DeepMultiple: A Deep Learning Model for RFID-based Multi-object Activity Recognition

Shunwen Shen[1], Lvqing Yang[1*], Sien Chen[2,3], Wensheng Dong[4], Bo Yu[4], Qingkai Wang[5,6]
[1]School of Informatics, Xiamen University, Xiamen, China
[2]School of Navigation, Jimei University, Xiamen, China
[3]School of Management, Xiamen University, Xiamen, China
[4]Zijin Zhixin(Xiamen)Technology Co., Ltd., Xiamen, China
[5]State Key Laboratory of Process Automation in Mining & Metallurgy, Beijing, China
[6]Beijing Key Laboratory of Process Automation in Mining & Metallurgy, Beijing, China
[*]The corresponding author email: lqyang@xmu.edu.cn

***Abstract*—Wireless sensing techniques for Human Activity Recognition (HAR) have been widely studied in recent years. At present, the research on HAR based on Radio Frequency Identification (RFID) is changing from the tag attachment method to the tag non-attachment method. Affected by multipath, the current solutions in tag non-attachment scenarios mainly focus on single-object activity recognition, which is not suitable for multi-object scenarios. To address these issues, we propose DeepMultiple, a novel tag non-attachment activity recognition model for multi-object. The model first preprocesses the raw signal with filter and phase calibration, then it applies dilated convolution in the frequency domain to extract multi-object activity features, finally *ProbSparse* is used to optimize the vanilla Transformer-based Encoder to enhance the activity recognition ability. We deployed a single reader and antenna for multi-object activity tracking during the experiments to reduce deployment difficulties. Extensive experimental results show that DeepMultiple can recognize ten types of multi-object activities with 98.12% precision under different challenging settings, which has excellent performance compared with several state-of-the-art methods.**

***Index Terms*—RFID, Multi-object, Human Activity Recognition, Deep Learning**

I. Introduction

With the development of deep learning technology, human activity recognition (HAR) has become one of the most important tasks in ubiquitous computing. It has attracted widespread attention from industry and academia. HAR based on cameras and portable wearable devices has problems such as high line-of-sight (LoS) requirements, unfriendly privacy protection(e.g. cameras), and real-time body attachment (e.g. wearable devices). RFID technology has become a new choice in the field of HAR as its low cost, small form size, and convenient deployment [1]. It has been widely used in various scenarios such as patient health monitoring [2], motion guidance in gyms [3], and activity monitoring on market shelves [4].

There are two tag placement methods for RFID-based HAR. (1) tag-attached method, tag-attached method employs RFID tags to attach to the human body (i.e. reference [5], reference [6]). (2) tag non-attachment method, tag non-attachment method apply multiple tags to fix in the environment as fixed references (i.e. reference [7]). The tag attachment method needs to attach RFID tags to specific objects or users to track their movements and infer their activities. This method has the inconvenience of equipment and privacy violations. Therefore, more and more researchers turn their attention to tag non-attachment method. But there are still the following challenges in the tag non-attachment scenario. First, it is easy to cause signal attenuation on the direct propagation link between the tag and the antenna due to environmental interference. So more antennas need to be brought in to reduce the multipath interference caused by the environment. Second, current research on RFID-based activity recognition mainly focuses on simple scenarios, that is, a person in an open environment. Since the interaction between multiple objects is more abundant, the interaction signals after backscattering are inevitably mixed. It will be more difficult to extend to multi-object activity recognition.

To address the above challenges, we propose DeepMultiple, a novel model for multi-object activity recognition in complex multipath environments. We use both RSSI and phase as input for getting more useful information from them. The model splits the preprocessed time-series data by sliding time windows and uses dilated convolution and selective attention mechanism to extract available features in complex multipath information to realize multi-object activity recognition. Besides, we deployed only a single antenna for activity identification during the experiment.

The contributions of this paper are summarized as follows.

- We propose a challenging multi-object activity recognition scenario without tag attachment. Compared to several state-of-the-art methods, experimental results demonstrate the superiority of our proposed model.
- To the best of the authors' knowledge, the DeepMultiple is the first model that applies dilated convolution in the RFID field to merge the spatial feature from different objects, and its validity is proven by ablation experiments.
- DeepMultiple optimizes the vanilla Transformer-based Encoder by *ProbSparse* attention mechanism, reduces the calculation of the parameters, and realizes multi-object HAR with only 2600 training samples.

DOI reference number: 10.18293/SEKE2023-138

II. Related Work

Tag-attached: Currently, there is a lot of work dedicated to human activity recognition based on RFID technology. Reference [10] attaches RFID tags to the user's back and recognizes the user's habitual sitting posture by establishing the correlation between the phase change of the tags and the sitting postures. TagBreathe [6] attaches RFID tags to users and measures the tiny movement of the human chest to monitor respiration. By attaching passive RFID tags on the dumbbells and leveraging the Doppler shift profile of the reflected backscatter signals, FEMO [5] provides an integrated free-weight exercise monitoring system. Rf-idraw [11] uses antenna array and beam steering technology to track the trajectory of the marked object. Tagoram [12] proposes a Differential Augmented Hologram (DAH) which will facilitate the instant tracking of the mobile RFID tag with millimeter accuracy. The above research is carried out around the tag attached to the detector, however, many activities do not directly interact with RFID-tagged objects. Empirical results have shown that RFID signals can be influenced by nearby human activities even if the objects are not moved [8].

Tag non-attachment: In recent years, more and more researchers have studied activity recognition in tag non-attachment scenarios. RFIPad [7] enables in-air handwriting without tags attached. RF-Care [13] aims to use passive RFID arrays to establish a tag non-attachment activity recognition for older people. Tagfree [8] is a pioneering work that aims to distinguish and use useful features contained in complex multipath information. Although the above research is based on tag non-attachment, more devices need to be used for accurate identification, which increases deployment costs. And these methods do not have the ability of multi-object recognition.

III. PRELIMINARIES

In this section, we will introduce the basic theory and data preprocessing methods in RFID.

A. RFID Communication Mechanism

Received Signal Strength Indicator (RSSI) . The wireless signal will be divided into multiple signals during the propagation process. After the reader sends signals to the surrounding environment, multiple signal signals are superimposed and returned to the reader. The signal strength received by the reader can be modeled as:

$$S = \sum_{i=1}^{N} h \cdot |F_i| \, e^{j\theta} \tag{1}$$

where h represents the attenuation coefficient in the signal propagation process. F_i is the amplitude of the i-th communication path signal, and θ is the phase of RFID signal. Due to the influence of multi-object interaction and multipath effects in the environment, the signal S consists of three parts: the S_{tag} returned directly by the tag, the S_{env} returned by the environment and the $S_{objects}$ returned by the multi-object interaction as shown in Fig.3.

$$S = S_{tag} + S_{env} + S_{objects} \tag{2}$$

From the above two formulas, we can conclude that in a complex multipath environment, the signal received by the reader is mixed with multiple signals, and it will be affected by distance and attenuation coefficient.

Phase. For a single-reader RFID propagation environment, the phase information of the signal received from the k-th tag provided by the reader is $\theta = mod\,(\theta_p + \theta_a + \theta_r + \theta_t, 2\pi)$, $\theta_p = 2\pi d/\lambda$. d is the propagation distance of RFID signal, λ represents the wavelength of the RFID signal. $\theta_a, \theta_r, \theta_t$ represent the phase jump caused by the antenna transmission circuit, receiver circuit and tag reflection characteristics.

B. Data Preprocessing

1) Phase Calibration: As shown in Fig.1(a) and Fig.1(b), we find that RSSI is less sensitive to human activity, but still detects fluctuations caused by activity, while the phase is very sensitive to information, but not accurate. The main reason for the inaccurate raw phase is caused by the frequency hopping mechanism. Previous research has shown that frequency hopping can lead to significant phase shifts due to the phase difference of the oscillators and the non-uniform frequency response of the antenna [19]. And according to PRELIMINARIES, we know that the phase shift will also be affected by transmission, receive, and reflection links.

We use phase unwrapping [15] and phase smoothing to eliminate this phase shift. The smoothing algorithm works by collecting an initial phase measurement, which takes about 10 seconds for a stationary tag. It can be written as:

$$\theta(t) = \theta_j(t) - \tilde{\theta_j} + \tilde{\theta_d} \tag{3}$$

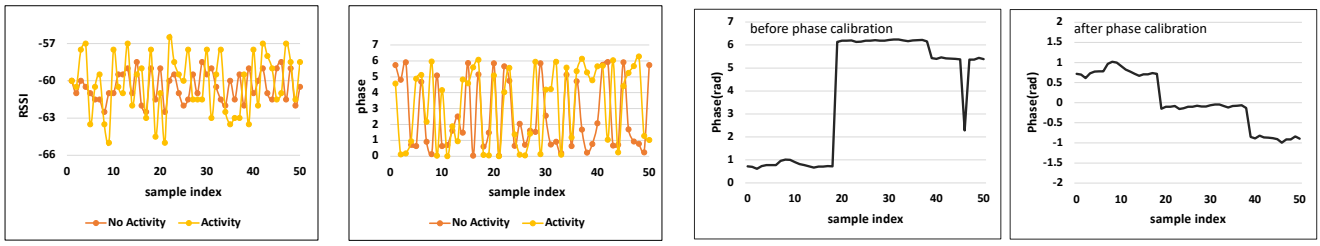
where $\theta_j(t)$ denote the measured phase at frequency f_j at time t. $\tilde{\theta_j}$ and $\tilde{\theta_d}$ represent the median values of phases measured in the last 10 seconds at frequency f_j and common frequency f_d (default to 922.625MHZ in this work). Fig.1(c) shows the results after phase unwrapping and phase smoothing.

2) Data Splitting and Resampling: We first split the collected long signal data into samples with width w (w = 5s in this work) and form the training set by them. Then, these samples are further divided into n (n = 10 in this work) non-overlapping timesteps $\mathbf{X}$ with time interval width $\boldsymbol{\tau}$ ($\boldsymbol{\tau}$ = 0.5s in this work), $\mathbf{X} = \{X_1, ..., X_t, ..., X_n\}$, X_t represents the data collected by k tags at timestep t, which can be further divided into $X_t = \{x_{t1}, x_{t2}, ..., x_{tk}\}$.

Due to the transmission characteristics of the RF signal, the RSSI and phase from each tag in the scenario cannot be read equally. We resample the input data to the same dimensions using linear interpolation.

3) Fast Fourier Transform: The raw RFID signals are a mixture of objects, too noisy to be directly understood and used. We use Fast Fourier Transform (FFT) to convert the time-domain data to the frequency domain for distinguishing the activity features from multiple objects.

4) Data Filtering: To select a more suitable filter in this scenario, our paper uses Gaussian Filter, Kalman Filter, Mean Filter, Median Filter, and Hampel Filter to process the raw



(a) RSSI changes in the scene (b) Phase changes in the scene (c) Illustration of phase calibration

Fig. 1. Data preprocessing module display diagram

RFID signal. In the experimental section, we further compare the performance of the model with different filters.

IV. Model

In this section, we give a detailed description of the DeepMultiple. The framework of DeepMultiple is shown in Fig.2.

A. Individual ConvLayer

DeepMultiple is arranged according to *n* timesteps as input. Since the structures of Individual ConvLayer at each timestep are the same, we focus on a single timestep with input $X_t = \{x_{t1}, x_{t2}, ..., x_{tk}\}$. Recall that x_{ti} represents the signals from the *i*-th tag at timestep *t* and the shape of x_{ti} is $d \times 2f$, where d present the tag measurement dimension, i.e. RSSI and phase, f is the dimension of frequency domain. For each timestep, x_{ti} is first fed into the dilated CNN with shape (1, conv1) to extract the multi-scale activity features from different objects in the frequency domain. Then it uses a 2d conv with shape (d, conv2) to merge features from different tags and a 2d conv with shape (1, conv3) to further learn the high-level relationship, with the output $v_{ti}^{(1)}$. After each convolution operation, DeepMultiple uses ReLu as the activation function and applies Batch Normalization to reduce internal covariate shift and vanishing gradients.

B. Flatten and Merge Layer

In Flatten and Merge Layer, we flatten $v_{ti}^{(1)}$ in different channels into $v_{ti}^{(2)}$ and concat *k* tags vector $\left\{v_{t1}^{(2)}, v_{t2}^{(2)}, ..., v_{tk}^{(2)}\right\}$ into a *k*-row matrix V_t, then we use conv2d with shape (k, conv4) to learn the intrinsic interactions among all *k* tags to generate the matrix V_t', furthermore 2d filters with (1, conv5) and (1, conv6) are applied to learn the high-level relationships, last we flatten the result into v_t' as the input to Transformer-based Encoder Variant Layer. Again, after each convolution layer, Batch Normalization and a ReLu activation are performed, and a MaxPool2d with stride = 2 is applied to compress the dimensions in the last convolution layer.

C. Transformer-based Encoder Variant Layer

A large amount of training data should be used in transformer due to its complex structure, which is diffcult to achieve in RFID filed. Previous research like [14] used simple RNN or GRU as the backbone network, but experiments have verified that these methods are not suitable for multi-object recognition without tags attachment. To this end, we optimize the vanilla Transformer-based Encoder structure and use sparse *ProbSparse* attention to effectively reduce the computational complexity and improve the recognition accuracy.

Since the vanilla self-attention mechanism uses atom operation, i.e. scaled dot product, causes the time complexity and memory usage per layer to be $\mathcal{O}\left(L^2\right)$.The use of the vanilla self-attention mechanism entails the computation of the scaled dot product of the input without discrimination, which will amplify the influence of noise in the data, hinder the performance of the model, and increase computational complexity. The *ProbSparse* self-attention mechanism proposed in [9] can be a good solution to this problem. It has shown that the distribution of self-attention probability is potentially sparse and achieves the $\mathcal{O}\left(L \log L\right)$ time complexity and $\mathcal{O}\left(L \log L\right)$ memory usage on dependency alignments. The *ProbSparse* self-attention mechanism is shown below.

$$\mathcal{A}\left(\mathrm{Q}, \mathrm{K}, \mathrm{V}\right) = \operatorname{softmax}\left(\frac{\overline{\mathrm{Q}}\mathrm{K}^{\top}}{\sqrt{\mathrm{d_k}}}\right)\mathrm{V} \tag{4}$$

where $\overline{Q}$ is a sparse matrix of the same size as q and it only contains the **Top-*u*** queries under the sparsity measurement $\overline{M}\left(q, K\right)$ controlled by a constant sampling factor *c*, we set $u = c \cdot \ln L_Q$, and $\mathcal{A}$ represents kernel smoother based on probability distribution. $\overline{M}\left(q, K\right)$ and $\mathcal{A}$ can be described as:

$$\overline{\mathrm{M}}\left(\mathrm{q_i}, \mathrm{K}\right) = \max_{\mathrm{j}}\left\{\frac{\mathrm{q_i}\mathrm{k_j}^{\top}}{\sqrt{\mathrm{d_k}}}\right\} - \frac{1}{\mathrm{L_K}}\sum_{\mathrm{j}=1}^{\mathrm{L_K}}\frac{\mathrm{q_i}\mathrm{k_j}^{\top}}{\sqrt{\mathrm{d_k}}} \tag{5}$$

Under the long tail distribution, we randomly draw sample $U = L_K \ln L_Q$ dot-product pairs to get $\overline{M}\left(q_i, K\right)$. Then, we select **Top-*u*** from them as $\overline{Q}$, so the *ProbSparse* self-attention time complexity and space complexity are $\mathcal{O}\left(L \log L\right)$.

$$\mathcal{A}\left(\mathrm{q_i}, \mathrm{K}, \mathrm{V}\right) = \sum_{\mathrm{j}}\frac{\mathrm{k}\left(\mathrm{q_i}, \mathrm{k_j}\right)}{\sum_{\mathrm{l}}\left(\mathrm{q_i}, \mathrm{k_l}\right)}\mathrm{v_j} = \mathbb{E}_{\mathrm{p}(\mathrm{k_j}|\mathrm{q_i})}\left[\mathrm{v_j}\right] \tag{6}$$

where q_i stands for the *i*-th row in Q, $p\left(k_j \mid q_i\right) = k\left(q_i, k_l\right) / \sum_l k\left(q_i, k_l\right)$ and $k\left(q_i, k_l\right)$ selects the asymmetric exponential kernel $\exp\left(q_i k_j^{\top} / \sqrt{d}\right)$. The self-attention and values are combined on the basis of the calculation of probabilities $p\left(k_j \mid q_i\right)$ to obtain the output.

We explain the Transformer-based Encoder Variant Layer with input v_t' at timestep *t* and start by applying the positional embedding [18] to introduce a concept of relative

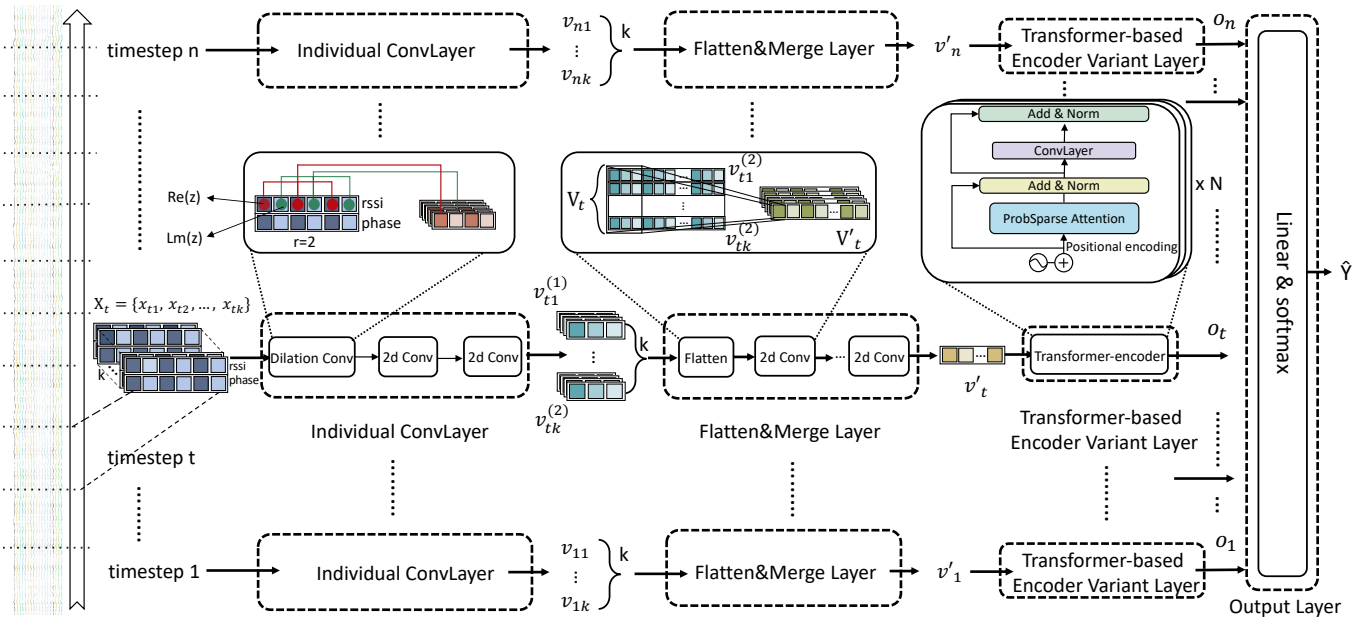


Fig. 2. Illustrating the overall structure of DeepMultiple, where Re(z) and Lm(z) represent the real part and imaginary part respectively

order between the features extracted at different timesteps. Then, we multiple v_t' with three different learnable matrices W^Q, W^K, W^V to get the query, key, value matrices Q, K, V. The **Top-*u*** important queries are selected according to (5) and the attention score is obtained by (6). Then we integrate attention score and input by skip-connection. Last, Layer Normalization is applied and then the normalized data are passed into the feedforward output layer to obtain the output $O = \{o_1, o_2, ..., o_t\}$.

D. Output Layer

The output after Transformer-based Encoder variant Layer is $O = \{o_1, o_2, ..., o_t\}$, we need a prediction activity $\hat{y}$, so a Linear layer, $\hat{y}_c = OA^\top + b$, is used to map O to $\hat{y}_c$ (where $c \subset C$, C is the set of all classes) and then normalize $\hat{y}_c$ by *softmax*, we select the max probability as follows.

$$P = \underset{c \in C}{argmax} \left(softmax \left(\hat{y}_c \right) \right) \tag{7}$$

We opt to use cross-entropy loss. It can be formulated as:

$$\mathcal{L} = -\sum_{c=1}^{N} y_c \log \left(P_c \right) \tag{8}$$

where y_c is the indicator variable, P_c is the probability that the predicted result belongs to class c, and N is the number of class categories.

V. Experiment

A. DataSet Description

We conduct extensive experiments based on the dataset collected by five volunteers (three males and two females) in two typical indoor environments, a laboratory and an empty room to complex and simple multipath environments, respectively. In each environment, we deploy a single reader and antenna then we attach a 3x3 tag array to a wall 1.5-2m above the ground. In our experiments, volunteers perform activities as shown in Fig.3. We tested ten activity scenarios of two people, as shown in Fig.4.

The hardware components of our experiment include an Impinj R700 reader, equipped with an ultrahigh-frequency UHF2599 antenna and SMARTRAC DogBone RFID tags.

We collected 3250 signal samples and name them RFAC DataSet, of which 80% were used as training sets, and 20% were used as test sets. To compare the impact of different filtering algorithms on model performance, we processed the raw data by different filters and constructed RFAC-Hampel, RFAC-Gaussian, RFAC-Kalman, RFAC-Median, and RFAC-Mean. In the subsequent experiments, the pre-processed dataset is applied to the baseline for performance comparison.

B. Experimental Details

In this section, we introduce the baseline, hyper-parameter tuning, and evaluation metrics used in this experiment.

Baseline: We chose five deep learning models used in the HAR field as a comparison, including the state-of-the-art models. For all the compared models, we only made minor adjustments to the shape of the input, and the model structures were implemented as provided by the authors. The CNN-stacked model, DeepConv [16], CNN-GRU, TagFree [8] and AttnSense [17] are used to compare the performance with DeepMultiple.

Hyper-parameter tuning: We conducted a grid search for learning rate and weight decay, the learning search range was $\{1e^{-2}, 1e^{-3}, 1e^{-4}, 1e^{-5}\}$, and the learning rate is finally set

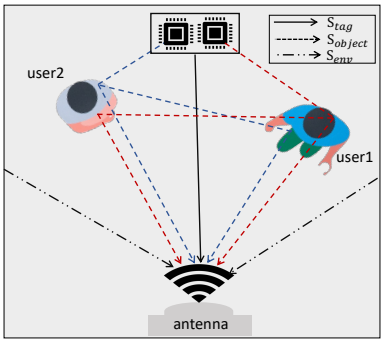


Fig. 3. Multi-object recognition scene

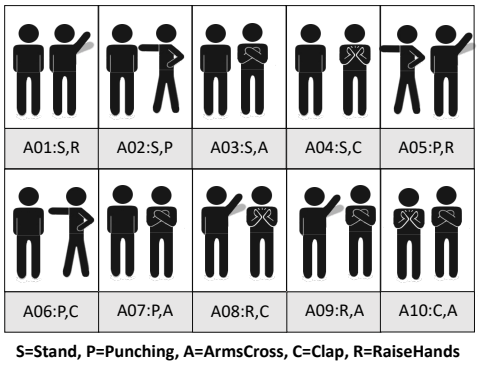


Fig. 4. Activity classification chart

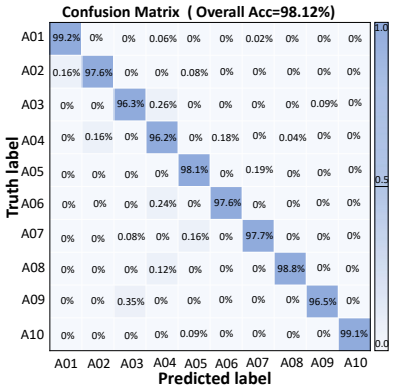


Fig. 5. Multi-object confusion matrix results

to $1e^{-3}$, the weight decay search range was $\{0.01, 0.1, 1\}$, and the weight decay was finally selected to be 0.1. During the training, we used the early stopping to train and set the batch size to 64 then trained a total of 500 epochs. We use Adam optimizer with a default value and normalize the input by zero-mean.

Metrics: In this work, we use Weighted-F1 as the metric in the evaluation and the results are averaged over the 5 runs. The formula of Weighted-F1 is described as follows.

$$\text{Weighted} - \text{F1} = \sum_{\text{i=1}}^{\text{C}} \text{w}^{(\text{i})} \frac{2 \cdot \text{Precision}^{(\text{i})} \cdot \text{Recall}^{(\text{i})}}{\text{Precision}^{(\text{i})} + \text{Recall}^{(\text{i})}} \quad (9)$$

where for a given class i, $w^{(i)}$ represents the proportion of this class in the total sample, $Precision^{(i)}$, $Recall^{(i)}$ are the precision and recall of the i-th class respectively.

C. Numerical Analysis

1) Performance of our model: As shown in Fig.5, the Confusion Matrix shows the overall performance of DeepMultiple with the kalman filter at an interval width of 0.5s. The overall accuracy rate is 98.12% and the accuracy rate of all ten activities is above 96%. It can be found that our model can extract important features well in single reader-antenna tag non-attachment scenarios.

2) Impact of time interval width and filter preprocessing: In order to compare the impact of the time interval and filter preprocessing, we evaluated the performance of the model with different filters at different time interval widths. From Fig. 6, we can find that the model performs best when the time interval width is set to 0.5s. The highest Weighted-F1 score is 0.981 with the Kalman filter and the lowest Weighted-F1 score is 0.967 with the mean filter under this setting. In the subsequent experiments, the default is to compare performance with the Kalman filter at an interval width of 0.5s.

3) Impact of phase calibration: In order to verify that phase unwrapping and phase smoothing can also improve the accuracy of activity recognition, we further evaluate the performance of the model with calibration and no-calibration. As shown in Fig.7, we can find that our model has achieved higher recognition accuracy with calibration, which directly demonstrates the effectiveness of our phase calibration.

TABLE I
WEIGHTED-F1 SCORES ON DIFFERENT ALGORITHMS

Model	Dataset					
	D1	D2	D3	D4	D5	D6
DeepMultiple	**0.941**	**0.976**	**0.968**	**0.981**	**0.973**	**0.962**
CNN	0.448	0.728	0.742	0.756	0.770	0.759
DeepConv [16]	0.509	0.790	0.834	0.867	0.796	0.851
CNN-GRU	0.520	0.848	0.892	0.900	0.890	0.906
TagFree [8]	0.562	0.829	0.890	0.909	0.878	0.889
AttnSense [17]	0.659	0.848	0.861	0.895	0.836	0.871

4) Compared Algorithms: In this section, DeepMultiple is compared with the baseline algorithm under the RFAC (D1), RFAC-Hampel (D2), RFAC-Gauss (D3), RFAC-Kalman (D4), RFAC-Median (D5) and RFAC-Mean (D6).

Through Table I, we find that the recognition accuracy of all models was improved after the filtering process, which proves the effectiveness of the filtering algorithm. In addition, we find that DeepMultiple always achieves the highest Weighted-F1 score regardless of filtering, which proves its strong robustness and generalization ability. In contrast, baselines without filtering have poor performance and cannot perform multi-object activity recognition in a complex multipath environment. There is no doubt that the performance of DeepMultiple is significantly better than the latest method in this field.

5) Ablation Study: To verify that the dilated convolution module and *ProbSparse* self-attention mechanism proposed by our model contribute to the effectiveness of activity recognition, we have performed two variants of our model, DeepMultiple_CConv using vanilla convolutions instead of dilated convolutions and DeepMultiple_CAttn using vanilla self-attention mechanism instead of *ProbSparse* self-attention. The result is shown in Fig.8.

By comparing the experimental performance of DeepMultiple, DeepMultiple_CConv and DeepMultiple_CAttn, it can be seen that the accuracy of the model using the dilated convolution has been improved in all ten activities. The overall recognition accuracy increased by 4.7%. The reason for this phenomenon is that vanilla convolution causes feature frag-

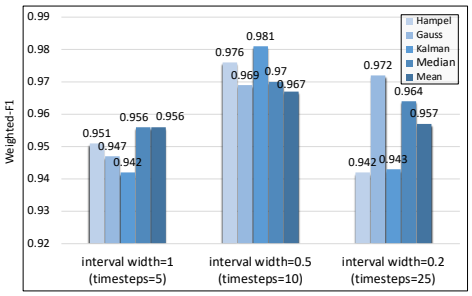


Fig. 6. Performance with different filter at differents time interval widths

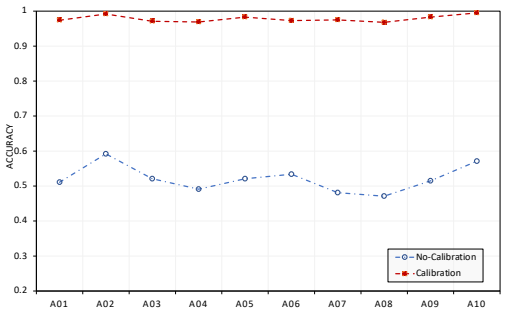


Fig. 7. The model performance with different filters

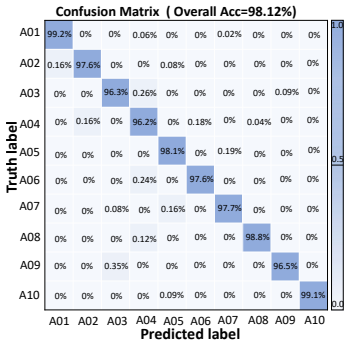


Fig. 8. Confusion matrix results

mentation when extracting features in the frequency domain, which can be solved by dilated convolution. Compared to the vanilla self-attention, using *ProbSparse* self-attention brings a 27.5% increase in global recognition accuracy. It can be concluded that the query sparsity of *ProbSparse* self-attention can effectively improve the performance of model recognition and prevent overfitting.

VI. Conclusion

In this paper, we propose DeepMultiple for multi-object activity recognition in complex multi-path environments. DeepMultiple not only applies the dilated convolution to feature extraction in the frequency domain, but it also reduces the calculation of the parameters with the selective *ProbSparse* attention and improves the recognition accuracy. Experimental results show that the accuracy of this model can reach 98.12%, which is superior to the state-of-the-art model in this field.

Acknowledgment

This paper is supported by the 2021 Fujian Foreign Cooperation Project(No. 2021I0001): Research on Human Behavior Recognition Based on RFID and Deep Learning; Horizontal project (Co-construction platform), 2023 Project of Xiamen University: Joint Laboratory of Public Safety and Artificial Intelligence(20233160C003); State Key Laboratory of Process Automation in Mining & Metallurgy, Beijing Key Laboratory of Process Automation in Mining & Metallurgy(No. BGRIMM-KZSKL-2022-14): Research and application of mine operator positioning based on RFID and deep learning; National Key R&D Program of China-Sub-project of Major Natural Disaster Monitoring, Early Warning and Prevention (No. 2020YFC1522604): Research on key technologies of comprehensive information application platform for cultural relic safety based on big data technology.

References

[1] J. Liu, M. Chen, S. Chen, Q. Pan, and L. Chen, "Tag-compass: Determining the spatial direction of an object with small dimensions," in *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*. IEEE, 2017, pp. 1–9.

[2] P. H. Frisch, "Rfid in today's intelligent hospital enhancing patient care & optimizing hospital operations," in *2019 IEEE international conference on rfid technology and applications (RFID-TA)*. IEEE, 2019, pp. 458–463.

[3] Z. Liu, X. Liu, and K. Li, "Deeper exercise monitoring for smart gym using fused rfid and cv data," in *IEEE INFOCOM 2020-IEEE Conference on Computer Communications*. IEEE, 2020, pp. 11–19.

[4] K. Ali, A. X. Liu, E. Chai, and K. Sundaresan, "Monitoring browsing activity of customers in retail stores via rfid imaging," *IEEE Transactions on Mobile Computing*, vol. 21, no. 3, pp. 1034–1048, 2020.

[5] H. Ding, L. Shangguan, Z. Yang, J. Han, Z. Zhou, P. Yang, W. Xi, and J. Zhao, "Femo: A platform for free-weight exercise monitoring with rfids," in *Proceedings of the 13th ACM conference on embedded networked sensor systems*, 2015, pp. 141–154.

[6] Y. Hou, Y. Wang, and Y. Zheng, "Tagbreathe: Monitor breathing with commodity rfid systems," in *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 2017, pp. 404–413.

[7] H. Ding, C. Qian, J. Han, G. Wang, W. Xi, K. Zhao, and J. Zhao, "Rfipad: Enabling cost-efficient and device-free in-air handwriting using passive tags," in *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 2017, pp. 447–457.

[8] X. Fan, W. Gong, and J. Liu, "Tagfree activity identification with rfids," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 1, pp. 1–23, 2018.

[9] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, "Informer: Beyond efficient transformer for long sequence time-series forecasting," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 12, 2021, pp. 11 106–11 115.

[10] L. Feng, Z. Li, and C. Liu, "Are you sitting right?-sitting posture recognition using rf signals," in *2019 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM)*. IEEE, 2019, pp. 1–6.

[11] J. Wang, D. Vasisht, and D. Katabi, "Rf-idraw: Virtual touch screen in the air using rf signals," *ACM SIGCOMM Computer Communication Review*, vol. 44, no. 4, pp. 235–246, 2014.

[12] L. Yang, Y. Chen, X.-Y. Li, C. Xiao, M. Li, and Y. Liu, "Tagoram: Real-time tracking of mobile rfid tags to high precision using cots devices," in *Proceedings of the 20th annual international conference on Mobile computing and networking*, 2014, pp. 237–248.

[13] Z. Xia, J. Liu, and S. Guo, "Rf-care: Rfid-based human pose estimation for nursing-care applications," in *2021 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. IEEE, 2021, pp. 1384–1389.

[14] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, "Deep learning for sensor-based activity recognition: A survey," *Pattern recognition letters*, vol. 119, pp. 3–11, 2019.

[15] K. Itoh, "Analysis of the phase unwrapping algorithm," *Applied optics*, vol. 21, no. 14, pp. 2470–2470, 1982.

[16] F. J. Ordóñez and D. Roggen, "Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition," *Sensors*, vol. 16, no. 1, p. 115, 2016.

[17] H. Ma, W. Li, X. Zhang, S. Gao, and S. Lu, "Attnsense: Multi-level attention mechanism for multimodal human activity recognition." in *IJCAI*, 2019, pp. 3109–3115.

[18] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[19] T. Wei and X. Zhang, "Gyro in the air: tracking 3d orientation of battery-less internet-of-things," in *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*, 2016, pp. 55–68.