

For WiFi-based HAR, conventional deep learning (DL) models (Xia, Huang, and Wang 2020; Abuhoureyah, Wong, and Isira 2024) struggle with identifying novel activities because these systems depend on static models that cannot adjust to emerging human activities. Thus, there is a pressing need to design DL models that can adeptly adapt to dynamically changing environments. Straightforward fine-tuning on DL models with new data without access to the original training data can lead to catastrophic forgetting, where previously learned knowledge is overwritten.

As illustrated in Figure 1, the model retrained on new activities (fall and lie) fails to recognize prior activities (walk and run). An alternative is to retrain a new model using data from both new and previous activities. However, storing historical data is challenging due to strict data privacy policies (Li et al. 2014) that limit unregulated storage and transfer, as well as the storage constraints of edge devices (Hernandez and Bulut 2020). These challenges require the development of HAR systems that can continually sensing without storing user data, thereby ensuring user privacy and system effectiveness.

Exemplar-free class-incremental learning (EFCIL) (Li and Hoiem 2017; Zhu et al. 2022; Goswami et al. 2024) aims to recognize both old and new classes without retaining exemplars from previous classes, addressing concerns regarding privacy and storage. While extensively explored in computer vision, applying existing EFCIL designed for computer vision tasks to WiFi-based HAR faces distinctive challenges. Unlike images, WiFi, being a wireless signal, undergoes subtle and time-sensitive changes due to human activ-

*Tao Deng and Siwei Feng are the corresponding authors.

ities, making stable feature extraction difficult. This challenge is exacerbated by the continuous and rapidly changing nature of WiFi data, which lacks clear spatial references. Consequently, it's crucial to enhance models that can concurrently capture spatial and temporal characteristics in time series to adapt to the dynamic nature of WiFi data. Moreover, existing EFCIL approaches often require significant computational resources and lengthy training times, limiting their practicality for resource-constrained edge devices. This creates an urgent need for lightweight, fast-training continual sensing models that can efficiently manage computation and storage resources for WiFi-based HAR.

To solve these challenges, in this paper we propose ConSense, a continual dynamic adaptive learning framework for WiFi-based HAR. To capture temporal and spatial relationships in sequential data, ConSense leverages the transformer architecture, which is particularly suitable for processing data with complex spatio-temporal characteristics. Additionally, ConSense preserves previously learned knowledge while integrating new information by **growing** with **dynamic model expansion** and **picking** with **selective retraining**. Specifically, we add small-scale trainable parameters, referred to as prefixes, within the multi-head self-attention (MHSA) layer. These prefixes are custom-designed and trained specifically for data corresponding to each individual task, allowing the model to effectively capture and retain key task-specific knowledge. This training strategy ensures that the unique features of each task are understood and preserved within the model architecture. In addition, a selective retraining strategy is employed, which dynamically adjusts the weights of neurons in the multilayer perceptron (MLP) based on their performance across different tasks. By monitoring the stability of neuron, this module identifies which aspects of information should be maintained over time and which should be adjusted to accommodate new data. This selective weighting not only helps the model acquire new knowledge without overwriting existing information but also enhances the system's overall adaptability. These strategies allow ConSense to maintain plasticity and stability during continual sensing while enabling fast training by updating a smaller set of task-specific parameters instead of retraining the entire model.

We validate our proposed framework on three publicly available WiFi datasets, confirming that ConSense exceeds the performance of other models while utilizing fewer parameters.

Related Work

Static DL models for WiFi cannot adapt to the dynamic environment where new activities are constantly introduced. In computer vision, some works proposed class incremental learning to solve this aspect. Class-incremental learning (CIL) is classified as three categories: 1) rehearsal-based method, which preserves knowledge by replaying exemplars from past tasks (Rebuffi et al. 2017; Hou et al. 2019; Wu et al. 2019); 2) regularization-based method, which uses penalties to maintain critical parameters but struggles with lengthy task sequences (Kirkpatrick et al. 2017; Yang et al. 2021; Saha, Garg, and Roy 2021), and 3) dynamic architecture-based method, which expands the model for new tasks but can be resource-intensive (Rusu et al. 2016; Verma et al. 2021; Douillard et al. 2022). Many existing CIL methods rely on storing exemplars from previous tasks to address catastrophic forgetting. However, the limited storage and computational power of commonly used edge devices, combined with privacy concerns around retaining user data, restrict the applicability of these methods in real-world scenarios.

Some works propose exemplar-free CIL (EFCIL) to avoid the need to retain exemplars. Li *et al.* (Li and Hoiem 2017) introduced knowledge distillation (KD) for CIL. However, KD has limited effects when only new data is used. Gao *et al.* (Gao et al. 2022) proposed a new framework that separates representation and classifier learning, thus improving model inversion to synthesize data for previous tasks. Asadi *et al.* (Asadi et al. 2023) introduced prototype-sample relation distillation by combining supervised contrastive loss (Khosla et al. 2020), self-supervised learning (Liu et al. 2021), and class prototype evolution techniques (De Lange and Tuytelaars 2021). By jointly learning representations and class prototypes, they effectively reduce catastrophic forgetting and maintain the relevance as well as the embedding similarity of old class prototypes. The above methods mainly used ResNet and other convolutional neural network (CNN)-based models. In general, the two dimensions of WiFi (time stamps and channel state) are fundamentally different from the two dimensions of images that contains spatial information. Using a two-dimensional kernel to extract spatial patterns from WiFi data would result in poor feature extraction.

Although many CNN and ResNet-based approaches have been developed for EFCIL, transformer-based EFCIL remains a relatively unexplored area. Roy *et al.* (Roy et al. 2023) adapted the transformer's MHSA layers with convolution operations for new tasks. However, their proposed method is unsuitable for WiFi data, as it relies on image augmentation, whereas WiFi augmentation strategies involve temporal delays and frequency shifts, leading to performance degradation. Zhang *et al.* (Zhang et al. 2023) and Ding *et al.* (Ding et al. 2023) investigated WiFi-based HAR using incremental learning. Zhang *et al.* employed retained exemplars and distillation loss to preserve activity knowledge, while Ding *et al.* introduced an enhancement CNN with attention and dual-loss functions. However, Ding's approach processes only one category at a time, limiting its applicability. To overcome the limitations of the above methods, we propose a new model specifically designed for WiFi-based HAR, which can more effectively adapt to the dynamic changes in WiFi data while reducing the need for data storage.

Method

In EFCIL, a model sequentially learns tasks $\{T_t\}_{t=1}^{T}$, each introducing a unique set of classes C_t, with no class overlap between tasks. The model only accesses the current task's training set $\{X_t, Y_t\}$, where X_t are samples and Y_t are corresponding labels, without storing previous exemplars. The objective of EFCIL is to maximize classification accuracy across all classes encountered up to T_t.

The proposed framework performs dynamic model ex-

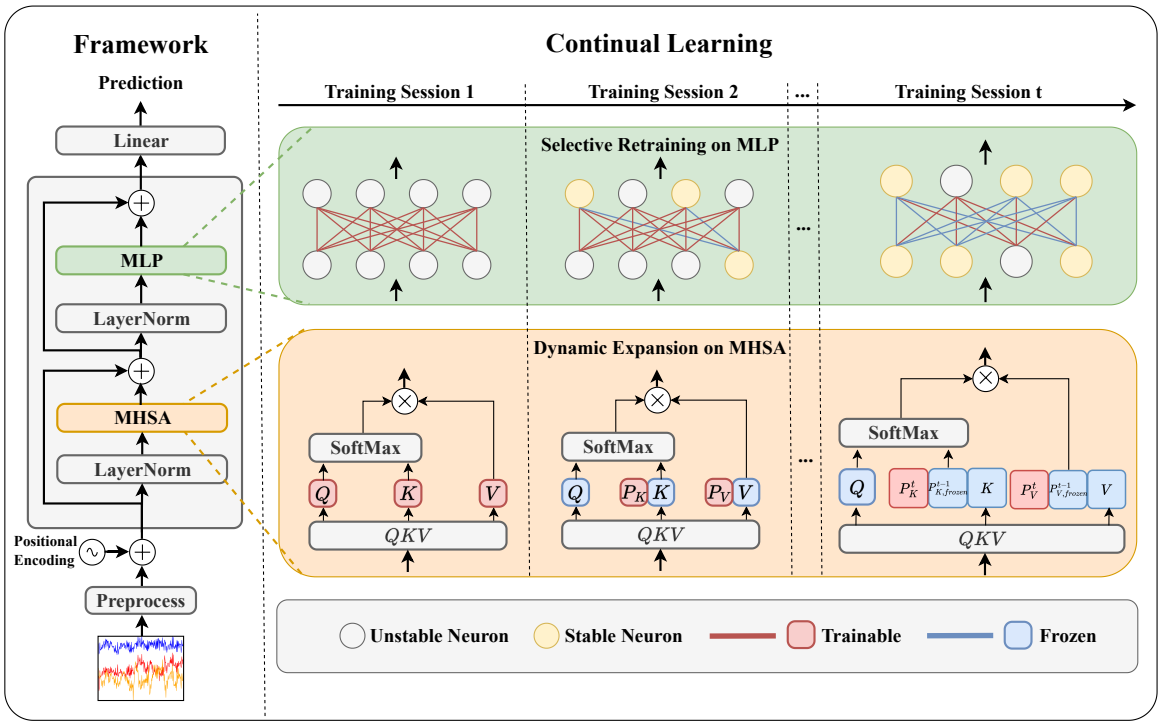


Figure 2: Architecture of ConSense. Left part contains the framework. Right part details how the model dynamically expands and selectively retrains during continual learning from training session 1 to training session t. As new tasks are introduced, the model dynamically expands with new prefixes in the MHSA layer. In the MLP, a selective retraining strategy is implemented to adjust neuron weights, preserving learned outcomes from stable neurons while updating unstable neurons to accommodate new tasks.

pansion and selective retraining to preserve learned knowledge and learn new classes. During incremental sessions, trainable prefixes are added to the MHSA layer, and a dynamic selective retraining strategy adjusts MLP weights based on neuron performance stability across tasks. The overall framework is illustrated in Figure 2.

Preliminaries

Input WiFi-based HAR utilizes channel state information (CSI) to capture subtle variations in signal characteristics like phase and amplitude, which arise due to environmental interactions and human movement. CSI effectively tracks these changes across multiple subcarrier frequencies, offering a precise method for activity recognition by analyzing how signals interact with physical obstacles and their dynamic alterations in a given space.

Architecture The proposed framework employs a transformer-based architecture, which consists of three main components: positional encoding, a MHSA layer, a MLP.

Sequential information is crucial for recognizing reverse actions such as sitting down and standing up. Traditional absolute or relative position encodings assign a unique and highly distinctive code to each individual point, which can introduce noise. To enhance the capture of sequential information for activities, we employ Gaussian range encoding (Li et al. 2021). This positional encoding method assigns multiple encoding ranges to each position in the data. It allows dynamic adjustments during training based on Gaussian distributions characterized by means μ and standard deviations σ.

MHSA's input sequence, $X \in \mathbb{R}^{n \times d}$, has undergone Gaussian range encoding, where n denotes the temporal dimension and d the spatial dimension of CSI, matching the original input CSI dimensions. Each attention head h_i, where i represents the index of the attention head, uses three matrices W_i^Q, W_i^K, and W_i^V to transform X into queries Q_i, keys K_i, and values V_i:

$$\text{Attention}(Q_i, K_i, V_i) = \text{softmax}\left(\frac{Q_i K_i^T}{\sqrt{d_k}}\right) V_i, \quad (1)$$

where d_k represents the dimension of the key, query and value vectors. $Q_i = XW_i^Q$, $K_i = XW_i^K$, $V_i = XW_i^V$. The final output is generated by concatenating the results from all heads and transforming them through a linear transformation, combining the individual outputs into a comprehensive representation for the entire sequence. The output dimension is the same as the input dimension.

The MLP used in our framework is a type of feedforward neural network, consisting of one input layer, two hidden layers, one output layer, and ReLU activation functions between each layer.

Training Procedure

The training pipeline of ConSense includes an initial training stage and incremental training stages. During the ini-

tial training stage T_1, the model learns the first batch of classes C_1, with all model parameters including the weights in the MHSA layer, the MLP and the classifier being trainable. When the initial training stage is completed, the parameters in the MHSA layer are frozen as frozen as $W_{frozen}^{(MHSA)} = \{W_{frozen}^{Q}, W_{frozen}^{K}, W_{frozen}^{V}\}$, representing the frozen weights for query, key and value from the MHSA layer. In incremental session t, we dynamically expand the MHSA layer and selectively retrain the MLP to balance plasticity and stability with minimal computation. For MHSA, new trainable prefixes $P^t = \{P_{K,frozen}^t, P_{V,frozen}^t\}$ are added, working with frozen weights $W_{frozen}^{(MHSA)}$ and previous task prefixes P_{frozen}^{t-1} to adapt to new classes C_t. $P_{frozen}^{t-1} = \{P_{K,frozen}^{t-1}, P_{V,frozen}^{t-1}\}$ represent the frozen key and value prefixes from the previous task. In the MLP, we average neuron activation values, identify stable neurons, and use freeze mask to update only unstable neurons. After completing task T_t, the previous task prefixes P_{frozen}^{t-1} are updated to incorporate the new prefixes from task T_t. Specifically, P_{frozen}^{t-1} is updated as follows: $P_{frozen}^t = \{P_{K,frozen}^t, P_{V,frozen}^t\}$, where $P_{K,frozen}^t = [P_K^t, P_{K,frozen}^{t-1}]$ and $P_{V,frozen}^t = [P_V^t, P_{V,frozen}^{t-1}]$, where [,] denotes the concatenate operation. Details on dynamic expansion on MHSA and selective retraining on MLP are described as follows.

Input For model input, CSI is structured into dimensions $T_{temp} \times C_{ch}$, where T_{temp} represents the temporal dimension, defined as $T_{temp} = vs$ (v represents the frequency of data packet collection, and s represents the sampling time for an action). The channel dimension C_{ch}, defined as $C_{ch} = eg$ (e represents the number of transmitting and receiving antenna combinations, and g represents the number of subcarriers per antenna pair).

Dynamic Expansion on MHSA The goal of acquiring new knowledge while retaining old knowledge without storing exemplars can be achieved by adding task-specific prefixes to the MHSA layers. This method allows knowledge transfer between tasks without significantly changing the model's original parameters. By keeping the model's parameters fixed and updating only the prefixes, the system avoids catastrophic forgetting while maintaining adaptability. Specifically, each MHSA layer has H attention heads, and adding prefixes to these layers enables class-incremental learning for new tasks. Concretely, we have

$$\begin{aligned} W_i^{K\prime} &= [P_K^t, P_{K,frozen}^{t-1}, W_{frozen}^K], \\ W_i^{V\prime} &= [P_V^t, P_{V,frozen}^{t-1}, W_{frozen}^V]. \end{aligned} \tag{2}$$

The output of a head in the self-attention layer is formulated as:

$$\text{head}_i = \text{Attention}(Q_i, K_i\prime, V_i\prime), \tag{3}$$

where $Q_i = XW_{frozen}^Q, K_i\prime = XW_i^{K\prime}, V_i\prime = XW_i^{V\prime}$, and Attention() is defined in Eq.1. To effectively integrate prior knowledge with new information, the model sequentially concatenates the new trainable prefixes P_K^t and P_V^t with the previously frozen prefixes $P_{K,frozen}^{t-1}$ and $P_{V,frozen}^{t-1}$. These concatenated prefixes are then merged with the consistently frozen weights W_{frozen}^K and W_{frozen}^V, employing concatenation to ensure a seamless transition and retention of learned features across tasks.

While prefixes are suitable for class-incremental learning tasks, their random initialization can lead to unstable performance due to varying initial weights. To address this, we took inspiration from parallel attention design (Yu et al. 2022), which uses a parallel adapter to stabilize prefixes. Specifically, with input sequence X, the prefix generation is formulated as:

$$P_K, P_V = Adapter(X) = Tanh(XW_{down})W_{up}, \tag{4}$$

where $Tanh$ is the activation function, and W_{down} and W_{up} are the parameters of the parallel adapter's scaling layers. W_{down} is a linear transformation layer that reduces the dimensionality of X, and W_{up} is another linear transformation that expands the transformed output.

Selective Retraining on MLP While the MHSA layer captures the temporal features of CSI signals through linear transformations, the added MLP layer introduces nonlinear transformations to better capture complex features. To prevent forgetting issues, we utilize a selective retraining strategy based on neuron activation in all MLP layers. This method involves three main steps: calculating each neuron's average activation value, identifying stable neurons, and generating freeze masks for parameter updates. The process is applied independently to each of the linear layers in the MLP.

First, given a training set, we calculate the average activation value $\bar{a}_p^{(l)}$ for each neuron in the l-th layer, defined as:

$$\bar{a}_p^{(l)} = \frac{1}{B} \sum_{q=1}^{B} a_p^{(q,l)}, \tag{5}$$

where B is the size of the training set, l denotes the layer index, and $a_p^{(q,l)}$ is the activation value of the p-th neuron in the l-th layer for the q-th sample.

Next, by comparing the current activation values with those from the previous task, we identify the set of stable neurons $S^{(l)}$ in each layer, which is:

$$S^{(l)} = \{p \mid \|\bar{a}_p^{(l,t)} - \bar{a}_p^{(l,t-1)}\|_2 \leq \epsilon\}, \tag{6}$$

where ϵ is a predefined threshold, and $\bar{a}_p^{(l,t)}$ and $\bar{a}_p^{(l,t-1)}$ represent the average activation values for the current and previous tasks in the l-th layer, respectively.

Finally, we generate the freeze mask set $M^{(l)} = \{M_W^{(l)}, M_b^{(l)}\}$ based on the set of stable neurons $S^{(l)}$ for each MLP layer. Specifically, $M_W^{(l)}$ and $M_b^{(l)}$ are mask matrices corresponding to the weight matrix $W^{(l)}$ and bias vector $b^{(l)}$ in the l-th layer, respectively, and are initialized with values set to one. For stable neurons in the set $S^{(l)}$, the corresponding values in $M_W^{(l)}$ and $M_b^{(l)}$ are set to zero.

During backpropagation, these masks are applied across all layers by identifying positions where $M_W^{(l)}$ and $M_b^{(l)}$ have

a value of zero. At these positions, the corresponding gradients of $W^{(l)}$ and $b^{(l)}$ are set to zero, ensuring that these parameters are not updated. Parameters that are not frozen continue to be updated normally.

In this manner, it reduces forgetting by preserving the weights of stable neurons and prevents the excessive computational load during new task learning.

Experiments

Datasets and Settings

Datasets with a limited number of categories are not suitable for evaluating class-incremental learning. Therefore, we selected the WiAR (Guo et al. 2019), MMFi (Yang et al. 2024), and XRF (Wang et al. 2024) datasets, which offer a broader range of categories. The statistics of these datasets are summarized in Table 1.

Dataset	Class	Size	Train	Test
WiAR	16	270×90	384	96
MMFi	27	10×342	2160	540
XRF	48	50×270	672	288

Table 1: Statistics of the evaluation datasets. The size of each dataset is denoted as $T_{temp} \times C_{ch}$, where T_{temp} represents the temporal dimension and C_{ch} represents the channel dimension.

WiAR consists of 480 CSI samples, evenly distributed across 16 distinct classes. We divided the dataset into training and testing subsets at a 4:1 ratio. After our processing, the sample size is 270 x 90. Additionally, we organized WiAR into two task types: short task and long task. The short task set includes 5 tasks: the first task covers 8 classes, while the following 4 tasks cover 2 classes each. In contrast, the long task set comprises 8 tasks, with each task consistently including 2 classes.

MMFi comprises 2700 CSI samples, evenly distributed across 27 classes. After our processing, the sample size is 10 x 342. In MMFi, the short task category includes a total of 6 tasks. The first task covers 12 classes, while each of the next five tasks covers 3 classes. In contrast, the long task category consists of 9 tasks, with each task handling 3 classes.

XRF initially includes 55 classes, with 7 dedicated to dual-person actions. We exclude these dual-person classes due to our focus on single-person activities, leaving 48 classes and 960 CSI samples. Each class contains 20 samples, with 14 samples per class allocated for training and the remaining 6 used for testing. After our processing, the sample size is 50 x 270. In XRF, the short task category consists of 5 tasks: the first task covers 24 classes, while each of the subsequent 4 tasks contains 6 classes. In contrast, the long task category is organized into 8 tasks, each responsible for analyzing 6 classes.

Baselines

We compare ConSense with five existing EFCIL methods: (1) LWF (Li and Hoiem 2017) uses knowledge distillation to mitigate forgetting. (2) PASS (Zhu et al. 2021) combines prototype augmentation with self-supervised learning to enhance memory of old classes. (3) R-DFCIL (Gao et al. 2022) synthesizes data for previous classes using model inversion and applies relation-guided representation learning to minimize the domain gap between synthetic and real data. (4) PRD (Asadi et al. 2023) introduces a new distillation loss to maintain the relevance of class prototypes during new task learning. (5) ConTraCon (Roy et al. 2023) modifies MHSA layer weights via convolutional operations to adapt the transformer architecture for new tasks.

Evaluation Metrics

We use two metrics, i.e., the average accuracy and average forgetting measure the performance of ConSense on all the classes seen so far. The accuracy after each task, denoted by A_t, represents the accuracy over all classes learned up to and including the t-th task. Subsequently, the average accuracy across all tasks, represented by $\bar{A}$, is expressed as: $\bar{A} = \frac{1}{N}\sum_{t=1}^{N} A_t$, where N represents the number of tasks. The average forgetting measure (Chaudhry et al. 2018) is used to estimate the forgetting of previous tasks. For each task t, the forgetting measure of predicting previous task k is denoted by f_k^t, which is expressed as $f_k^t = \max_{z\in\{1,\ldots,k-1\}}(\alpha_{z,t} - \alpha_{z,t})$, where $\alpha_{m,j}$ represents the accuracy of task j after training task m. The average forgetting measure represents the forgetting measure of the last task, denoted by $\bar{F}$, which is expressed as $\bar{F} = \frac{1}{N-1}\sum_{k=1}^{N-1} f_k^N$.

Implementation Details

We set the number of Gaussian distributions in the positional encoding to 10. The values of μs are uniformly distributed across the temporal dimension for various datasets as follows. For WiAR, they range from 13.5 to 256.5 with a step size of 27. For MMFi, they range from 0.5 to 9.5 with a step size of 1. For XRF, they range from 2.5 to 47.5 with a step size of 5. The standard deviation of the Gaussian distributions on all the datasets is uniformly set to 8. The number of stacks in the module is set to 1. The input dimensions for the three datasets are set to 90, 342, and 270, respectively, while maintaining a consistent number of heads at 9 for each, and employing a dropout rate of 0.1.

Our method is implemented by PyTorch (Paszke et al. 2019) and trained on NVIDIA A5000 GPU with 32GB memory. The optimizer chosen is Adam (Kingma and Ba 2014), with an initial learning rate of 0.001 and a batch size of 16. The model's training cycle is set to 50 epochs.

Comparative Results

Performance Comparison Tables 2 and 3 present the results of the average accuracy $\bar{A}$ and the average forgetting $\bar{F}$, respectively. From the two tables, we observe that ConSense significantly outperforms other methods on all the datasets. Specifically, in the long task sequences of WiAR dataset, the average accuracy of ConSense surpasses that of other methods by nearly 30%. In the short task sequences of MMFi dataset, the average accuracy improvement exceeds 30%. Especially compared to LWF, the advantage of ConSense

Method	Replay Data	WiAR			MMFi			XRF		
		Params	$N=5$	$N=8$	Params	$N=6$	$N=9$	Params	$N=5$	$N=8$
LWF	-	18.52M	40.77	37.26	18.52M	33.81	30.54	18.52M	33.30	29.77
PASS	-	11.32M	59.15	40.96	11.32M	45.19	39.29	11.32M	48.65	35.93
R-DFCIL	Synthetic	12.81M	60.63	57.76	12.81M	50.95	47.83	12.81M	49.81	44.30
PRD	-	11.75M	64.58	60.23	11.75M	54.46	52.22	11.75M	54.44	51.57
ConTraCon	-	3.60M	-	48.58	2.50M	-	44.08	2.10M	-	41.03
ConSense	-	**3.35M**	**91.66**	**89.85**	**1.92M**	**84.42**	**71.97**	**1.50M**	**66.19**	**65.79**

Table 2: The average accuracy $\bar{A}$ (%) comparison of Our ConSense with other five methods on WiAR, MMFi, and XRF with short task ($N=5$ or $N=6$) and long task ($N=8$ or $N=9$). Params refers to the initial number of parameters of a model, measured in millions. None of the methods utilize real historical data for replay. R-DFCIL employs synthetic data to simulate the replay data.

Method	WiAR		MMFi		XRF	
	$N=5$	$N=8$	$N=6$	$N=9$	$N=5$	$N=8$
LWF	31.46	31.38	31.47	33.07	32.91	29.98
PASS	24.51	28.94	22.64	25.68	20.49	28.74
R-DFCIL	22.84	24.83	20.27	24.74	21.05	27.63
PRD	20.30	21.69	19.30	19.34	24.34	20.90
ConTraCon	-	28.15	-	29.88	-	25.59
ConSense	**14.31**	**12.89**	**16.28**	**17.99**	**19.51**	**18.09**

Table 3: The average forgetting measure $\bar{F}$ (%) comparison of Our ConSense with other five methods on WiAR, MMFi, and XRF with short and long tasks (lower is better).

reaches 50%. In addition, ConSense achieves a forgetting rate of less than 20% for both short and long tasks on the three datasets, and outperforms other methods. The two insights manifest that ConSense effectively balances plasticity and stability. The reason is that in ConSense, we utilize MHSA and positional encoding. This design particularly adapts to the characteristics of time-series data, such as the patterns and intensity of signal changes in CSI. These features pose challenges to traditional image-based network architectures, like Resnet, which primarily optimizes for spatial feature extraction and struggles with the dynamic characteristics of time-series data. However, for other methods, the knowledge distillation approach of LWF does not fare well in dynamically changing environments, and the synthetic data approach of R-DFCIL fails to accurately capture the true characteristics of CSI. PRD attempts to mitigate forgetting by maintaining relationships between class prototypes, but the high dynamism and complexity of CSI may render this prototype-based method ineffective. ConTraCon uses the Transformer architecture to adapt to new tasks, but its success depends on the effectiveness of its attention mechanism. If this mechanism fails to capture the temporal and frequency domain characteristics of CSI, the results may fall short. Moreover, its entropy-based task prediction, which relies on image enhancement techniques, is unsuitable for CSI, as CSI variations like temporal delays and frequency shifts don't correspond to visual changes.

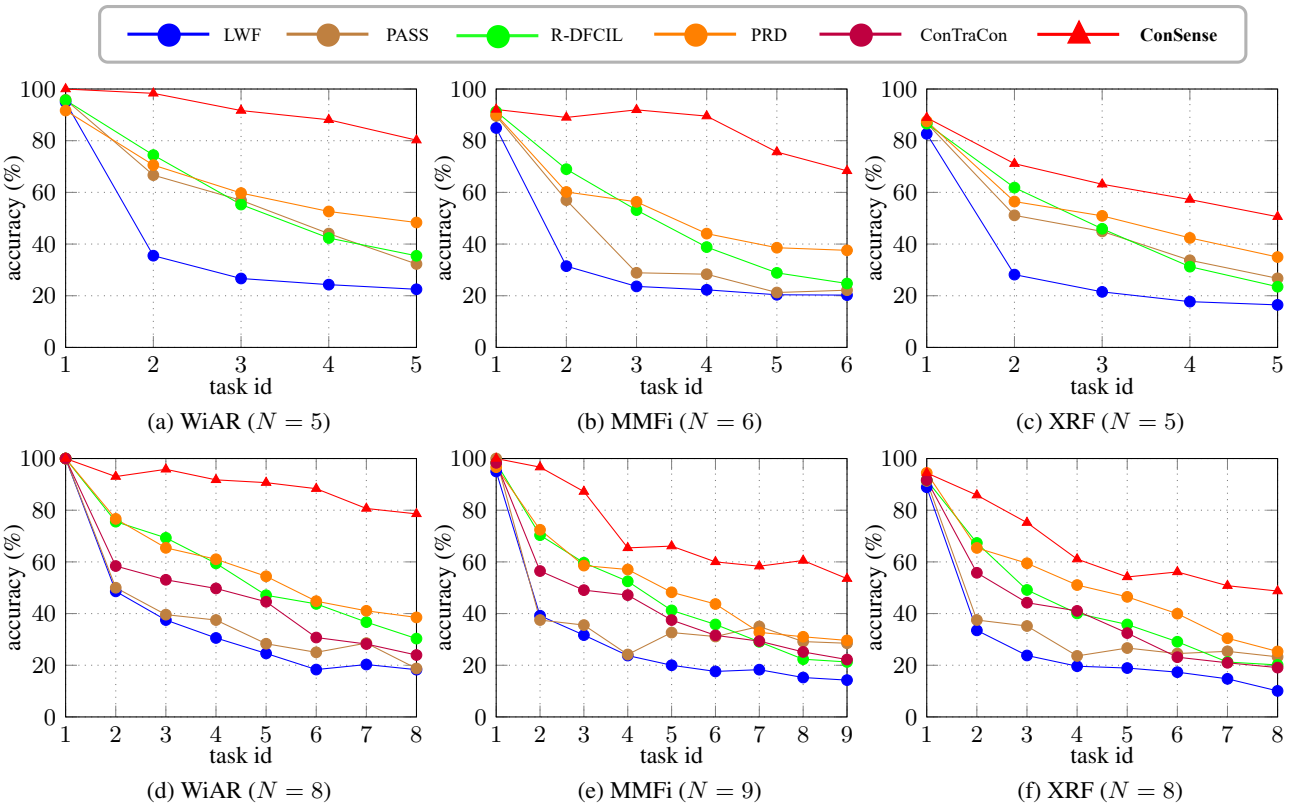
Parameters Comparison Moreover, ConSense exhibits a marked reduction in model parameters compared to other methods. Specifically, on the WiAR dataset, the parameter count of ConSense is comparable to ConTraCon, but at least three times less than other methods. Notably, on the MMFi and XRF datasets, the parameter count of ConSense is even only one-sixth of that of other methods excluding ConTraCon. This advantage manifests that ConSense is especially suitable for the deployment of edge devices, e.g., WiFi-based HAR terminal.

Accuracy of Each Task In Figure 3, we compare the accuracy of ConSense with other methods on each task on three datasets. We observe that the accuracy of the initial task for all the methods is comparable, and the accuracy of ConSense significantly outperforms other methods in subsequent tasks. This demonstrates that our method achieves a better balance between knowledge forgetting and acquisition when dealing with CSI. In addition, the performance gain of ConSense compared to other methods widens with the increase of the number of tasks. For example, in Figure 3(e), at the fifth task, the gain of ConSense compared to PRD is approximately 10%, while at the ninth task, the gain has reached 20%. This trend emphasizes the effectiveness of ConSense in handling extended task sequences, highlighting its robustness in continually sensing.

Ablation Test

Dataset	Strategy 1	Strategy 2	A_T	$\bar{A}$
WiAR	×	×	36.45	52.08
	✓	×	51.03	67.70
	×	✓	44.79	59.37
	✓	✓	**78.58**	**89.85**
MMFi	×	×	25.91	46.27
	✓	×	38.87	60.15
	×	✓	34.24	54.78
	✓	✓	**53.52**	**71.97**
XRF	×	×	22.56	40.96
	✓	×	41.66	53.12
	×	✓	30.90	48.60
	✓	✓	**48.71**	**65.79**

Table 4: Ablation study of two strategies for long task on three datasets. Strategy 1 represents dynamic expansion on MHSA. Strategy 2 represents selective retraining on MLP.



(a) WiAR ($N = 5$) (b) MMFi ($N = 6$) (c) XRF ($N = 5$)

(d) WiAR ($N = 8$) (e) MMFi ($N = 9$) (f) XRF ($N = 8$)

Figure 3: Accuracy comparison of ConSense with other five methods on each task. The x-axis represents the t-th task, and the y-axis represents the accuracy of the t-th task, i.e., A_t.

Effect of Dynamic Expansion and Selective Retraining Table 4 shows that the two strategies, dynamic expansion on MHSA and selective retraining on MLP, significantly enhance the performance of ConSense in long-term EFCIL ablation experiments. More specifically, dynamic expansion significantly enhances the accuracy of the last task and average task accuracy over all the datasets. It achieves this by adding trainable prefixes to the multi-head self-attention layers, thereby emphasizing its robustness against forgetting and adaptability to new tasks. Compared to dynamic expansion, while the performance gains observed with selective retraining are less pronounced, it still contributes to model performance enhancement, particularly through stabilizing trained parameters in specific scenarios.

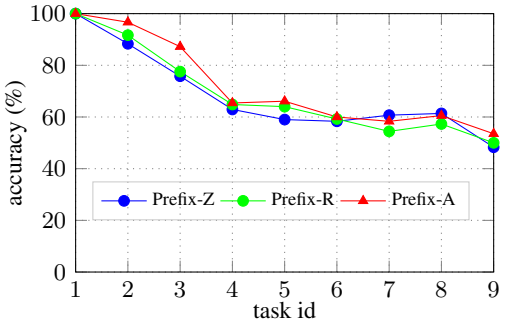


Figure 4: Ablation study of the impact of parallel adapter for long task on MMFi.

Effect of Parallel Adapter Figure 4 shows that the parallel adapter initialization (Prefix-A) significantly outperforms zero (Prefix-Z) and random (Prefix-R) initializations in ConSense. It achieves an average accuracy of 71.97%. In contrast, Prefix-R attains 69.00% accuracy, while Prefix-Z achieves 68.30%. This highlights the effectiveness of the parallel adapter in enhancing prefix handling and overall model performance.

Conclusion

We propose ConSense, a lightweight and fast-adapted exemplar-free class incremental learning framework for WiFi-based HAR. By leveraging the transformer architecture, ConSense effectively handles the challenges of continual learning in dynamic environments. The framework's key innovations, including dynamic model expansion on MHSA and selective retraining on MLP, enable fast training by focusing on integrating new information while preserving previously acquired knowledge. Comparative tests on three datasets show that ConSense improves average accuracy by over 10% across all tasks and maintains a forgetting rate below 20%, outperforming existing methods. Moreover, it reduces model parameters significantly, making it ideal for resource-constrained environments. Ablation studies highlight the effectiveness of two strategies and parallel adapters in enhancing stability and accuracy. Future efforts will focus on real-world applications and further optimization for edge deployments.

Acknowledgments

This work is supported by Natural Science Foundation of Jiangsu Province, China (Grant No. BK20230477 and BK20230482), National Science Foundation of China (NSFC No. 62302328 and No. 62106167), and the Priority Academic Program Development of Jiangsu Higher Education Institutions, Suzhou Frontier Science and Technology Program (Project SYG202310).

References

Abuhoureyah, F. S.; Wong, Y. C.; and Isira, A. S. B. M. 2024. WiFi-based human activity recognition through wall using deep learning. *Engineering Applications of Artificial Intelligence*, 127: 107171.

Asadi, N.; Davari, M.; Mudur, S.; Aljundi, R.; and Belilovsky, E. 2023. Prototype-sample relation distillation: towards replay-free continual learning. In *International Conference on Machine Learning*, 1093–1106.

Chaudhry, A.; Dokania, P. K.; Ajanthan, T.; and Torr, P. H. 2018. Riemannian walk for incremental learning: Understanding forgetting and intransigence. In *Proceedings of the European Conference on Computer Vision*, 532–547.

De Lange, M.; and Tuytelaars, T. 2021. Continual prototype evolution: Learning online from non-stationary data streams. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 8250–8259.

Ding, X.; Zhong, Y.; Wu, S.; Jiang, C.; and Xie, W. 2023. Passive sensing for class-incremental human activity recognition. *IEEE Geoscience and Remote Sensing Letters*, 20: 1–5.

Douillard, A.; Ramé, A.; Couairon, G.; and Cord, M. 2022. Dytox: Transformers for continual learning with dynamic token expansion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9285–9295.

Gao, Q.; Zhao, C.; Ghanem, B.; and Zhang, J. 2022. R-dfcil: Relation-guided representation learning for data-free class incremental learning. In *European Conference on Computer Vision*, 423–439.

Ge, Y.; Taha, A.; Shah, S. A.; Dashtipour, K.; Zhu, S.; Cooper, J.; Abbasi, Q. H.; and Imran, M. A. 2022. Contactless WiFi sensing and monitoring for future healthcare-emerging trends, challenges, and opportunities. *IEEE Reviews in Biomedical Engineering*, 16: 171–191.

Goswami, D.; Soutif-Cormerais, A.; Liu, Y.; Kamath, S.; Twardowski, B.; van de Weijer, J.; et al. 2024. Resurrecting Old Classes with New Data for Exemplar-Free Continual Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 28525–28534.

Guo, L.; Wang, L.; Lin, C.; Liu, J.; Lu, B.; Fang, J.; Liu, Z.; Shan, Z.; Yang, J.; and Guo, S. 2019. Wiar: A public dataset for wifi-based activity recognition. *IEEE Access*, 7: 154935–154945.

Hernandez, S. M.; and Bulut, E. 2020. Lightweight and standalone IoT based WiFi sensing for active repositioning and mobility. In *IEEE International Symposium on" A World of Wireless, Mobile and Multimedia Networks"*, 277–286.

Hou, S.; Pan, X.; Loy, C. C.; Wang, Z.; and Lin, D. 2019. Learning a unified classifier incrementally via rebalancing. In *Proceedings of the IEEE/CVF Conference on Computer vision and pattern recognition*, 831–839.

Jobanputra, C.; Bavishi, J.; and Doshi, N. 2019. Human activity recognition: A survey. *Procedia Computer Science*, 155: 698–703.

Khosla, P.; Teterwak, P.; Wang, C.; Sarna, A.; Tian, Y.; Isola, P.; Maschinot, A.; Liu, C.; and Krishnan, D. 2020. Supervised contrastive learning. *Advances in Neural Information Processing Systems*, 33: 18661–18673.

Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kirkpatrick, J.; Pascanu, R.; Rabinowitz, N.; Veness, J.; Desjardins, G.; Rusu, A. A.; Milan, K.; Quan, J.; Ramalho, T.; Grabska-Barwinska, A.; et al. 2017. Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences*, 114(13): 3521–3526.

Li, B.; Cui, W.; Wang, W.; Zhang, L.; Chen, Z.; and Wu, M. 2021. Two-stream convolution augmented transformer for human activity recognition. In *Proceedings of the AAAI Conference on Artificial intelligence*, volume 35, 286–293.

Li, H.; Sun, L.; Zhu, H.; Lu, X.; and Cheng, X. 2014. Achieving privacy preservation in WiFi fingerprint-based localization. In *Ieee Infocom 2014-IEEE Conference on Computer Communications*, 2337–2345.

Li, Z.; and Hoiem, D. 2017. Learning without forgetting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(12): 2935–2947.

Liu, X.; Zhang, F.; Hou, Z.; Mian, L.; Wang, Z.; Zhang, J.; and Tang, J. 2021. Self-supervised learning: Generative or contrastive. *IEEE Transactions on Knowledge and Data Engineering*, 35(1): 857–876.

Lolla, S.; and Zhao, A. 2019. WiFi motion detection: A study into efficacy and classification. In *IEEE Integrated STEM Education Conference*, 375–378.

Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems*, 32.

Qian, K.; Wu, C.; Yang, Z.; Liu, Y.; and Jamieson, K. 2017. Widar: Decimeter-level passive tracking via velocity monitoring with commodity Wi-Fi. In *Proceedings of the 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, 1–10.

Rebuffi, S.-A.; Kolesnikov, A.; Sperl, G.; and Lampert, C. H. 2017. icarl: Incremental classifier and representation learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2001–2010.

Roy, A.; Verma, V. K.; Voonna, S.; Ghosh, K.; Ghosh, S.; and Das, A. 2023. Exemplar-free continual transformer with convolutions. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 5897–5907.

Rusu, A. A.; Rabinowitz, N. C.; Desjardins, G.; Soyer, H.; Kirkpatrick, J.; Kavukcuoglu, K.; Pascanu, R.; and Hadsell, R. 2016. Progressive neural networks. *arXiv preprint arXiv:1606.04671*.

Saha, G.; Garg, I.; and Roy, K. 2021. Gradient projection memory for continual learning. *arXiv preprint arXiv:2103.09762*.

Verma, V. K.; Liang, K. J.; Mehta, N.; Rai, P.; and Carin, L. 2021. Efficient feature transformations for discriminative and generative continual learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 13865–13875.

Wang, F.; Lv, Y.; Zhu, M.; Ding, H.; and Han, J. 2024. XRF55: A Radio Frequency Dataset for Human Indoor Action Analysis. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 8(1): 1–34.

Wu, Y.; Chen, Y.; Wang, L.; Ye, Y.; Liu, Z.; Guo, Y.; and Fu, Y. 2019. Large scale incremental learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 374–382.

Xia, K.; Huang, J.; and Wang, H. 2020. LSTM-CNN architecture for human activity recognition. *IEEE Access*, 8: 56855–56866.

Yang, J.; Huang, H.; Zhou, Y.; Chen, X.; Xu, Y.; Yuan, S.; Zou, H.; Lu, C. X.; and Xie, L. 2024. Mm-fi: Multi-modal non-intrusive 4d human dataset for versatile wireless sensing. *Advances in Neural Information Processing Systems*, 36.

Yang, Y.; Zhou, D.-W.; Zhan, D.-C.; Xiong, H.; Jiang, Y.; and Yang, J. 2021. Cost-effective incremental deep model: Matching model capacity with the least sampling. *IEEE Transactions on Knowledge and Data Engineering*, 35(4): 3575–3588.

Yu, B. X.; Chang, J.; Liu, L.; Tian, Q.; and Chen, C. W. 2022. Towards a unified view on visual parameter-efficient transfer learning. *arXiv preprint arXiv:2210.00788*.

Zhang, Y.; He, F.; Wang, Y.; Wu, D.; and Yu, G. 2023. CSI-based cross-scene human activity recognition with incremental learning. *Neural Computing and Applications*, 35(17): 12415–12432.

Zhu, F.; Zhang, X.-Y.; Wang, C.; Yin, F.; and Liu, C.-L. 2021. Prototype augmentation and self-supervision for incremental learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5871–5880.

Zhu, K.; Zhai, W.; Cao, Y.; Luo, J.; and Zha, Z.-J. 2022. Self-sustaining representation expansion for non-exemplar class-incremental learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9296–9305.