remote sensing

MDPI

Article

A Mutiscale Residual Attention Network for Multitask Learning of Human Activity Using Radar Micro-Doppler Signatures

Yuan He, **Xinyu Li** * and **Xiaojun Jing**

Key Laboratory of Trustworthy Distributed Computing and Service (BUPT), Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China; yuanhe@bupt.edu.cn (Y.H.); jxiaojun@bupt.edu.cn (X.J.)
* Correspondence: lixinyu@bupt.edu.cn

Received: 19 September 2019; Accepted: 30 October 2019; Published: 4 November 2019

Abstract: Short-range radar has become one of the latest sensor technologies for the Internet of Things (IoT), and it plays an increasingly vital role in IoT applications. As the essential task for various smart-sensing applications, radar-based human activity recognition and person identification have received more attention due to radar's robustness to the environment and low power consumption. Activity recognition and person identification are generally treated as separate problems. However, designing different networks for these two tasks brings a high computational complexity and wastes of resources to some extent. Furthermore, there are some correlations in activity recognition and person identification tasks. In this work, we propose a multiscale residual attention network (*MRA-Net*) for joint activity recognition and person identification with radar micro-Doppler signatures. A fine-grained loss weight learning (FLWL) mechanism is presented for elaborating a multitask loss to optimize *MRA-Net*. In addition, we construct a new radar micro-Doppler dataset with dual labels of activity and identity. With the proposed model trained on this dataset, we demonstrate that our method achieves the state-of-the-art performance in both radar-based activity recognition and person identification tasks. The impact of the FLWL mechanism was further investigated, and ablation studies of the efficacy of each component in *MRA-Net* were also conducted.

Keywords: smart sensing; human activity recognition; person identification; multitask learning; radar micro-Doppler signatures

1. Introduction

With the great development of the Internet of Things (IoT), short-range low-power radar sensors for smart sensing are attracting increasingly more interest mainly due to the advantages of robustness to weather and lighting conditions, penetrability of obstacles, low power consumption and protecting visual privacy [1,2]. IoT is treated as the future trend in the global technological development after the Internet, and it connects all things to the Internet for information sensing and exchange, and data computing [3]. For the smart sensing function of IoT, human activity recognition and person identification are crucial steps, and they have been widely adopted in indoor real-time positioning, activity monitoring, elderly fall detection, and so forth.

Person identification and activity classification have both been investigated in prior work [4]. Most of these systems are based on optical cameras for collecting information. However, optical devices have many limitations such as sensitivity to light or weather conditions and a high demand for computational resources, which hinders the application of such devices for human activity recognition and identification, especially with portable devices that have limited computational capabilities. Radar is considered to be a powerful approach to overcome these drawbacks. Furthermore, radar can be applied to more scenarios and effectively protect visual privacy.

Recently, radar-based person identification and activity recognition have attracted increasing attention. When a human target is moving in the line of sight of a radar, the returned signals are modulated by the movement. In addition to the main Doppler shift caused by the human torso, the movements of body parts form micro-Doppler (MD) shifts, which is commonly called a target MD signature. Generally, radar MD signatures are both target and action specific, and can hence be used to recognize targets and classify activities. For example, Vandersmissen et al. [5] demonstrated that MD features are able to characterize individual humans in realistic scenarios, which makes radar-based person identification possible. Handcrafted features such as extreme frequency ratio, torso frequency, period of motion and total Doppler bandwidth [6] are extracted from MD spectrograms to characterize human motions. However, the activity recognition and person identification solutions based on hand-crafted features are not reliable enough since Doppler and MD signatures are individual and action specific. In addition, professional knowledge is indispensable during the manual feature extraction process. Hence, automated and optimized feature extraction approaches are desired.

Recent trends in deep learning have led to a renewed interest in radar-based person identification and activity recognition due to its capability of automatically extracting features and encouraging precision. Furthermore, with the development of GPUs, it is possible to process vast amounts of data in a limited time via parallel computing techniques. Convolutional neural networks (CNNs) play a vital role in numerous deep-learning-based systems, and they have been successfully applied in the areas of object detection, image classification and so forth [7,8]. Although not as intuitive as natural images, an MD spectrogram, which is a two-dimensional representation of radar signals, could also be analyzed as an image. In this circumstance, a CNN that is adept at learning embedding from two-dimensional images is often utilized to process radar MD spectrograms [5,9].

Although closely related, activity recognition and person identification are generally treated as separate problems. It has recently been demonstrated that learning correlated tasks simultaneously can enhance the performance of individual tasks [10,11]. Joint activity recognition and person identification has two advantages. (1) Sharing the model between the two tasks accelerates the learning and converging process compared with applying it to a single task. (2) Multiple labels supply more information about the dataset, which is capable of regularizing the network during training. Furthermore, the computational complexity can be improved by sharing the feature extractor between the activity recognition and person identification tasks. Motivated by these advantages, we go one step beyond separate person identification and activity recognition by proposing a novel CNN-based multitask framework to complete the two tasks simultaneously. Multitask learning (MTL) [12] aims at leveraging the relatedness among the tasks and learning embedding of each task synchronously to improve the generalization performance of the main task or all tasks. In this paper, our goal is to enhance both tasks with an MTL-based deep neural network.

We propose a multiscale residual attention network (*MRA-Net*) for joint person identification and activity recognition with radar. As shown in Figure 1, *MRA-Net* is composed of two parts: feature extractor and multitask classifier. *MRA-Net* firstly extracts a common embedding from the MD signature, and then the embedding is input into fully connected (FC) layers to perform the classification of each task. In the CNN-based feature extractor, two scales of convolutional kernels are applied for extracting different-grained features from the input. The shared features entangle attributes for both activity recognition and person identification. Furthermore, the residual attention mechanism [13] is adopted in the feature extractor to facilitate the feature learning process. Finally, a fine-grained loss weight learning (FLWL) mechanism is proposed in contrast to the methods in previous work that either treat each task equally [14] or obtain the loss weights by greedy search [15].

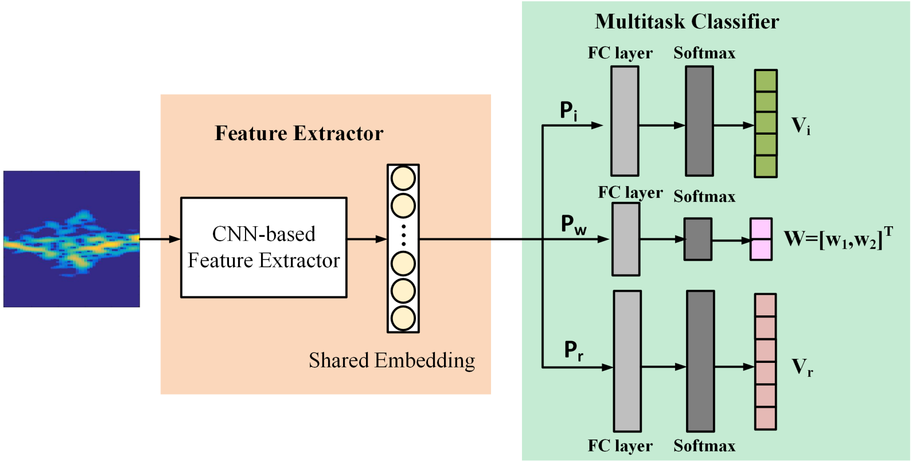

Figure 1. The CNN-based framework of the proposed *MRA-Net*. The model is composed of two parts: feature extractor and multitask classifier. In the multitask classifier part, there are three branches: activity recognition branch P_r, person identification branch P_i and FLWL branch P_w. There is a FC layer and a Softmax layer in P_i and P_w, respectively, and V_i and V_r denote the corresponding output vectors that are utilized for the final classifications of the two tasks. W denotes the output vector that is utilized for the automatic loss weight learning.

In summary, our contributions mainly include the following three aspects.

- A novel multiscale residual attention network, named as *MRA-Net*, is proposed to jointly perform human identification and activity recognition tasks based on radar MD spectrograms. *MRA-Net* outperforms the state-of-the-art methods for both tasks by jointly recognizing human activities and identifying persons.
- A fine-grained loss weight learning mechanism is proposed to automatically search for proper loss weights rather than equalizing or manually tuning the loss weight of each task.
- Extensive experiments with state-of-the-art results have validated the feasibility of radar-based joint activity recognition and person identification, as well as the effectiveness of *MRA-Net* towards this issue.

The rest of this paper is organized as follows. Section 2 briefly introduces the related work of person identification, human activity recognition and multitask learning. Section 3 details the proposed deep multitask learning network. A measured radar micro-Doppler signature dataset is described in Section 4. The performance metrics and implementation details are presented in Section 5. Experimental results and analysis are provided in Section 6. Section 7 concludes this paper.

2. Related Work

2.1. Person Identification

Person identification is a key technology in various fields, for example, terrorist attack preventing, criminal seeking and defense. In prior work, person identification always depends on biological or vision-based features. Liang et al. [16] proposed a method that combines inaccurately estimated human pose and inferred confidence metric for cross-view person identification. A fingerprint recognition system that applies information from multiple feature extractors is built for person identification [17]. Recently, WIFI has been applied to person identification [18]. Meanwhile, radar-based person identification has attracted more attention. Compared with other sensors, radar is more stable in weak light and bad weather conditions. It is penetrable and able to protect visual privacy. In addition, radar systems do not need any tag attached to the human body. Vandersmissen et al. [5] utilized a low-power FMCW radar for indoor person identification based on gait characteristics. Cao et al. [19] applied radar MD signatures for person identification with a deep CNN.

2.2. Activity Recognition

There are two main categories for human activity recognition methods [20]: vision-based and sensor-based methods. Vision-based methods take advantage of the high resolution of optical sensors and the rapidly evolving computer vision techniques, and fruitful results have been obtained. Zhang et al. [21] proposed an active action proposal model that aims to find actions through continuously adjusting the temporal bounds in a self-adaptive way. A reinforcement learning algorithm is also adopted in this work. Wearable sensors are another commonly used tool for activity recognition and have achieved high accuracies [22]. However, this method requires subjects to wear sensors in a strict way to ensure correct operations, and the sensors are always on the body, which makes people uncomfortable and burdened. Following the work in [23], radar-based activity recognition approaches have been gradually receiving attention. Seyfioğlu et al. [9] proposed a deep convolutional autoencoder architecture for similar aided and unaided human activities with MD signatures. Le et al. [24] developed a Bayesian-optimized CNN model for human motion classification with a Doppler radar.

2.3. Multitask Learning

MTL is an inductive transfer mechanism that aims to train tasks in parallel and learn sufficiently generalized representations. MTL can be combined with various learning algorithms and is applied in many fields. HyperFace [14], which is a multitask framework, is proposed for face detection, landmark localization, pose estimation, and gender recognition. The idea of MTL is also incorporated in object detection algorithms [25]. In addition, Miranda-Correa et al. [26] applied a multitask cascaded network for predicting affect, personality, mood and social context with EEG signals.

Prior MTL approaches either treat the loss weights equally or utilize a manual greedy search to train the model. Kendall et al. [27] utilized homoscedastic uncertainty, which is explained as task-dependent weighting, to combine multiple loss functions. The efficiency of the proposed method in learning scene geometry and semantics has been demonstrated with three tasks. Yin et al. [28] proposed a dynamic-weighting scheme to learn the loss weights of each side task automatically. In our work, we propose an FLWL mechanism that makes *MRA-Net* automatically learn weights for both tasks.

3. Multiscale Residual Attention Network

In this paper, we propose *MRA-Net*, which is a multiscale residual attention network for joint radar-based person identification and activity recognition. The unified network is learned end-to-end and is optimized with a multitask loss. The feature extractor part of *MRA-Net* is illustrated in Figure 2. In this part, multiscale learning and residual attention learning mechanisms are employed to facilitate the feature extracting process. Specifically, the feature extractor is composed of three blocks, and there are three branches in every block: coarse-scale learning branch, fine-scale learning branch and residual attention branch. For the multitask classifier, we propose an FLWL mechanism to automatically set the loss weights for MTL.

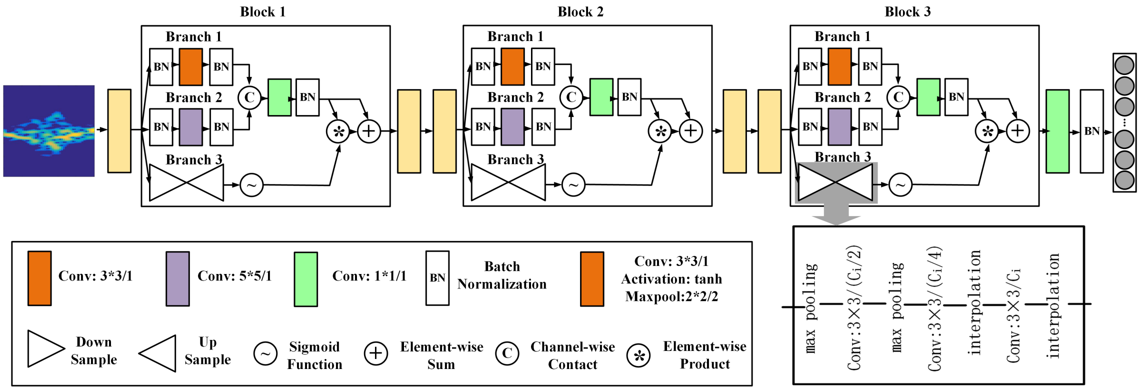

Figure 2. The feature extractor part of *MRA-Net*, which is composed of three blocks. There are three branches in each block: coarse-scale learning branch, fine-scale learning branch and residual attention learning branch. All branches are able to facilitate the feature learning process. In the CNN-based feature extractor, the convolution operation with a kernel size of 3 × 3 and a stride of 1 is denoted as 3 × 3/1, and the pooling operation is denoted as the same way.

3.1. Multiscale Learning

Multiscale learning mechanism is able to extract features from various granularities and learn more efficient representations [29]. In *MRA-Net*, we apply two types of convolution kernels with different receptive fields: the 3 × 3 kernel is used for fine-scale learning in Branch 1, while the 5 × 5 kernel is utilized for coarse-scale learning in Branch 2 (see Figure 2). The intuition of employing multiscale convolution kernels is that the MD characteristics of different activities vary. For example, the MD effect of "box" is more delicate, while the MD effect of "walk" is stronger. The receptive fields of different convolution kernels match the MD features of different scales, as illustrated in Figure 3. Due to the complementary information between different scales of convolution kernel, multiscale learning can significantly improve the performance of *MRA-Net*.

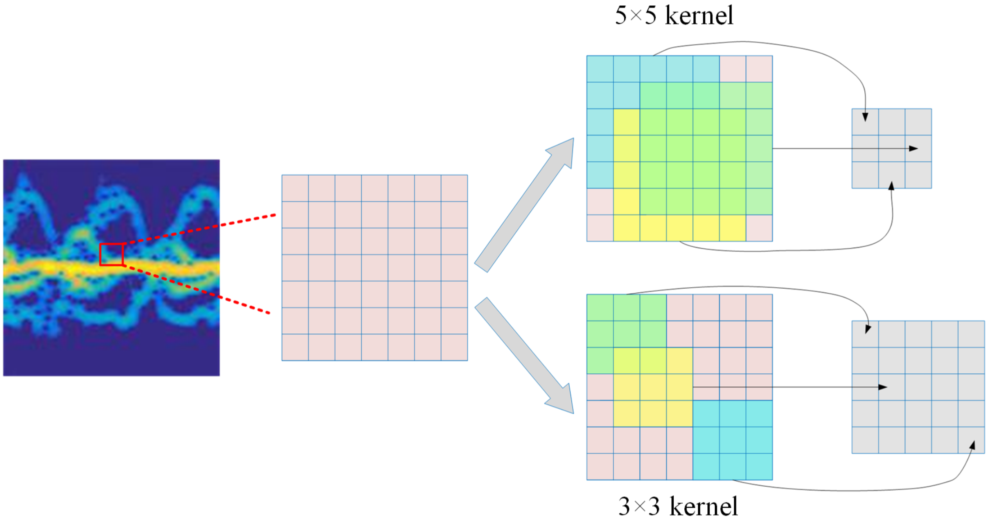

Figure 3. Convolution operations with different convolution kernels. The 5 × 5 kernel is suitable to learn coarse-scale features while the 3 × 3 kernel is suitable to learn fine-scale features in MD signatures.

In addition, the 1 × 1 kernel is also adopted for fusing the features in every block. By flexibly adjusting the number of channels, the 1 × 1 convolution kernel is capable of significantly increasing the nonlinear characteristics of the network without a loss of resolution and realizing cross-channel interaction and information integration.

3.2. Residual Attention Learning

We introduce the residual attention learning mechanism into our model to make *MRA-Net* learn more attention-aware representations from the input MD signatures. The output of every block in Figure 2 can be denoted as:

$$O_b(x) = (M(x) + 1) \cdot f(O_{p1}(x) + O_{p2}(x)) \tag{1}$$

where x represents the input of the block; $M(x)$ represents the residual attention mask; $O_{p1}(x)$ and $O_{p2}(x)$ represent the outputs of the coarse-scale and fine-scale learning branches, respectively; f represents the convolutional operation with a kernel size of 1×1; and $O_b(x)$ represents the output of the block.

In this paper, we tend to learn discriminative embedding from MD signatures for both radar-based activity recognition and person identification tasks. In a CNN, the convolution operation is achieved by sliding the convolution kernel over the feature map. Thus, the feature learning process treats each area of the input equally. However, it is obvious that the MD frequency parts in an MD signature are more representative of the corresponding activity, and thus should receive greater attention. To this end, the elaborate residual attention learning is adopted to make *MRA-Net* focused, as shown in Branch 3 in Figure 2. Residual attention learning is composed of two components: residual learning mechanism and mixed attention mechanism [13]. The bottom-up, top-down feedforward residual attention mechanism is realized by multiple stacked attention modules that generate attention-aware features and aim at guiding more discriminative feature representations. The stacked structure is the basis of mixed attention mechanism, where different types of attention can be obtained from different attention modules. Due to the obvious performance drop caused by module stacking, residual learning mechanism is adopted for optimizing the deep model.

3.3. Fine-Grained Loss Weight Learning

To make the multitask classifier recognize activities and identify persons more accurately, we present the FLWL mechanism in the multitask classifier, aiming to automatically assign a proper loss weight to each task and to simultaneously retain good multitask classification performance.

Given a training set T composed of N MD signatures and their labels: $T = \{S_n, L_{rn}, L_{in}\}_{n=1}^{N}$, where S_n denotes the nth input MD signature, L_{rn} denotes the corresponding label of human activity, and L_{in} represents the corresponding label of person identification. $X_n \in \mathbf{R}^{d \times 1}$ represents the high-level embedding vector of S_n:

$$X_n = g(S_n; \Theta), \tag{2}$$

where Θ denotes all parameters to be optimized in the feature extractor layers of *MRA-Net*, and g denotes the nonlinear mapping from the input signature to the shared embedding.

For a multitask classifier, the loss weight setting for the tasks are vital to the classification performance. During the multitask training process, suppose that the training losses of the activity recognition task and person identification task are represented as $Loss_r$ and $Loss_i$, respectively. Then, the overall loss is computed as the weighted sum of the two individual losses, which is written as,

$$\begin{aligned} Loss_{overall} = w_r \times \sum_{n=1}^{N} Loss_r(S_n, L_{rn}) \\ + w_i \times \sum_{n=1}^{N} Loss_i(S_n, L_{in}) \end{aligned} \tag{3}$$

where $Loss_{overall}$ represents the overall loss of the model, w_r represents the loss weight parameter of the activity recognition task, and w_i represents the loss weight parameter of the person identification task.

The conventional approach to set the loss weights is the greedy search algorithm, which is affected by the search step size. A large step size causes the search process difficulty in converging, while a small step size is time-consuming. Under this circumstance, we first utilize a greedy search algorithm for initialization and locate the rough ratio range of the two loss weights. Several typical ratio values r_w are adopted for the experiment and the results are illustrated in Figure 4. As shown in this figure, the person identification task is more sensitive to r_w compared with the activity classification task. Furthermore, when r_w is between $\frac{2}{3}$ and 1, the accuracies of person identification and activity classification both retain at a high level.

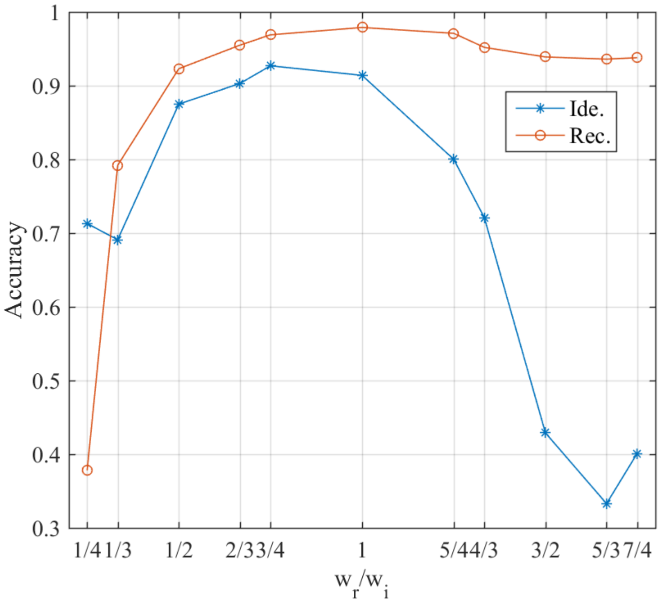

Figure 4. The initialization process of loss weights. From the accuracy curves of the activity recognition task and person identification task under several typical ratio values r_w, it is shown that $[\frac{2}{3}, 1]$ is the proper initial range for r_w.

Based on the result of the rough greedy search, we elaborate the multitask classifier illustrated in Figure 1. Specifically, in addition to the two branches P_i and P_r for activity recognition and person identification, we propose another branch P_w for automatic weight learning. Suppose that $\alpha_p \in \mathbf{R}^{d\times 2}$ and $\beta_p \in \mathbf{R}^{2\times 1}$ are the weight matrix and bias vector in the FC layer of P_w, respectively. Then, the output is fed into a *Softmax* layer,

$$\mathbf{w} = Softmax(\alpha_p^T X_n + \beta_p) \tag{4}$$

Consequently, the output of P_w is $\mathbf{w} = [w_1, w_2]^T$, where $w_1 + w_2 = 1$ and $0 \leq w_{1,2} \leq 1$.

Then, we design an overall loss function $Loss_{overall}$ for *MRA-Net* as follows:

$$\begin{aligned} Loss_{overall} = {} & (2 + max(w_1, w_2))Loss_r \\ & +(2 + min(w_1, w_2))Loss_i \end{aligned} \tag{5}$$

where $Loss_i$ denotes the cross-entropy loss function of the person identification task and $Loss_r$ denotes the cross-entropy loss function of the activity classification task. Therefore, the weight ratio r_w can be expressed as:

$$r_w = \frac{w_r}{w_i} = \frac{2 + min(w_1, w_2)}{2 + max(w_1, w_2)} = \frac{2 + min(w_1, w_2)}{3 - min(w_1, w_2)} \tag{6}$$

where w_r represents the loss weight for activity recognition and w_i represents the loss weight for person identification. In this way, *MRA-Net* is optimized under the limit of r_w, which is between $\frac{2}{3}$ and 1. Then, the fine-grained and optimal weights are automatically assigned for both tasks. With the proposed FLWL algorithm, the multitask classifier is able to automatically assign the loss weights and learn discriminative features for both tasks subsequently.

4. Dataset Description

In this work, we construct a measured radar MD signature dataset and verify the performance of *MRA-Net* with an FLWL mechanism. The measured MD signature data are collected with a UWB radar module named *PulsON 440* (see Figure 5a), and the recording parameters are given in Table 1. The experiment is performed in an indoor environment. The radar is placed at a height of 1 m, and activities are performed in the line of sight of the radar. The measurement range of the radar is between 1.5 m and 7.5 m. The experimental deployment diagram is illustrated in Figure 5b. In each experimental scenario, a subject performs a specified activity continuously for approximately 1.5 s. Thirty scenarios corresponding to five activities performed by six subjects are included in this dataset. Basic characteristics of the six subjects are recorded in Table 2, and the five activities are listed as follows: (a) directly walking towards/away from the radar (walk); (b) boxing while standing in place (box); (c) directly running towards/away from the radar (run); (d) jumping forward (jump); and (e) running in circle (circle).

Table 1. Radar configuration parameters.

Center Frequency	4.0 GHz
Chirp Bandwidth	1.8 GHz
Pulse Repetition Frequency (PRF)	290 Hz
Coherent Processing Interval (CPI)	0.2 s

Table 2. Subject information.

	Sub #1	Sub #2	Sub #3	Sub #4	Sub #5	Sub #6
Gender	male	male	male	female	male	female
Age	23	25	23	23	23	24
Height (cm)	173	178	172	166	188	169
Weight (kg)	73	71	75	66	92	52

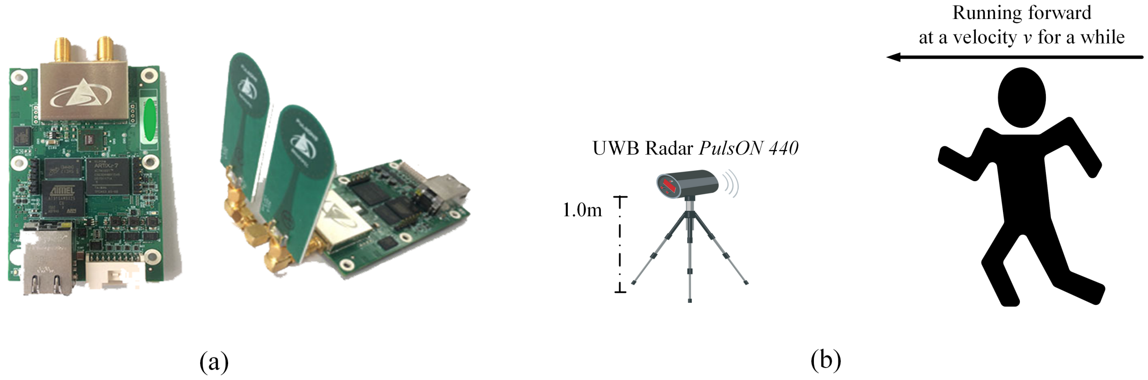

Figure 5. (**a**) Photo of UWB radar module named *PulsON 440*; and (**b**) experimental deployment diagram where a tested person is running towards the radar at a velocity v.

The radar data preprocessing is shown in Figure 6. Background clutter suppression via the moving target indicator approach is firstly performed on the raw data, and short-time Fourier transform (STFT) with a sliding window of 1 s is conducted subsequently. To take full advantage of the continuous radar motion data acquired in each scenario, the overlap of consecutive frames when using the sliding window is 0.36 s. Then, the radar motion data acquired in a scenario are transformed into a series of MD signatures. Several typical MD signatures in the dataset are shown in Figure 7. Each datum represents an activity that lasts 1 s, and the radical velocity ranges of all activities are between −5.14 m/s and 5.14 m/s. Consequently, a dataset composed of approximately 7498 spectrograms is obtained, and the concrete distribution is shown in Table 3. Then, we resize the spectrograms into 100 × 150 and feed them into the models.

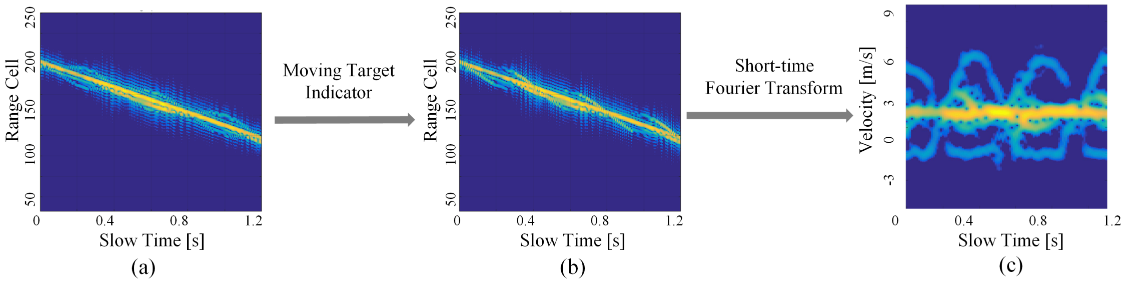

Figure 6. Radar data preprocessing: (**a**) Raw radar data; (**b**) radar data with background clutter suppression; and (**c**) radar MD signature.

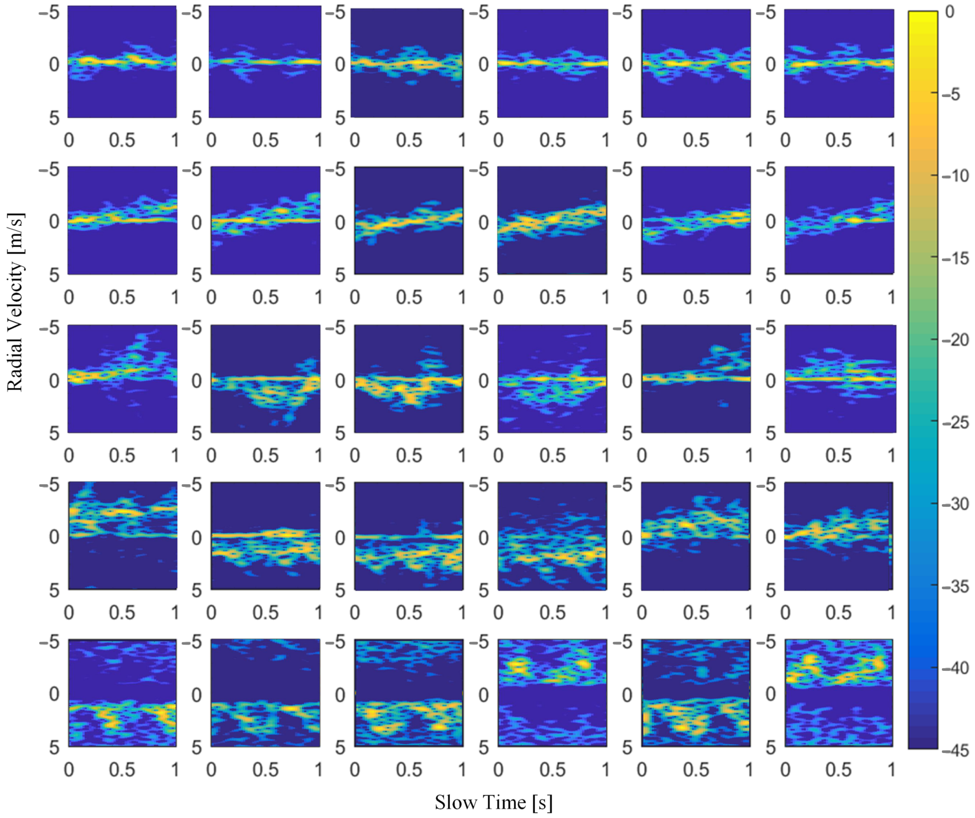

Figure 7. Typical MD signatures of five activities performed by six subjects. From top to bottom: box, circle, jump, run and walk. From left to right: Sub #1, Sub #2, Sub #3, Sub #4, Sub #5 and Sub #6. In each spectrogram, the radial velocity range is from −5.14 m/s to 5.14 m/s, and the activity duration is 1 s. It is shown that every piece of MD signature is both activity-specific and individual-unique.

Table 3. Dataset composition.

Num. Act. Sub.	Walk	Run	Jump	Circle	Box
Sub #1	340	338	166	396	266
Sub #2	388	282	160	356	197
Sub #3	233	237	220	430	158
Sub #4	208	211	100	242	141
Sub #5	256	175	190	320	145
Sub #6	177	414	197	316	239

5. Evaluation and Implementation

5.1. Performance Metrics

Due to the imbalance existing in the dataset, an appropriate performance metric is indispensable for evaluating the performance of activity recognition and person identification. As illustrated in Table 3, for several items, such as "jump" performed by Sub #1 and #2, and "walk" performed by Sub #6, there are only half as many as the other items, which causes a data imbalance issue. The overall accuracy is not able to measure the performance of these skewed classes. In this circumstance, in addition to *Accuracy*, more metrics should be introduced for comprehensively evaluating the performance of the two tasks. Four types of measurements, namely True Negatives (TN), False Positives (FP), False Negatives (FN) and True Positives (TP), are often utilized for assessing the performance of machine learning algorithms. TP, FP, TN, and FN are the numbers of instances of true positive, false positive, true negative and false negative, respectively. Then, *Accuracy*, *Precision*, *Recall* and *F1-score* for the two tasks are calculated with TN, FN, TP and FP as follows:

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP} \tag{7}$$

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

$$F1 - score = 2 \times \frac{Recall \times Precision}{Recall + Precision} \tag{10}$$

where *F1-score* is an indicator designed to comprehensively consider *Precision* and *Recall*.

5.2. Implementation Details

The model was implemented on Tensorflow, which is developed by Google Brain. The model was trained in a fully supervised way, and the gradients were backpropagated from the softmax layers. The network parameters were updated with an Adaptive Moment Estimation (Adam) optimizer, and the mini-batch size was 128. The cross-entropy function was adopted to compute the losses between the predictions and the targets for each task. In addition, we added an L2 penalty to the losses. All weights and biases were randomly orthogonally initialized with a learning rate of 5×10^{-5}, and the momentum was set to 0.9. We trained the model for 400 epochs. All experiments were performed on 4 Ti 1080 GPUs with 11 GB of memory, applying CUDA for acceleration.

6. Experiments and Discussion

6.1. Experimental Results

We employed five-fold cross validation on *MAR-Net*, and Table 4 shows the test F1-scores of *MRA-Net* for the person identification and activity recognition tasks. The architecture of *MRA-Net* can be used for not only MTL but also single-task learning. When applying *MRA-Net* to single-task learning, such as the activity classification task, the P_w and P_i of MAR-Net in Figure 1 are expired and P_r remains. In this way, *MAR-Net* is transformed into a single-task learning model. As shown in Table 4, irrespective of whether it is used in MTL or single-task learning, *MRA-Net* is capable of providing good performance for the two tasks, with the F1-scores of above 90%. As illustrated in Equation (10), F1-score is more comprehensive, which derived from precision and recall. A high F1-score means that the corresponding precision and recall are high enough. Consequently, it is indicated that the proposed *MRA-Net* architecture is effective and robust enough in both single-task learning and MTL. Specifically, the F1-score of activity classification is stable between 97% and 98% in both multitask and single-task learning. Meanwhile, the performance of person identification in MTL is better than that in single-task learning, with a margin of 4.53% in F1-score. This result shows that the shared representations learned in MTL are more generic and better facilitate the person identification task.

Table 4. Test F1-scores of *MRA-Net* for multitask learning and single-task learning.

	Activity Recognition	Person Identification
MRA-Net for multitask learning	98.29%	95.87%
MRA-Net for activity recognition	97.61%	×
MRA-Net for person identification	×	91.34%

Next, Figure 8 displays the confusion matrix of activity recognition and person identification in MTL. As shown in Figure 8a, Sub #5 is the easiest person to be identified according to his activities. From Table 2, we know that Sub #5 has a larger radar cross-section (RCS) because of his physiological properties, such as height and weight; thus, the backscattering echoes of his activities are more intense and more distinguishable. Additionally, Sub #1 and #3 are easily confused, which is probably due to their similar physiological properties and behavior styles. In Figure 8b, "walk" is the easiest activity to be recognized, which demonstrated the MD signature of "walk" is more discriminative than the other activities. "Circle" has a lower recall than the others because when a person is running in a circle in front of the radar, the aspect angle between the person and the radar changes dynamically. Consequently, the produced MD signatures are changeable and more difficult to recognize [30]. In addition, "run" and "circle" are more confused due to the action similarity.

Then, we investigated the performance of the five activities for person identification, respectively. Figure 9 illustrates the F1-scores of "walk", "run", "box", "jump" and "circle" for person identification. This figure shows that the five activities are all able to be utilized for identifying persons with F1-scores of more than 90%, which demonstrates the feasibility of MD-based person identification with the five activities. In detail, the spectrogram of "walk" is the most efficient for person identification with an F1-score of 96.50%, indicating that the motion of "walk" probably retains more personal information than the other counterparts. By contrast, the F1-score of the spectrogram of "box" for person identification is the lowest, approximately 91.38%. Since the RCS of "box" is smaller than those of the other motions, its MD signature is not as obvious as those of the others. Consequently, the identification performance with the MD signatures of "box" is slightly worse.

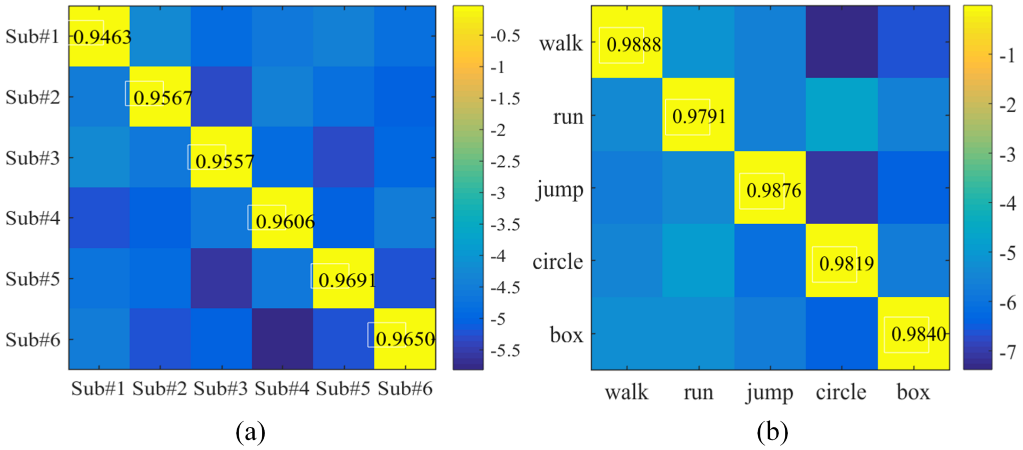

Figure 8. (**a**) Confusion matrix of the person identification task. (**b**) Confusion matrix of the activity recognition task. To illustrate the classification performance more clearly with confusion matrix, color is used to indicate the value *log(Recall)* instead of *Recall*.

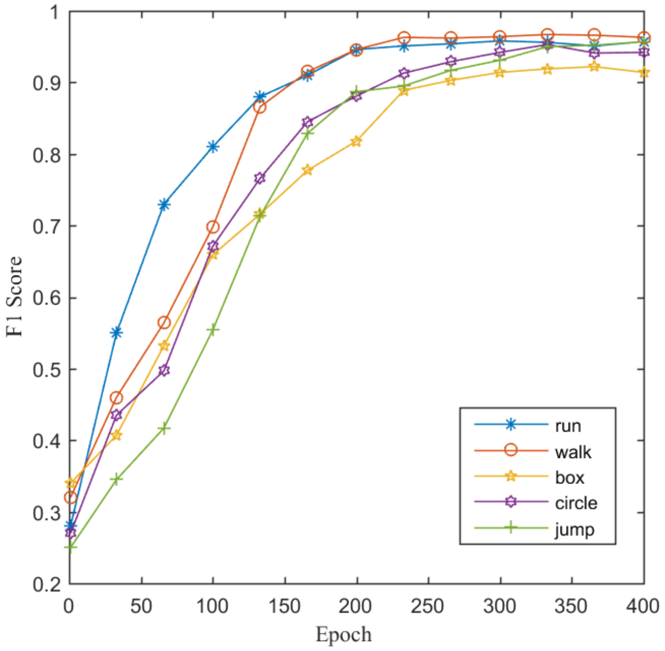

Figure 9. F1-score curves of five activities for person identification. "Run" has the highest F1-score, which indicates that the spectrogram of "run" is the most efficient for person identification among the five activities.

6.2. Comparison with the State-of-the-Art

To show the advantage of combining radar-based activity recognition and person identification tasks together with the MTL mechanism, we compare the performance of *MRA-Net* for the two tasks with that of several state-of-the-art methods. At present, the deep convolutional neural networks (DCNNs) in [5,19] are two typical models for radar-based person identification, and we selected them as baselines. The result is shown in Figure 10a. Our *MRA-Net* outperforms all the methods, with an approximately 4% higher F1-score than the DCNN in [19] and an approximately 8% higher F1-score than the DCNN in [5]. Furthermore, the DCNN in [23], the convolutional autoencoder (CAE) in [9] and the stacked long short term memory (LSTM) in [31] were treated as baselines for the radar-based activity recognition task, and the comparison results are illustrated in Figure 10b.

We found that, for the activity recognition task, *MRA-Net* with the MTL mechanism also achieves the best performance among these methods. Specifically, among the three baselines, the CAE obtains the best performance, with an F1-score of 95.08%, while the stacked LSTM obtains the lowest F1-score. Furthermore, the performance of the proposed *MRA-Net* is better than the CAE, with a margin of approximately 3.21% in F1-score. The results above indicate that *MRA-Net* with MTL is able to obtain a better performance for the two tasks than the state-of-the-art approaches. However, for both of the tasks, *MRA-Net* converges slightly slower than the other baselines, which is probably due to the parameter optimization complexity caused by MTL.

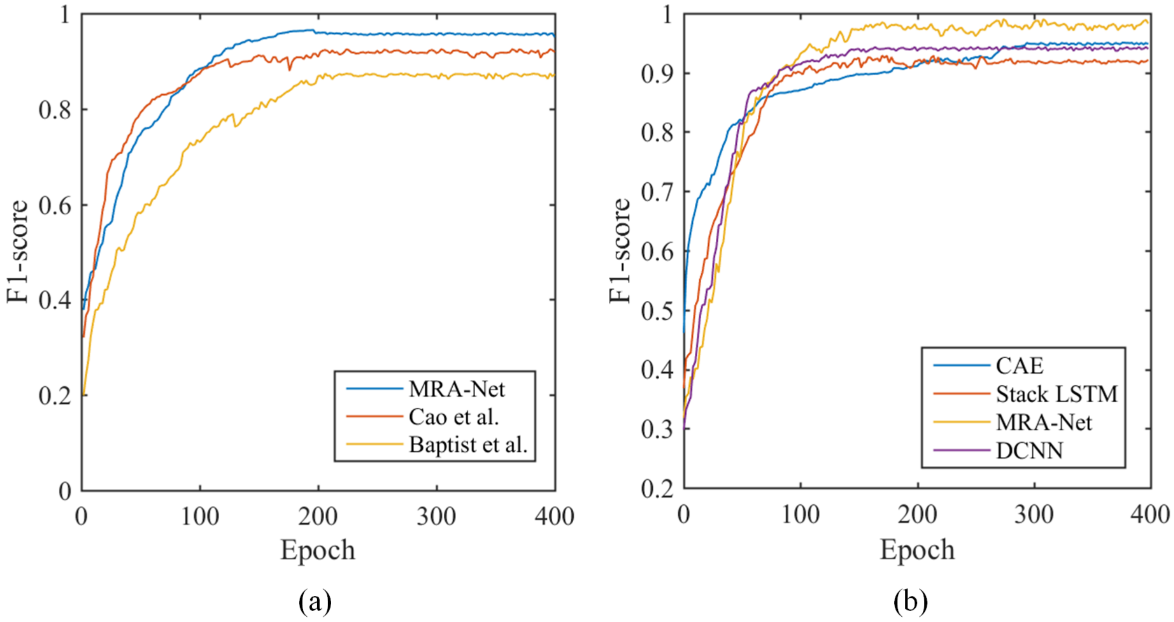

Figure 10. (**a**) Performance comparison for person identification on test dataset. (**b**) Performance comparison for activity recognition on test dataset. The proposed *MRA-Net* for joint activity recognition and person identification outperforms the state-of-the-art single-task approaches.

Subsequently, we compared the proposed *MRA-Net* with another MTL network, *JMI-CNN* [32], for the person identification and activity classification tasks. Additionally, to more comprehensively compare the two networks, Gaussian white noise (GWN) with different signal noise ratios (SNRs) was added on the MD signatures of the dataset. The performance of the two networks for both person identification and activity classification tasks under different SNRs was investigated. The results are illustrated in Figure 11. On the whole, the proposed *MRA-Net* outperforms *JMI-CNN* for MTL, indicating that that *MRA-Net* is more robust and generalized. For the person identification task, the F1-scores of *MRA-Net* are higher than *JMI-CNN* under various SNRs. Especially, when the SNR is greater than 15 dB, the F1-scores of *MRA-Net* are higher than 90%. For the activity recognition task, both of the networks obtain good performance of more than 95% in F1-score. When SNR is 5 dB, the two networks achieve almost the same classification F1-score, and, when SNR is 0 dB, *JMI-CNN* outperforms *MRA-Net* with a small margin. It is indicated that, when utilized for MTL, the proposed network outperforms *JMI-CNN* more markedly on the person identification task.

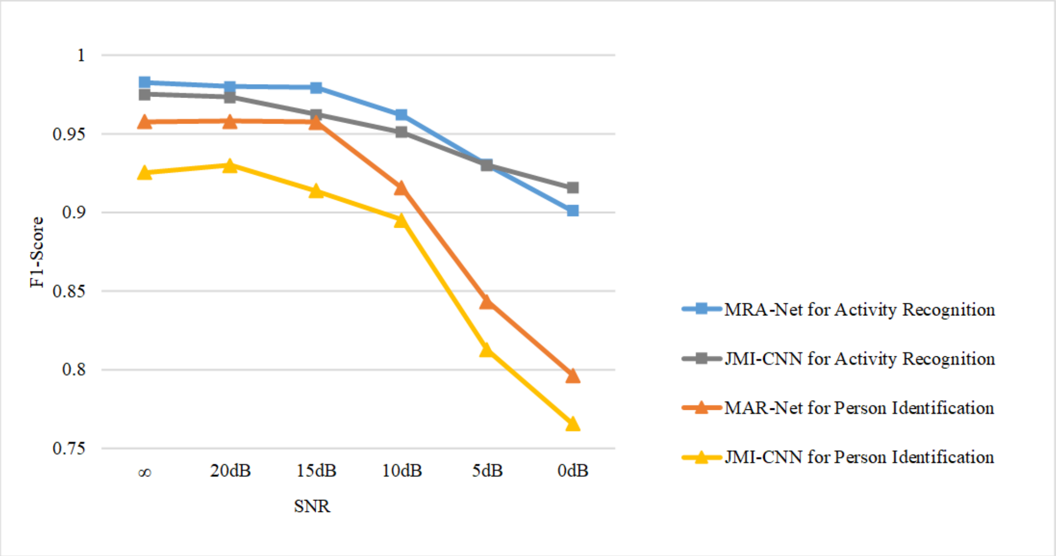

Figure 11. Performance comparison of *MRA-Net* and *JMI-CNN* for MTL under different SNRs.

6.3. Fine-grained Loss Weight Learning

In MTL, setting appropriate loss weights and elaborating a loss function are of crucial importance for model optimization. How to set the loss weight of each task for MTL remains an open question. Prior work either treats the tasks equally or acquires the loss weights via greedy search [28]. However, finding the optimal weights for all tasks via greedy search is time consuming or practically impossible. In this paper, our proposed FLWL mechanism is able to automatically learn the loss weight for each task and bring a better performance for MTL. First, we randomly selected a batch of MD signatures from the test data and fed them into the trained model, and then we visualized the loss weights of these data for the person identification and activity classification tasks, as illustrated in Figure 12a. As shown in Figure 12a, for each task, the loss weight assigned to different data is different, as described by the yellow and blue polylines. This result indicates that the optimal r_ws of different MD data are not exactly the same. Furthermore, we count the assigned r_ws of all test MD data and show the results in a bar chart, as illustrated in Figure 12b. The values of r_w in the bar chart are selected according to the relationship between w_r and w_i, as illustrated in Equation (6). When w_r increases uniformly from 2.0 to 2.5 at a rate of 0.1, r_w increases from $\frac{2}{3}$ to 1. Figure 12b demonstrates that r_ws are mostly between $\frac{23}{27}$ and $\frac{12}{13}$, and secondly between $\frac{21}{29}$ and $\frac{11}{14}$. Based on this result, we manually select several representative loss weight ratios and list the performance for MTL under these ratios, as illustrated in Table 5. It is indicated that the proposed FLWL mechanism is able to assign proper loss weights for each task to obtain a better performance. Although the performance of activity recognition is not obviously improved, the person identification task with the FLWL mechanism obtains the highest F1-score compared with greedy search. Additionally, the proposed mechanism is more efficient and labor-saving.

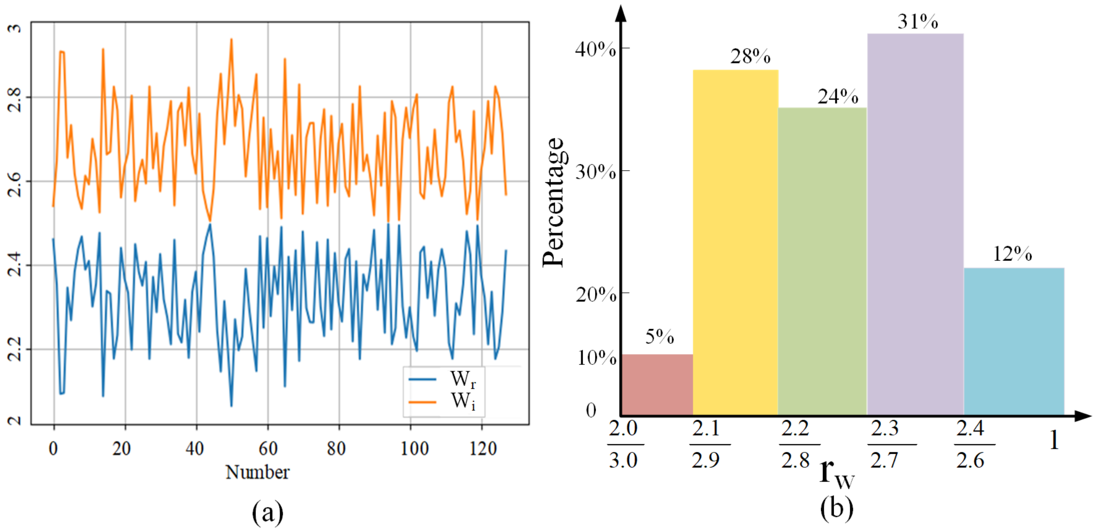

Figure 12. (**a**) Visualization of the loss weights for the person identification and activity classification tasks. (**b**) Bar chart for the statistical result of r_w. It is indicated that the automatically assigned loss weights for each MD signature vary, and most r_w are between $\frac{2.3}{2.7}$ and $\frac{2.4}{2.6}$.

Table 5. Performance comparison in F1-score for MTL with different loss weight ratios.

Multitask Learning		**Activity Recognition**	**Person Identification**
Greedy search	$r_w = \frac{2}{3}$	95.72%	88.97%
	$r_w = \frac{3}{4}$	97.28%	91.37%
	$r_w = \frac{11}{14}$	**98.43** %	94.85%
	$r_w = \frac{12}{13}$	96.13%	93.16%
	r_w =1	97.45%	90.23%
FLWL mechanism		98.29%	**95.87** %

6.4. Ablation Study

To demonstrate the necessity and effectiveness of the components in the model, some ablation studies on the performance of *MRA-Net* for activity classification and person identification are performed, as shown in Table 6. The contributions of three components (coarse-scale learning, fine-scale learning and residual attention learning) to MTL are investigated. Incorporating the three components into *MRA-Net* is able to greatly improve the performance for both activity recognition and person identification tasks, with limited increase of computational effort. From Rows (1)–(3) in Figure 12, we can find that applying multiscale learning alone does not significantly improve results. By contrast, residual attention learning is able to obtain more improvements. For example, when comparing Rows (2) and (5), residual attention learning obtains obvious improvements of 3.66% in F1-score for the activity recognition task, and 4.02% in F1-score for the person identification task. At the same time, the residual attention learning mechanism brings a higher computational complexity, increasing the execution time by 83.03%. Additionally, the fine-scale learning with a 3 × 3 convolution kernel offers more improvements for both tasks than the coarse-scale learning with a 5 × 5 convolution kernel when the residual attention learning mechanism is employed. Moreover, as expected, the incorporation of all three components results in the highest F1-scores for both tasks. Although the execution time of the structure in Row (6) is 136.03% longer than that of the structure in Row (1), the performance of the two tasks is improved obviously. The analysis of Table 6 indicates that all of the three components are necessary, with obvious improvements of performance and acceptable increases of execution time.

Table 6. Ablation study on *MRA-Net*.

	Multiscale Learning		Residual Attention learning Mechanism	F1-Score of Activity Recognition	F1-Score of Person Identification	Execution Time
	Coarse Scale	Fine Scale				
(1)	√	×	×	92.96%	89.75%	2.31 s
(2)	×	√	×	94.52%	90.14%	−13.96%
(3)	√	√	×	96.83%	91.30%	+89.28%
(4)	√	×	√	97.71%	93.89%	+83.03%
(5)	×	√	√	98.18%	94.06%	+75.41%
(6)	√	√	√	**98.29%**	**95.87%**	+136.03%

'Execution time' refers to the duration of the model to be trained once by all of the data in the dataset. We treat the execution time of the structure in Row (1) as a baseline time, and the execution time of the other structures is represented as an increment of the baseline time. For example, −13.96% denotes the execution time of the structure in Row (2) is 13.96% shorter than the baseline.

7. Conclusions

In this paper, a novel end-to-end neural network *MRA-Net* for joint activity classification and person identification with radar MD signatures was proposed. We explored the correlation between activity classification and person identification, and take advantage of the MTL mechanism to share computations between the two tasks. Multiscale learning and the residual attention mechanism were adopted in *MRA-Net* to learn more fully from the input MD signatures. Furthermore, instead of the conventional greedy search algorithm, we proposed an FLWL mechanism, which is also suitable for other multitask systems. We constructed a new radar MD dataset, with dual activity and identity labels for each piece of data, to optimize the proposed model.

The experiments showed that the proposed *MRA-Net* for joint learning achieved good performance with F1-scores of 98.29% for activity recognition and 95.87% for person identification. It outperforms not only *MRA-Net* for single-task learning but also some state-of-the-art radar-based activity recognition and person identification methods. In addition, the proposed FLWL mechanism further improves the performance of *MRA-Net*. The ablation studies indicated the efficacy of the components in the feature extractor of *MRA-Net*. In future work, we intend to further investigate the proposed model and design more reasonable multitask architectures for joint radar-based activity recognition and person identification. Additionally, more radar applications for smart sensing in IoT will be explored.

Author Contributions: Y.H. proposed the main algorithm of this research and wrote the manuscript. Y.H. and X.L. prepared the data used in the experiments. X.L. did the experiments and made further analysis. X.J. revised the paper and provided many useful modification suggestions. All authors discussed and reviewed the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China under Grant 61901049 and the BUPT Basic Research Funding 500419757.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

IoT	Internet of Things
MRA-Net	Multiscale Residual Attention Network
FLWL	Fine-grained Loss Weight Learning
MD	micro-Doppler
CNN	Convolutional Neural Networks
MTL	Multitask Learning

Fully Connected	FC
UWB	Ultra-wideband
PRF	Pulse Repetition Frequency
CPI	Coherent Processing Interval
TN	True Negatives
FP	False Positives
FN	False Negatives
TP	True Positives
Adam	Adaptive Moment Estimation
RCS	Radar Cross-Section
DCNN	Deep Convolution Neural Network
LSTM	Stacked Long Short Term Memory
CAE	Convolutional Autoencoder
GWN	Gaussian White Noise
SNR	Signal Noise Ratio

References

1. Lien, J.; Gillian, N.; Karagozler, M.E.; Amihood, P.; Schwesig, C.; Olson, E.; Raja, H.; Poupyrev, I. Soli: Ubiquitous gesture sensing with millimeter wave radar. *ACM Trans. Graph. (TOG)* **2016**, *35*, 142. [CrossRef]
2. Zou, Y.; Liu, W.; Wu, K.; Ni, L.M. Wi-Fi radar: Recognizing human behavior with commodity Wi-Fi. *IEEE Commun. Mag.* **2017**, *55*, 105–111. [CrossRef]
3. Lee, I.; Lee, K. The Internet of Things (IoT): Applications, investments, and challenges for enterprises. *Bus. Horizons* **2015**, *58*, 431–440. [CrossRef]
4. Brutti, A.; Cavallaro, A. Online Cross-Modal Adaptation for Audio–Visual Person Identification With Wearable Cameras. *IEEE Trans. Hum.-Mach. Syst.* **2017**, *47*, 40–51. [CrossRef]
5. Vandersmissen, B.; Knudde, N.; Jalalvand, A.; Couckuyt, I.; Bourdoux, A.; De Neve, W.; Dhaene, T. Indoor Person Identification Using a Low-Power FMCW Radar. *IEEE Trans. Geosci. Remote. Sens.* **2018**, *56*, 3941–3952. [CrossRef]
6. Kim, Y.; Ling, H. Human activity classification based on micro-Doppler signatures using a support vector machine. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 1328–1337.
7. Liu, K.; Liu, W.; Gan, C.; Tan, M.; Ma, H. T-C3D: Temporal convolutional 3D network for real-time action recognition. In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018.
8. Kim, J.H.; Hong, G.S.; Kim, B.G.; Dogra, D.P. deepGesture: Deep learning-based gesture recognition scheme using motion sensors. *Displays* **2018**, *55*, 38–45. [CrossRef]
9. Seyfioğlu, M.S.; Özbayğglu, A.M.; Gurbuz, S.Z. Deep convolutional autoencoder for radar-based classification of similar aided and unaided human activities. *IEEE Trans. Aerosp. Electron. Syst.* **2018**, *54*, 1709–1723. [CrossRef]
10. Yan, Y.; Ricci, E.; Subramanian, R.; Liu, G.; Lanz, O.; Sebe, N. A multi-task learning framework for head pose estimation under target motion. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 1070–1083. [CrossRef]
11. Liu, A.; Su, Y.; Nie, W.; Kankanhalli, M.S. Hierarchical Clustering Multi-Task Learning for Joint Human Action Grouping and Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 102–114. [CrossRef]
12. Ruder, S. An overview of multi-task learning in deep neural networks. *arXiv* **2017**, arXiv:1706.05098.
13. Wang, F.; Jiang, M.; Qian, C.; Yang, S.; Li, C.; Zhang, H.; Wang, X.; Tang, X. Residual attention network for image classification. *arXiv* **2017**, arXiv:1704.06904.
14. Ranjan, R.; Patel, V.M.; Chellappa, R. Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *41*, 121–135. [CrossRef] [PubMed]
15. Tian, Y.; Luo, P.; Wang, X.; Tang, X. Pedestrian detection aided by deep learning semantic tasks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 8–10 June 2015.

16. Liang, G.; Lan, X.; Zheng, K.; Wang, S.; Zheng, N. Cross-View Person Identification by Matching Human Poses Estimated with Confidence on Each Body Joint. In Proceedings of the AAAI Conference on Artificial Intelligence (AAAI), New Orleans, LA, USA, 2–7 February 2018.
17. Ali, M.M.; Mahale, V.H.; Yannawar, P.; Gaikwad, A. Fingerprint recognition for person identification and verification based on minutiae matching. In Proceedings of the IEEE 6th International Conference on Advanced Computing (IACC), Bhimavaram, India, 27–28 February 2016; pp. 332–339.
18. Zeng, Y.; Pathak, P.H.; Mohapatra, P. WiWho: Wifi-based person identification in smart spaces. In Proceedings of the 15th International Conference on Information Processing in Sensor Networks, Vienna, Austria, 11–14 April 2016; p. 4.
19. Cao, P.; Xia, W.; Ye, M.; Zhang, J.; Zhou, J. Radar-ID: Human identification based on radar micro-Doppler signatures using deep convolutional neural networks. *IET Radar Sonar Navig.* **2018**, *12*, 729–734. [CrossRef]
20. Huang, X.; Dai, M. Indoor device-free activity recognition based on radio signal. *IEEE Trans. Veh. Technol.* **2017**, *66*, 5316–5329. [CrossRef]
21. Zhang, T.; Li, N.; Huang, J.; Zhong, J.X.; Li, G. An Active Action Proposal Method Based on Reinforcement Learning. In Proceedings of the 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; pp. 4053–4057.
22. Wang, J.; Chen, Y.; Hao, S.; Peng, X.; Hu, L. Deep learning for sensor-based activity recognition: A survey. *Pattern Recognit. Lett.* **2019**, *119*, 3–11. [CrossRef]
23. Kim, Y.; Moon, T. Human detection and activity classification based on micro-Doppler signatures using deep convolutional neural networks. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 8–12. [CrossRef]
24. Le, H.T.; Phung, S.L.; Bouzerdoum, A.; Tivive, F.H.C. Human Motion Classification with Micro-Doppler Radar and Bayesian-Optimized Convolutional Neural Networks. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 2961–2965.
25. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017.
26. Mioranda-Correa, J.A.; Patras, I. A Multi-Task Cascaded Network for Prediction of Affect, Personality, Mood and Social Context Using EEG Signals. In Proceedings of the IEEE International Conference on Automatic Face & Gesture Recognition, Xi'an, China, 15–19 May 2018; pp. 373–380.
27. Kendall, A.; Gal, Y.; Cipolla, R. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. *arXiv* **2017**, arXiv:1705.07115.
28. Yin, X.; Liu, X. Multi-task convolutional neural network for pose-invariant face recognition. *IEEE Trans. Image Process.* **2018**, *27*, 964–975. [CrossRef]
29. Cai, Z.; Fan, Q.; Feris, R.S.; Vasconcelos, N. A unified multi-scale deep convolutional neural network for fast object detection. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 354–370.
30. Vignaud, L.; Ghaleb, A.; Le Kernec, J.; Nicolas, J.M. Radar high resolution range & micro-doppler analysis of human motions. In Proceedings of the 2009 International Radar Conference "Surveillance for a Safer World" (RADAR 2009), Bordeaux, France, 12–16 October 2009; pp. 1–6.
31. Wang, M.; Zhang, Y.D.; Cui, G. Human motion recognition exploiting radar with stacked recurrent neural network. *Digit. Signal Process.* **2019**, *87*, 125–131. [CrossRef]
32. Lang, Y.; Wang, Q.; Yang, Y.; Hou, C.; Liu, H.; He, Y. Joint Motion Classification and Person Identification via Multi-Task Learning for Smart Homes. *IEEE Internet Things J.* **2019**. [CrossRef]