



mm2Sleep: Highly generalized dual-person sleep posture recognition using FMCW radar



Yicheng Yao ^{a,b,c,d}, Hao Zhang ^{a,b}, Pan Xia ^{a,b}, Changyu Liu ^{a,b}, Fanglin Geng ^{a,b}, Zhongrui Bai ^g, Lidong Du ^{a,e}, Xianxiang Chen ^{a,e}, Peng Wang ^{a,e,*}, Weifeng Yao ^{f,*}, Ziqing Hei ^{f,*}, Zhen Fang ^{a,b,e,*}

^a Aerospace Information Research Institute, Chinese Academy of Sciences (AIRCAS), Beijing, China

^b School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing, China

^c National Key Laboratory of Electromagnetic Effect and Security on Marine Equipment, China

^d Nanjing Marine Radar Institute, China

^e Personalized Management of Chronic Respiratory Disease, Chinese Academy of Medical Sciences, Beijing, China

^f Department of Anesthesiology, Third Affiliated Hospital of Sun Yat-sen University, Guangzhou, Guangdong Province, China

^g School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China

ARTICLE INFO

Keywords:

Sleep posture recognition

FMCW radar

Data augmentation

Deep learning

ABSTRACT

Identifying and monitoring sleep posture is significant for reducing breathing pause events, monitoring sleep disorder patients, and avoiding bedsores. The traditional solution is to install cameras, wear wearable devices, or install pressure sensors. These methods are affected by obstructions such as blankets and also affect patient comfort. Radar sensors can effectively solve the problems of traditional solutions. However, existing radar-based sleep posture recognition systems can only be used in single-person scenarios and cannot simultaneously recognize the posture of two people. In addition, they also have the problem of low generalization. Therefore, we propose *mm2Sleep*, a system that uses millimeter-wave radar to recognize the sleeping posture of two people simultaneously. We use FMCW radar to calculate the human azimuth heatmap, and design a neural network to distinguish each person's posture adaptively. The proposed model is a dual-channel parallel structure, including position embedding layer, position segmentation module, feature extractor and predictor. We embed the position information into the radar signal map so that the position segmentation module can automatically identify the positions of two users. In addition, we propose multiple data augmentation methods for radar azimuth heatmap to increase the diversity of data. The proposed model can achieve high-accuracy sleep posture recognition without any calibration data from new users. We conducted experiments using data from 20 subjects. The results show that the F1 score of the proposed model is 0.874 without using any calibration data from new users.

1. Introduction

Sleep is closely related to human health, and sleep quality is related to the risk of cardiovascular disease, sleep apnea syndrome, and other diseases [1]. Sleep posture affects the quality of sleep and is a vital sign for the diagnosis of diseases. For example, the supine posture increases the frequency and severity of abnormal breathing events [2]. Infrequent changes in sleep posture may lead to pressure sores in postoperative patients [3]. Pressure sores can damage the patient's skin tissue,

increasing the patient's risk. In addition, compared with the non-prone posture, the prone posture was significantly associated with sudden death in patients with epilepsy [4]. Therefore, sleep posture monitoring can help monitor patients' physical condition, help patients with treatment, and reduce the health risk of patients.

Sleep monitoring devices are divided into contact devices and non-contact devices [5]. The contact method uses devices such as pressure sensors [6], activity recorders [7], and polysomnography (PSG) [8]. This type of method has high monitoring accuracy but requires a large

* Corresponding authors at: Aerospace Information Research Institute, Chinese Academy of Sciences (AIRCAS), Beijing, China (P. Wang) and (Z. Fang).

E-mail addresses: yaoyicheng19@mails.ucas.ac.cn (Y. Yao), zhanghao190@mails.ucas.ac.cn (H. Zhang), xiapan17@mails.ucas.edu.cn (P. Xia), liuchangyu21@mails.ucas.ac.cn (C. Liu), gengfanglin20@mails.ucas.ac.cn (F. Geng), zhongrui.bai@sjtu.edu.cn (Z. Bai), duld@aircas.ac.cn (L. Du), chenxx@aircas.ac.cn (X. Chen), wangpeng01@aircas.ac.cn (P. Wang), yaowf3@mail.sysu.edu.cn (W. Yao), heizq@mail.sysu.edu.cn (Z. Hei), fangzhen@aircas.ac.cn (Z. Fang).

number of sensors to be worn on the mattress or user's body, which can be uncomfortable for the user. Traditional non-contact sleep monitoring mainly uses computer vision-based solutions, using cameras [9] or infrared cameras [10] as sensors. This type of method usually invades users' privacy, and users may be concerned that their privacy will be leaked and, therefore, unwilling to use such devices. In addition, such devices are susceptible to light interference and may cause system performance degradation due to users covering themselves with blankets.

The RF-based sensor is a contactless device that does not require the user to wear the device [11]. The RF-based sensor does not invade the user's privacy and can work in dark conditions. In addition, the radio frequency signal has a penetration characteristic and can penetrate the cover of the user [12]. The radio frequency sensors have been used to identify human activity and monitor the physical signs of subjects [13,14]. The existing studies on sleeping posture recognition based on radio frequency have low generalization, and their performance degrades significantly when the model is applied to new users [15]. A common solution is to collect part of the data adjustment model of new users, but this will bring difficulties in promoting the model [16]. In addition, previous work could only recognize the sleeping posture of one person and could not recognize the sleeping posture of two people in bed simultaneously. Sharing a bed is a common phenomenon in daily life, and it is important to monitor the sleeping posture of two users on the bed at the same time. In the two-person scenario, the radar receiving signal contains the composite signal of the two users, which may lead to the failure of the traditional single-sleeping position recognition method. Monitoring the sleeping posture of two users at the same time increases the complexity of data processing and requires solving the technical difficulties of signal separation. In addition, using a single radar to monitor the sleeping posture of two users simultaneously can effectively reduce hardware costs and installation complexity, which means lower input and maintenance costs.

This paper introduces mm2Sleep, a sleep posture recognition system based on FMCW radar. The system can recognize the sleep posture of two users on the bed simultaneously and maintain high-accuracy sleep posture recognition without any new user data calibration. The proposed method uses FMCW radar with multiple input multiple outputs (MIMO) antenna array to estimate the user's azimuth heatmap. We propose a dual-channel neural network to extract each user's signals and predict each user's sleep posture. In addition, we propose various data augmentation methods based on radar azimuth heatmaps to improve the diversity of data. In particular, the proposed data augmentation method can enhance each user's signal and simulate a richer situation.

The main contributions of this paper are summarized as follows:

1. We design an easy-to-use dual-person sleep posture recognition system to achieve high-accuracy sleep posture recognition without needing model calibration using new user data. To our knowledge, this is the first study to achieve dual-person sleep posture recognition using radar.
2. We design a dual-channel neural network to recognize the sleep posture of two people simultaneously. We set up a position segmentation module to learn the masks of two users, filter the original data, and extract the information of the two users separately.
3. We research and apply six data augmentation methods for radar azimuth heatmaps, improving the model's generalization ability. We conduct ablation experiments to analyze and evaluate the effectiveness of different data augmentation methods.

The rest of the paper is organized as follows. Section 2 discusses the related work in sleep posture recognition. Section 3 introduces the data collection and processing procedure. Section 4 describes the methodology of this paper. Section 5 shows the experimental results and evaluations of the proposed framework. Finally, section 6 concludes the paper.

2. Related Works

2.1. Contact Equipment-Based sleep posture recognition

The method based on the contact device is to place the sensor in the position in contact with the human body [17–19], and its sensor types mainly include pressure sensor array and acceleration sensor [20]. Acceleration sensor-based or gyroscopic sensor-based methods can accurately capture the angle and motion of the body by wearing the sensor on the waist or chest, thereby inferring the sleeper's posture. For example, Chang *et al.* [18] wore a three-axis accelerometer on the user's body to recognize sleeping posture based on the direction of the accelerometer. This method can affect the user's sleep comfort and may cause the device to fail if the user forgets to charge it. Another contact-based solution is to install a sensor array on the mattress to monitor sleeping posture based on pressure and other information sensed by the sensor array [17,19]. Pressure sensors embedded in the mattress can sense the pressure distribution of the human body in different sleeping posture. When people change the sleeping posture, different body parts will have different pressure on the mattress, by analyzing these pressure changes, you can judge the posture of the sleeper. This type of solution will change the feel of the mattress and affect the user's sleep comfort. In addition, this type of solution requires a large number of sensors to maintain high accuracy.

The advantage of sleep position recognition method based on contact device is that it can provide accurate sleep position data, which is helpful to study sleep quality and improve sleep environment. However, prolonged exposure to the sensor can cause discomfort and face user compliance issues, and the pressure array-based approach adds cost.

2.2. Vision-Based sleep posture recognition

Vision-based sleep posture monitoring systems use RGB cameras [21], infrared cameras [22], or depth cameras [23] to obtain images or videos of users and then use deep learning models to recognize user sleeping postures. Ordinary cameras can be used to collect images in the visible range, which is suitable for environments with good light lines. Infrared cameras can work at night or in low-light environments, not limited by light conditions, and are often used for sleep monitoring. The depth camera can not only obtain two-dimensional images, but also obtain spatial information, and the recognition of sleeping position is more accurate. This method has the advantages of non-invasive and no direct contact with the human body, so it has a wide application prospect in smart home, health monitoring and other fields. For example, Liu *et al.* [21] used an RGB camera to obtain a top-down view of the human body to detect the person's position and then recognize the sleeping posture. However, vision-based methods are usually limited by three factors. First, the performance of RGB camera-based methods is usually affected by lighting. Second, when the user covers the blanket or quilt, the system's performance will significantly reduce [21,22]. Third, users may be concerned that their privacy will be leaked and, therefore, unwilling to install cameras in their bedrooms.

2.3. Rf-based sleep posture recognition

RF-based sleep posture recognition methods do not have the limitations of wearable devices and cameras and can be better integrated into users' daily lives. In recent years, researchers have used RF sensors in activity recognition [24], sleep monitoring [25], physiological signal monitoring [26], and other fields. The sensors used in RF-based sleep posture recognition technology include radar, RFID [15], and WIFI devices [27,28]. Liu *et al.* [15] extracted the user's breathing information by deploying three RFID tags near the user. They recognized the sleeping posture by analyzing the clarity of the breathing pattern extracted from the three RFID tags. This method depends on the deployment of tags and requires additional tripods to be installed when no walls are next to the

bed. In addition, they reported that their model's performance significantly decreased when directly applied to new users.

WIFI-based methods analyze channel state information (CSI) to recognize sleeping posture because the impact of CSI is different under different postures. For example, Liu *et al.* [28] extracted features from the CSI time series and used machine learning methods such as SVM to recognize sleeping posture. They analyzed the CSI signal in the frequency domain to simultaneously extract the breathing signals of two people in bed because different breathing rates produce different peaks in the spectrum. However, they could not extend this method to tasks that recognize the sleeping posture of two people simultaneously because they could not extract features separately for each user.

Radar sensors can perceive the distance and orientation of users. Researchers recognize user posture by analyzing radar position heat maps or analyzing distance information changes caused by turning over [16,29–31]. Kiriazi *et al.* [31] and Islam *et al.* [30] proposed to use dual-frequency radar to identify sleep posture. They classify sleep postures by setting radar reflection parameter thresholds or using machine learning classifiers. Although they have achieved high accuracy, they use the same subjects' data to calculate the threshold or training model, and the accuracy obtained can not reflect the performance of the model in the use of new users. Piriyajitakonkij *et al.* [29] used impulse radio ultra-wideband (IR-UWB) radar to capture the time-distance signal of user sleep posture transitions. They used neural networks to recognize the transition process of four sleep postures from the time-domain and frequency-domain information of the analyzed signal. However, complete posture recognition requires recognition of at least six posture transition processes (supine to side, supine to prone, side to supine, side to prone, prone to side, and prone to supine). If left-side and right-side sleeping postures are to be recognized, 12 posture transition processes need to be recognized. Lai *et al.* used IR-UWB radar for sleep posture recognition, and they respectively tested the accuracy of sleep posture recognition under single radar configuration, dual radar configuration and three-radar configuration [32]. Their method achieved up to 0.586 accuracy in a single radar configuration, 0.808 accuracy in a dual radar configuration, and 0.651 accuracy in a three-radar configuration. Using multiple radars increases accuracy, but it also increases user costs. Yue *et al.* [16] used neural networks to analyze the position heat map to infer the sleep posture of the subject. Their model only targets single-person scenarios and cannot be extended to dual-person scenarios. In addition, their method requires collecting data from new users to calibrate the model to maintain high performance when used by new users. Luo *et al.* proposed a sleeping position recognition method based on single-input and single-output FMCW radar [33]. They placed FMCW radar above the bed and calculated distance Doppler graphs to identify sleeping positions. In the two-user scenario, the signal of the two users will be aliasing in the distance and Doppler dimension of the distance Doppler graph, so that the method is not suitable for the two-user scenario.

The above RF sensors for sleeping position recognition include RFID, WIFI, Doppler radar, IR-UWB radar and FMCW radar. In the above sensors, methods based on RFID, WIFI and Doppler radar identify sleeping positions by extracting the user's breathing signal or channel status. However, the signals of the two subjects are aliased in the signals extracted by RFID, WIFI, and Doppler radar, so separating the signals of different users based on the signals from these sensors is challenging. In addition, the reflected signals from different parts of the user's body are also aliased in the signal. IR-UWB and FMCW radars can detect reflected signals at different distances, and can set up multi-transmit and multi-receive antenna arrays to detect echoes at different angles. These two radar sensors have the ability to resolve signals at different positions in space, and are more suitable for resolving signals from different users. FMCW radars can be modulated with a wide bandwidth and thus provide a high range resolution. In addition, compared with IR-UWB radar, FMCW radar is simpler in structure, smaller in size and lower in cost. Inspired by the above research, we propose a dual-channel neural

network, which can recognize the sleep posture of two subjects simultaneously. In addition, we propose various data augmentation methods for radar azimuth heatmaps, which simulate more abundant situations and improve the model's generalization. The proposed system can realize high-accuracy two-person sleep posture recognition without using the new user data to calibrate the model.

2.4. Data augmentation

In order to ensure a good performance of the model, deep learning models require a large amount of data for training. However, data acquisition and labeling consume a lot of time and resources in many cases. In many real-world scenarios, sufficient training data is usually not available [34]. Data augmentation can effectively alleviate the problem of insufficient training data and thus improve the performance of the model. The main goal of data augmentation is to increase the quantity, quality and diversity of training data. Typical data augmentation operations include flipping, rotating, scaling, cropping, adding noise to the data, etc. [35]. Another classical data augmentation method is data generation, which is used to improve the accuracy and generalization ability of the model by generating diverse and rich data. For example, data generation using variational auto-encoders [36] and generative adversarial networks [37].

For radar signals, the original received signal is the collection of all the reflected echoes in the environment. The original radar signal is not intuitive, and it is difficult to obtain the user's sleeping position signal in different scenarios by processing the original radar signal. For example, the distance and angle of the user relative to the radar will change during sleep, but it is difficult to process the raw radar signal to simulate the distance and angle changes. In contrast, radar azimuth heat maps contain intuitive information about the user's distance and azimuth relative to the radar. The data augmentation of the radar azimuth heat map can more intuitively simulate the radar azimuth heat map of users in different attitudes. Therefore, in this paper, the radar azimuth heat map is enhanced to expand the data.

3. Data collection and Preprocessing methodology

3.1. Radar device

In this work, we used TI's IWR6843ISK radar sensor. The sensor has three transmitting antennas and four receiving antennas, which can transmit and receive frequency-modulated continuous wave signals from 60 GHz to 64 GHz. The azimuth field of view of the sensor is 120 degrees, and the elevation field of view is 30 degrees. The signal transmitted by radar can be expressed as the following formula:

$$S_{tx}(t) = \exp\left(j2\pi\left(f_0 t + \frac{B}{2T} t^2\right)\right) \quad (1)$$

Where f_0 is the starting frequency of the transmitted signal, T is the chirp duration, B is the bandwidth. The signal received after the delay of τ is:

$$S_{rx}(t - \tau) = \exp\left(j2\pi\left(f_0(t - \tau) + \frac{B}{2T}(t - \tau)^2\right)\right) \quad (2)$$

$$\tau = \frac{2R}{c} \quad (3)$$

where R is the distance from the reflector to the radar, c is the speed of light. The signal is processed by mixer and low-pass filter, and the obtained intermediate frequency signal is:

$$S(t) = \exp\left(j2\pi\left(\frac{2BR}{Tc} t + f_0 \frac{2R}{c}\right)\right) \quad (4)$$

We set the starting frequency of the radar to 60 GHz, set the idle time to 80us, set the ADC start time to 7us, set the ramp end time to 60us, set

the frequency slope to 66.59 MHz/us, set the TX start time to 1us, set the ADC sampling points to 100, and set the ADC sampling rate to 2000ksps. The period of a frame is 20 ms. In a frame, three transmitting antennas cyclically transmit chirps in a time-division multiplexing manner, with each antenna transmitting 32 chirps. Based on the time-division multiplexing of transmitting antennas, 12 virtual receiving antennas can be formed at the receiving end. Radar parameter configuration is shown in Table 1.

3.2. Acquisition of radar azimuth heatmap

The problem we propose is to identify the sleep posture of two users in bed at the same time. Based on this requirement, we need to distinguish the signals of the two users. The proposed radar system has three transmitting antennas and four receiving antennas, and eight virtual receiving antennas can be formed in the horizontal direction. These eight virtual receiving antennas allow us to estimate the position of the user using the azimuth estimation method and form a distinguishable signal containing two users on the range-azimuth heatmap. FFT transform is performed on each receiving chirp $S(t)$ to obtain the signal $f(n)$ containing the distance information of the target. In order to eliminate the influence of static objects, each signal is subtracted from the mean value of all signals received by the antenna in 5 s:

$$f_i(n) = f_i(n) - \frac{1}{N} \sum_{j=1}^N f_j(n) \quad (5)$$

Where i is the chirp index within 5 s, and n is the index of the sampling points in a chirp.

In order to obtain a high-quality azimuth heat map, we configure the transmitting antenna as a time division multiplexing model and form a virtual antenna array at the receiver. We use Minimum Variance Distortionless Response (MVDR) to estimate the direction of arrival of the signal and obtain high-quality azimuth heat map [38]:

$$P(n, \theta) = \frac{1}{a^H(\theta) R(n)^{-1} a(\theta)} \quad (6)$$

$$R(n) = E[X(n)X(n)^T] \quad (7)$$

$$a(\theta) = [1, e^{-j\pi\sin(\theta)}, e^{-j2\pi\sin(\theta)}, \dots, e^{-j(K-1)\pi\sin(\theta)}] \quad (8)$$

Where n is the index of range dimension, θ is the index of angle dimension, $X(n)$ is the data matrix of the virtual antenna array, $a(\theta)$ is the direction vector, $E[\cdot]$ represents the expectation of the signal, and K is the number of virtual receiving antennas. In addition, we perform a log transformation on the azimuth heatmap to amplify the details of the heatmap. Fig. 1 shows the radar azimuth heat maps corresponding to the four sleep postures.

3.3. Data collection

A total of 20 subjects participated in this study, including six females and 14 males. The subjects' heights ranged from 1.5 to 1.9 m, and their weights ranged from 45 kg to 100 kg. Subjects' body mass index (BMI)

ranged from 17.6 to 29.8. We collect data in three rooms. In each experiment, we adjust the position of the bed and radar in the room to simulate different environmental arrangements. We installed the radar directly above the center of the bed, tilted downward to ensure the human body was within the radar's field of view. Fig. 2 shows the layout of the experimental scenario. We collected four sleeping postures: supine, prone, left lateral, and right lateral. Each experiment included one or two subjects to simulate single and double-user scenarios. The dataset collected in the i -th experiment is denoted as $D^i = (D_1^i, D_2^i)$. Dataset D_1^i represents pre-collected data, while dataset D_2^i represents test data. When the data from the i -th experiment is used for training, datasets D_1^i and D_2^i are added to the training set. When the data from the i -th experiment is used for testing, we test the model using data from dataset D_2^i . We collect two sets of data in each experiment because we have set up an ablation experiment to explore the impact of using new user data to calibrate the model on performance. When calibrating the model using new user data, we use data from dataset D_1^i . It should be noted that the proposed method does not include model calibration and does not use data from dataset D_1^i of test users to calibrate the model. We conducted 13 groups of experiments, including seven groups of two-person scene experiments and six groups of single-person scene experiments. The subjects in each experiment were different.

After collecting each sample, we asked the subject to change their position and posture. During the experiment, we asked the subject to randomly change the position of their arms and legs to ensure the diversity of postures. When collecting D_1^i , we asked the subjects to perform the postures in a prescribed order. There were 16 combinations of the two users' postures for scenarios involving two subjects, with four samples collected for each combination. For single-subject scenarios, we collected four samples for each posture. When collecting D_2^i , we asked the subjects to randomly change their position and posture after each sample was collected. For scenarios involving two subjects, we collected at least 160 samples. For scenarios involving a single subject, we collected at least 80 samples. Approval of all ethical and experimental procedures and protocols was granted by the Ethics Committee of Chinese PLA General Hospital.

3.4. Data processing

The distance between a person and the radar is variable. In order to reduce the impact of position on the model, it is necessary to identify the person's position and extract the region of interest in the signal. Fig. 3 shows the process of extracting the region of interest. First, we average the azimuth heatmap in the angle dimension to obtain the distribution of the signal on the distance axis. Second, we select the coordinate point with the highest power on the distance axis as the user's position. Third, we extract the signal within a range of 50 on the distance axis around the user's distance coordinate as the region of interest in the signal map. This method determines the user's position based on energy and can reduce the impact of position on the model.

The power of pixels in the radar azimuth heatmap is also affected by the distance between the user and the radar. In order to eliminate the modulation of the signal by distance, we use the Min-Max normalization method to map the values of the region of interest signal P_{ROI} to the [0,1] interval:

$$\tilde{P}_{ROI} = \frac{P_{ROI} - \text{Min}(P_{ROI})}{\text{Max}(P_{ROI}) - \text{Min}(P_{ROI})} \quad (9)$$

4. Proposed methodology

4.1. Method Overview

The proposed dual-person sleep posture recognition model consists of a position embedding layer, a position segmentation module, two

Table 1
Radar parameter configuration.

*Configuration Parameter	Value	Configuration Parameter	Value
Start Frequency	60.0 GHz	ADC Start Time	7.0us
Idle Time	80.0us	ADC Samples	100
Frequency Slope	66.59 MHz/us	Sample Rate	2000ksps
TX Start Time		Ramp End Time	60.0us
Frame Period	20 ms	No of Chirp Loops	32

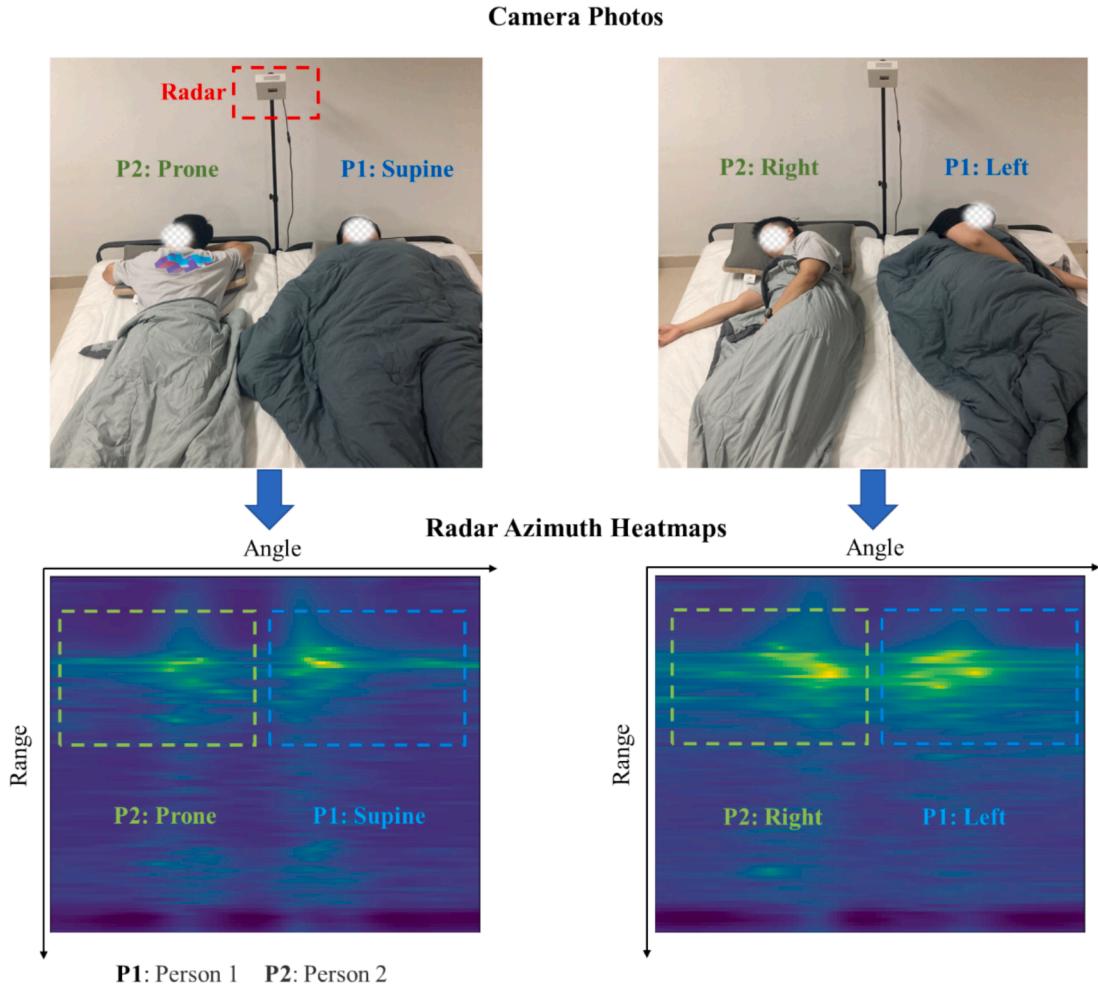


Fig. 1. Radar azimuth heat maps corresponding to four sleep postures in a two-person scene.



Fig. 2. The layout of the experimental scenario.

feature extractors, and two predictors. The model is shown in Fig. 4. The position embedding layer embeds position information into the input azimuth heatmap. The proposed dual-person sleep posture recognition model is a dual-channel model that separately judges the sleep posture of each user. The position segmentation module includes two mask

generation networks that extract the signals of two users separately based on the heatmap embedded with position information. The feature extractor is used to extract the spatial features of the azimuth heatmap. The predictor recognizes sleep posture based on the features extracted by the feature extractor.

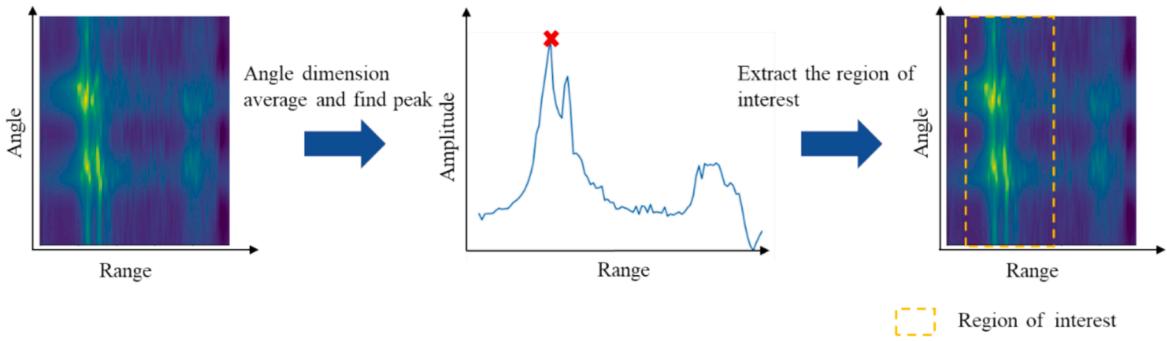


Fig. 3. The proposed process for extracting the region of interest.

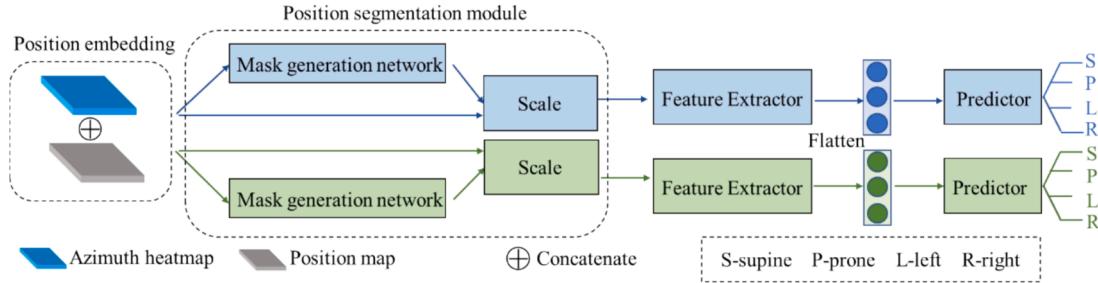


Fig. 4. The framework of the proposed dual-person sleep posture recognition model.

In addition, to increase the richness of the data, we propose various data augmentation methods. Specifically, we propose a dual-person positioning-based data augmentation method that enhances the signals of two users separately. The proposed data augmentation method simulates the signals after changes in distance and angle for each user, as well as the signals after changes in the position of different body parts of the user, and also simulates the signals after being affected by different levels of noise. By combining various proposed data augmentation methods, the diversity of the data is improved, significantly improving the performance and generalization of the model.

4.2. Neural network structure

4.2.1. Position embedding

Recognizing the sleeping posture of two people requires us to extract the features of each user separately. The convolution kernel of the convolutional neural network is shared, and the signals of the two people are extracted through the same convolution kernel. Therefore, it is difficult for the convolution operation to extract the signals of the two users separately. In this paper, we proposed a position embedding method to embed position information into the azimuth heatmap to help the neural network distinguish the signals of different users. First, we generate the abscissa matrix X and ordinate matrix Y , and normalize them:

$$X = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 2 & 2 & \cdots & 2 \\ \vdots & \vdots & \ddots & \vdots \\ M & M & \cdots & M \end{bmatrix}, Y = \begin{bmatrix} 1 & 2 & \cdots & N \\ 1 & 2 & \cdots & N \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 2 & \cdots & N \end{bmatrix} \quad (10)$$

$$X = \frac{X - \text{mean}(X)}{\text{std}(X)}, Y = \frac{Y - \text{mean}(Y)}{\text{std}(Y)} \quad (11)$$

Where M and N are the lengths of the distance dimension and angle dimension of the radar azimuth heatmap, respectively. We stack the radar azimuth heatmap with the abscissa matrix and ordinate matrix as the new sample and input it into the neural network model. The new sample contains position information, which can assist the neural

network in extracting the signals of users at different positions and help distinguish between two users.

4.2.2. Position segmentation module

As the number of convolutional layers increases, the receptive field of the output features becomes larger, and the features of the two users are easily confused. To solve this problem, we propose a position segmentation module to separate the signals of the two users in the azimuth heatmap. Fig. 5 shows the network structure of the proposed position segmentation module. In the position segmentation module, we set up two mask generation networks to extract the signals of the two users separately. The mask generation network module is composed of 2D convolution, and its input channel dimension is 3. The mask generation network contains four convolutional layers. The first two convolutional layers have a kernel size of 3×5 and use padding to keep the size of the signal map unchanged. The last two convolutional layers have a kernel size of 1, which is equivalent to a multi-layer perceptron on the channel dimension to predict the value of the mask at the current pixel. In addition, the output of the last convolutional layer is activated by the Sigmoid activation function to generate a mask with values in the range of 0–1. The original azimuth heatmap is multiplied by the mask to extract the signal of the interested target and filter out the signal of other people. The position segmentation module generates two masks m_1 and m_2 to extract the signals of the two users separately. Through the dual-channel mask generation networks, we hope that the first mask generation network will extract a mask related to the first user and filter out the signal of the second user and vice versa. It is necessary to design some loss function to constrain the output of the position segmentation module:

$$L_m = \text{mean}(m_1) + \text{mean}(m_2) \quad (12)$$

$$L_c = \|\tilde{P}_{ROI} - \tilde{P}_{ROI} \circ m_1 - \tilde{P}_{ROI} \circ m_2\|_2^2 \quad (13)$$

$$L_{dis} = \text{mean}\left(\text{abs}\left(\tilde{P}_{ROI} \circ m_1 - \tilde{P}_{ROI} \circ m_2\right)\right) \quad (14)$$

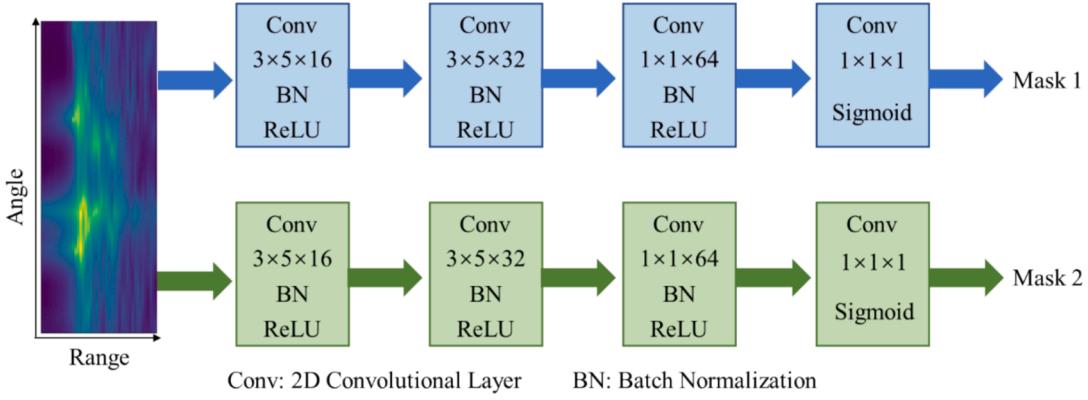


Fig. 5. The network structure of the proposed position segmentation module.

$$L_{dir} = 1 - \frac{\tilde{P}_{ROI}^{\circ} m_1 \bullet \tilde{P}_{ROI}^{\circ} m_2}{\max(\|\tilde{P}_{ROI}^{\circ} m_1\|_2, \|\tilde{P}_{ROI}^{\circ} m_2\|_2, 1e-8)} \quad (15)$$

The network is optimized according to the loss function L_m to make the effective area of the mask generated by the network smaller and extract the area most relevant to the user. The network is optimized according to the loss function L_c , requiring that the sum of the two signal maps filtered by the mask approaches the original signal map. Loss function L_c can constrain the area of the effective area of the mask not to be too small due to the optimization of loss function L_m . L_{dis} calculates the difference in distance between two filtered signals, while L_{dir} calculates the difference in direction between two filtered signals. We subtract these two losses from the loss function, which requires that the two filtered signals be different in distance and direction, helping to reduce the overlap of the mask.

4.2.3. Feature extractor

The proposed feature extractor is a ResNet [39] structure model built using 2D convolution. The ResNet model can alleviate the problem of gradient disappearance and allow us to design deeper networks. The feature extractor contains four residual blocks, a total of 8 2D convolutional layers. The kernel size is 3*3, and the convolution stride of the first three residual blocks is 2*2 to reduce the size of the data. The output channel numbers of the four residual blocks are 32, 64, 128, and 256, respectively.

4.2.4. Predictor

The predictor is a multilayer perceptron [40] with three layers and output feature lengths of 256, 256, and 5, respectively. The prediction results of the prediction layer include five categories: no one, supine, prone, left lateral and right lateral. The label of an input sample can be represented as (L_1, L_2) . In a two – person scenario, we set the L_1 of the sample as the posture label of the first user and the L_2 as the posture label of the second user. In a single – person scenario, we set L_1 as the user's posture label and L_2 as "no one." We use cross-entropy as the classification loss of the model:

$$L_{cls} = \sum_c -[y_c^1 \bullet \log(p_c^1) + y_c^2 \bullet \log(p_c^2)] \quad (16)$$

Where y_c^1 is the sleep posture label of the first user, y_c^2 is the sleep posture label of the second user, p_c^1 is the prediction of the first predictor, and p_c^2 is the prediction of the second predictor. The total loss of the model is:

$$L_{total} = L_{cls} + \alpha^* L_m + \beta^* L_c - \gamma^* (L_{dis} + L_{dir}) \quad (17)$$

Where α , β , and γ are hyperparameters that control the weight of the loss function, and their values are set based on experience. In this paper, the values of all three hyperparameters are 0.5.

4.3. Data augmentation

Deep learning requires rich and large amounts of data to train models. However, due to the limited number of human subjects, human resources, and time costs, datasets often can only cover some situations. The body shape, sleeping position, and sleeping posture of new users may differ from those of training set users, making it difficult for the model to generalize to new users, resulting in low accuracy when new users use the model. To solve this problem, we propose six data augmentation methods, including Range Shift (RS), Angle Shift (AS), Angle Flip (AF), Angle-Warping (AW), Location-Based Range Shift (LB-RS), and Location-Based Angle Shift (LB-AS). LB-RS and LB-AS locate the positions of the two users and perform data augmentation on the signals of the two users separately. The results of data augmentation are shown in Fig. 6. We use combinations of six data augmentation methods to generate new training data.

4.3.1. Range Shift (RS) and angle Shift (AS)

RS is achieved by moving the azimuth heat map along the positive and negative directions on the distance axis, with the moving distance being a random value between 0 and the set maximum RS distance. AS is achieved by moving the azimuth heat map along the angle dimension on the angle axis, with the moving angle being a random value between 0 and the set maximum AS angle. RS aims to simulate signals at different distances, while AS aims to simulate signals at different angles.

4.3.2. Angle Flip (AF)

AF simulates mirrored postures by flipping the azimuth heat map along the angle axis. AF can double the amount of data. After processing the azimuth heat map with AF, the labels of the samples need to be changed. The labels for supine and prone do not need to be changed. The label for the left lateral needs to be changed to the label for the right lateral, and the label for the right lateral needs to be changed to the label for the left lateral. AF simulates richer positions and postures of people on the bed. For example, AF simulates signals where two subjects on the bed swap positions and signals with different postures.

4.3.3. Angle-Warping (AW)

AW is achieved by randomly shifting the angle signals of each range. The angle signals corresponding to each range represent the angle distribution at different distances. The head of a person is closest to the radar, while the feet are farthest from the radar. AW aims to simulate signals after the angle change of different body parts. For example, when a person's legs are placed to the left and right, their angle relative to the radar also changes. The moving angle used in this paper is a random value between 0 and the set maximum AW angle. The maximum AW angle is less than the maximum AS angle because the range of movement of body parts is smaller than the range of movement of people on the

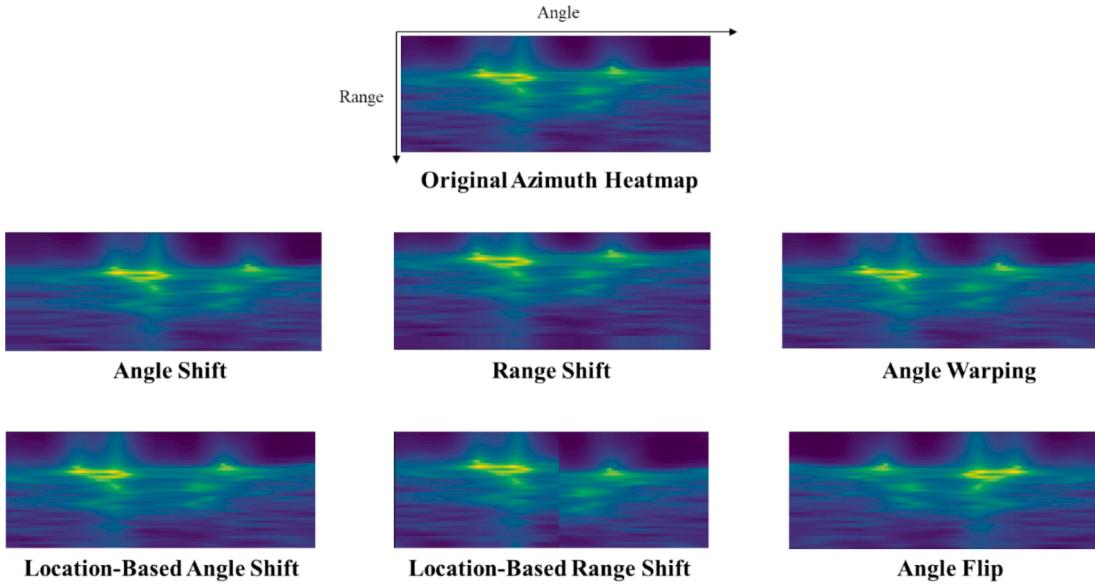


Fig. 6. The effects of the proposed data augmentation methods.

bed.

4.3.4. Location-Based range Shift (LB-RS) and Location-Based angle Shift (LB-AS)

LB-RS and LB-AS locate the angles corresponding to the two users and enhance the data of the two users separately, with the midpoint of the angles of the two users as the boundary. Fig. 7 shows the calculation process of LB-RS and LB-AS. First, we use 2D-CFAR to detect the position of the human body [41]. Second, we use the DBSCAN clustering algorithm to cluster the detected points [42]. When the number of clusters is one or positioning fails, data augmentation is not performed. When two or more clusters are detected, the angles of the center points of the two clusters with the largest interval are used as the positions of the two users. Third, take the midpoint of the positions of the two users as the dividing line and perform RS and AS on each user's signal separately.

5. Experimental results and discussion

5.1. Metrics

We use the F1 score to evaluate the performance of the system. The F1 score takes into account both the recall and precision of the model and is an effective indicator for evaluating the model. We calculate the F1 score for each category and calculate the average of all category F1 scores. The calculation of the F1 score is as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (18)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (19)$$

$$\text{F1 score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (20)$$

5.2. Evaluation setting

In the 13 experiments, there are seven two-person scenario experiments and six single-person scenario experiments, with different subjects in each experiment. In order to evaluate the performance of the model in new users, we use a 13-fold cross-validation method, using data from 12 experiments to train the model and data from the remaining experiment to test the model. The data of the i -th user is $D^i = (D_1^i, D_2^i)$, where D_1^i is pre-collected data and D_2^i is test data. For single-person scenarios, D_1^i contains 16 samples, 4 samples for each posture. For two-person scenarios, D_1^i contains 64 samples, 4 samples for each posture combination. We tested the performance of the model when calibrating the model using the pre-collected data D_1^i of the test users. When calibrating, we added the D_1^i data of the test users to the training set. In the case of no calibration, the test user's data does not participate in training.

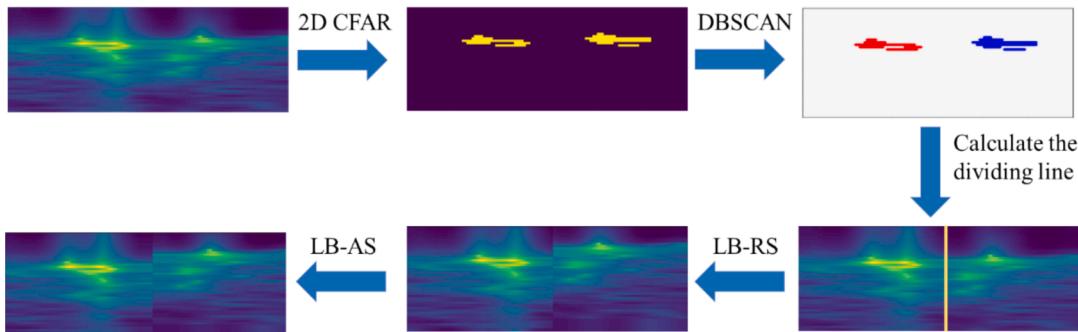


Fig. 7. The calculation process of LB-RS and LB-AS.

5.3. Evaluation of mm2Sleep's performance

Table 2 shows the model's performance on whether or not to use the data of new users to calibrate the model. We evaluated the F1 scores of four sleep postures and the average F1 score. Please note that when calibrating the model, we added pre-collected data D_1^i from test user data $D^i = (D_1^i, D_2^i)$ to the training data to train the model. In the setting without calibration, the average F1 score of the model is 0.874. In the calibration setting, the average F1 score of the model is 0.912, which is only 0.038 higher than the model without calibration. The table shows that the proposed model has high generalization and only slightly degraded performance when not using new user data to calibrate the model.

Fig. 8 shows the confusion matrix of the model when calibrating and not calibrating the model. The confusion matrix shows the average results of cross-validation. The figure shows that the supine posture is easily confused with the prone posture, and the left lateral posture is easily confused with the right lateral posture. The radar signal has a lower resolution than the camera, and the model may give an incorrect judgment when two postures are close. Millimeter-wave radar cannot obtain fine-grained pictures of the human, and it is even difficult to obtain accurate body contours, so different sleep postures are easily confused. After using the data of new users to calibrate the model, the confusion of sleep postures is improved. One reason is that the model has learned the knowledge of new users and is better at recognizing new user sleep postures.

Table 2 shows the results of the ablation experiment of the model. The table shows the model's performance before and after using or removing the position embedding and position segmentation modules. First, we tested the model's performance after only removing the position embedding. In the setting where only the position embedding is removed, the input to the position segmentation module is a radar azimuth heatmap that does not stack coordinate matrices. Second, we tested the model's performance after only removing the position segmentation module. We used the azimuth heatmap after position embedding as the input to the model. Third, we tested the model's performance after simultaneously removing both position embedding and position segmentation modules.

Table 3 shows that after removing both components simultaneously, the average F1 score of the model decreased by 0.022. Using position embedding in conjunction with the position segmentation module can learn masks containing human body position information in azimuth heatmaps to extract signals from each user separately, reducing signal confusion. In the setting where only position embedding is removed, the model's performance decreased by 0.045. Due to the lack of positional information, the quality of masks generated by the position segmentation module decreased, and the model could not distinguish between two users, resulting in decreased performance. In the setting where only the position segmentation module is removed, the model's performance decreased by 0.03. Due to the lack of the position segmentation module, signals from two users were confused in extracted features. In addition, samples after position embedding contain additional positional information, which is not helpful for posture recognition, and model performance is slightly lower than when not using position embedding and position segmentation modules. The position embedding and position segmentation modules work together to generate masks for each user to filter the raw azimuth heatmap rather than filtering the azimuth heatmap after position embedding. The filtered signal contains only one user

signal and does not contain positional coordinate information to minimize interference from other users and positional factors on the model.

Fig. 9 shows the masks and filtered signals output by the model when using position embedding and not using position embedding. After using position embedding, the mask can successfully identify the areas of two users, with almost no overlapping areas. When not using position embedding, the mask cannot identify the areas of two users. In addition, not using position embedding will cause the model to learn incorrect masks, resulting in a loss of information in the filtered signal and a decrease in model performance.

Table 4 shows the results of the ablation experiment of the data augmentation methods. The table shows the model's performance after using or removing one or several proposed data augmentation methods. First, we tested the model's performance using only one proposed data augmentation method to evaluate the performance improvement brought by each data augmentation method alone. Second, we tested the performance improvement by combining two or more data augmentation methods. **Table 4** shows that AF improved by 0.146 in the F1 score, which is the most helpful data augmentation method for performance improvement. LB-AS brought the smallest performance improvement, increasing the F1 score by 0.03. All data augmentation methods are helpful for performance improvement. In tests using a single data augmentation method, AF delivered the largest performance gain, meaning that AF is able to simulate richer sleep postures. RS, AS, and AW all enhance the data by shifting the signal in the distance dimension and the angle dimension, but do not simulate new user postures. AF, by flipping the angle, can simulate left-side sleep as right-side sleep, and right-side sleep as left-side sleep. As a result, AF can generate richer information about the sleeping postures, thus providing the greatest performance improvement to the model. In addition, using a combination of multiple data augmentation methods to train the model will further improve the model's performance, and the performance is higher than using only one data augmentation method. The performance of the model after using both AS and RS is better than the performance of the model after using only AS and RS, which means that combining the two methods can enhance the data more effectively. On this basis, the performance of the model is further improved after using AS, RS, LB-AS, and LB-RS at the same time, which means that LB-AS and LB-RS can further enrich the data by operating on the signal area of a single user. The model trained using all data augmentation methods has an F1 score increase of 0.194 compared to a model that does not use any data augmentation. The huge improvement in performance means that the proposed data augmentation methods effectively simulate richer sleep postures, increase the diversity of data, and therefore improve the model's generalization.

We compared the proposed method with other RF-based sleep posture recognition systems. Since there is no public dataset, we cannot test our system on a public dataset. In addition, the proposed method is the first dual-person sleep posture recognition model, and it is challenging to reproduce other single-person recognition models on our dataset. Therefore, we compare the performance reported in other research with the proposed model. We compare whether new user data calibration is used, whether it is a single-person scene and the reported performance. In the comparison, we separately calculated the performance of the proposed model in the single-person scene and the accuracy in the dual-person scene.

As shown in **Table 5**, the proposed model is the first to solve the problem of dual-person sleep posture recognition. When calibrating the model with test user data, the accuracy of TagSleep [15] in the single-person scene is 98.9 %, the accuracy of BodyCompass [16] is 83.7 %, the accuracy of the model in [31] is 100.0 %, the accuracy of the model in [30] is 98.4 %, and the accuracy of the proposed method is 96 %. Without calibrating the model with new user data, the accuracy of TagSleep drops to below 40 %, while the proposed method only has a slight performance decline, indicating that the proposed method has high generalization. In addition, the proposed method is the first sleep

Table 2
The F1 score of the model in an experiment with or without calibration.

Posture Method	Supine	Prone	Left	Right	Mean
Calibration	0.917	0.901	0.926	0.903	0.912
Not Calibration	0.878	0.872	0.888	0.859	0.874

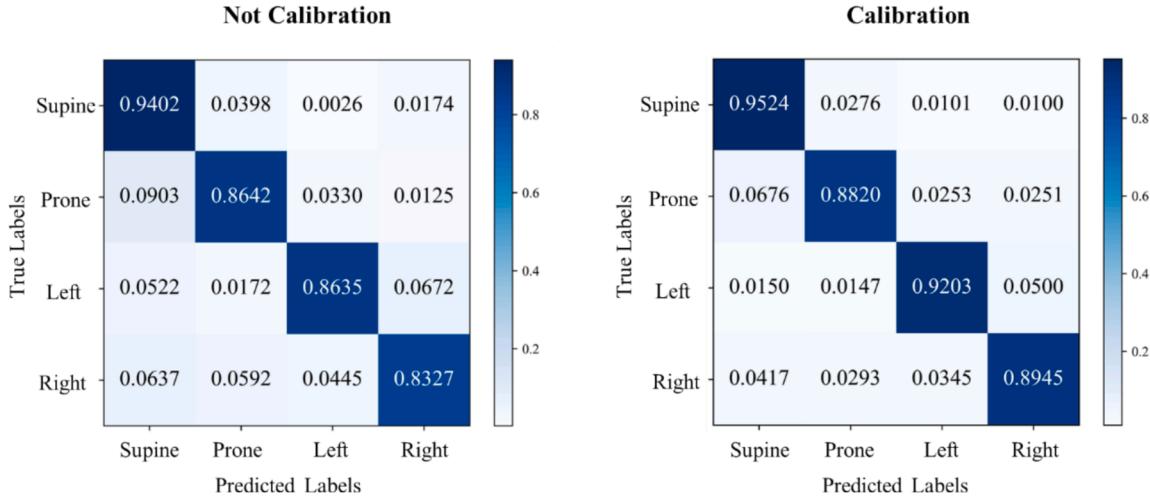


Fig. 8. The confusion matrix of the model when calibrating and not calibrating the model.

Table 3
F1 score of the model in ablation experiments.

Posture Method	Supine	Prone	Left	Right	Mean
Full Model	0.878	0.872	0.888	0.859	0.874
w/o M1	0.867	0.823	0.825	0.801	0.829
w/o M2	0.865	0.837	0.846	0.829	0.844
w/o M1 and M2	0.875	0.852	0.852	0.830	0.852

* M1: Position embedding, M2: Position segmentation module.

posture recognition model that supports dual-person scenes. In the dual-person scene, the proposed method can achieve an accuracy of 83.5 % in new users.

When the two targets are close to each other, there may be no obvious dividing line between the signals of the two users in the radar azimuth heat map. In view of the above situation, we test the performance of the model when the subjects are close to each other. In this experiment, we require two users to be close to each other in the process of data collection, and the bodies of the two users need to touch. We asked users to randomly change their sleep posture after each sample, and a total of 200 samples were collected. We use these 200 samples to test the performance of the model when two users are close to each other.

Table 6 shows the experimental results. When the two subjects were close to each other, the F1 scores of the proposed model for supine, prone, left and right recumbent were 0.780, 0.682, 0.731 and 0.752 respectively, and the average F1 score was 0.736. Compared with the experimental results without limiting the distance between users, the F1 score of the proposed model decreased by 0.097 when two users were close together. When the users are close to each other, the signals of the

two users in the radar signal map of some samples are aliased, so it is difficult to separate them correctly. In the future work, we will study the region of interest separation model when two users are close together, so as to further improve the performance of the model when the subjects are close to each other.

This paper explores a two-person sleeping posture recognition task without using data from new subjects to calibrate the model. In practical applications or subsequent research, we believe that there are two methods that are promising to further improve the accuracy of the model. First of all, the cost of radar sensors can be increased, and radar sensors with more transmitting and receiving antennas can be used to obtain higher-resolution radar azimuth heat maps, thus improving the model's performance. Second, it is possible to combine the respiratory signal with the radar azimuth heat map used in this paper and use the

Table 4
The F1 score of the model in the ablation experiment of data augmentation method.

Posture Method	Supine	Prone	Left	Right	Mean
No data augmentation	0.738	0.689	0.638	0.654	0.680
RS(A1)	0.814	0.723	0.716	0.794	0.762
AS(A2)	0.809	0.719	0.707	0.724	0.740
AF(A3)	0.841	0.807	0.863	0.793	0.826
AW(A4)	0.802	0.670	0.703	0.676	0.713
LB-AS(A5)	0.777	0.682	0.666	0.715	0.710
LB-RS(A6)	0.815	0.737	0.738	0.759	0.762
A1 + A2	0.833	0.791	0.768	0.781	0.793
A5 + A6	0.861	0.762	0.759	0.748	0.783
A1 + A2 + A5 + A6	0.852	0.783	0.785	0.838	0.814
A1 + A2 + A3 + A4 + A5 + A6	0.878	0.872	0.888	0.859	0.874

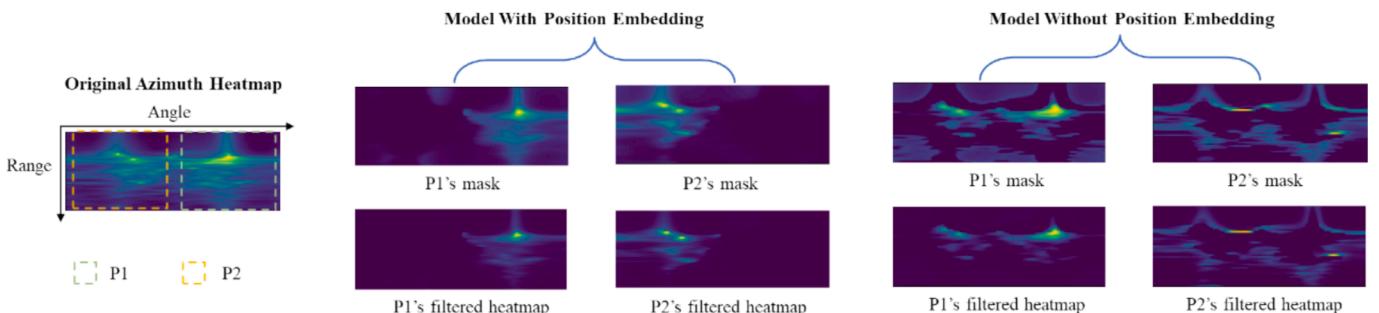


Fig. 9. Masks and filtered signals output by the model when using and not using position embedding.

Table 5

Comparison of the proposed method with other sleep posture recognition studies based on RF.

Method	Support for two-person scenario?	Accuracy (One-person scenario)		Accuracy (Two-person scenario)	
		Calibration	Not calibration	Calibration	Not calibration
Kiriazi <i>et al.</i> [31]	No	100.0 %	N/A	N/A	N/A
Islam <i>et al.</i> [30]	No	98.4 %	N/A	N/A	N/A
TagSleep [15]	No	98.9 %	Less than 40 %	N/A	N/A
BodyCompass [16]	No	83.7 %	N/A	N/A	N/A
mm2Sleep	Yes	96.0 %	91.5 %	86.4 %	83.5 %

Table 6

The F1 score of the model at different distances between subjects.

Posture Distance	Supine	Prone	Left	Right	Mean
D1	0.859	0.838	0.865	0.768	0.833
D2	0.780	0.682	0.731	0.752	0.736

*D1: Normal distance, D2: The subjects were close to each other.

complementary characteristics of the information contained in the two signals to improve the model's performance.

Sleeping posture recognition systems are designed to monitor the subject while the body is in a static state. In our previous research, we introduced an automatic body motion detection algorithm that effectively identifies human motion [43]. In practical implementation, the sleeping posture recognition system first detects body movement, then employs the sleeping posture recognition model proposed in this paper to monitor when the user is in a static state. Additionally, when the system detects human movement again or after a certain time interval, the sleeping posture detection model proposed in this paper is used to determine whether the user's sleeping posture has changed.

6. Conclusion

This study proposes a dual-person sleep posture recognition model mm2Sleep. The model is a dual-channel deep learning model, using a single FMCW radar to recognize the sleep posture of two people simultaneously. To our knowledge, this is the first model that enables high-precision sleeping posture recognition on new subjects without the need to collect their data for calibration. We propose 6 data augmentation methods to enhance the diversity of data. In particular, we propose a method to perform data augmentation on two users separately. The proposed data augmentation methods effectively improve the accuracy and generalization of the model. Tests on a dataset containing 20 subjects show that the proposed method achieves an F1 score of 0.874 for sleep posture recognition in new users. We believe this work as an easy-to-use dual-person sleep posture recognition system can help doctors and patients better manage their condition.

CRediT authorship contribution statement

Yicheng Yao: Writing – original draft, Validation, Methodology, Investigation, Formal analysis, Conceptualization. **Hao Zhang:** Visualization. **Pan Xia:** Data curation. **Changyu Liu:** Investigation. **Fanglin Geng:** Investigation, Data curation. **Zhongrui Bai:** Methodology. **Lidong Du:** Supervision. **Xianxiang Chen:** Supervision, Investigation. **Peng Wang:** Data curation. **Weifeng Yao:** Resources, Investigation. **Ziqing Hei:** Resources, Investigation. **Zhen Fang:** Writing – review & editing, Supervision, Project administration, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China [grant number 62071451, 62331025, U21A20447], and CAMS Innovation Fund for Medical Sciences [grant number 2019-I2M-5-019].

Data availability

Data will be made available on request.

References

- [1] X. Li, Y. Gong, X. Jin and P. Shang, "Sleep posture recognition based on machine learning: A systematic review", *Pervasive and Mobile Computing*, p. 101752, 2023.
- [2] A. Oksenberg, D.S. Silverberg, The effect of body posture on sleep-related breathing disorders: facts and therapeutic implications, *Sleep Med. Rev.* 2 (3) (1998) 139–162.
- [3] C. Gorecki, S.J. Closs, J. Nixon, M. Briggs, Patient-reported pressure ulcer pain: a mixed-methods systematic review, *J. Pain Symptom Manage.* 42 (3) (2011) 443–459.
- [4] J.A. Liebenthal, S. Wu, S. Rose, J.S. Ebersole, J.X. Tao, Association of prone position with sudden unexpected death in epilepsy, *Neurology* 84 (7) (2015) 703–709.
- [5] F. Deng, et al., Design and Implementation of a Noncontact Sleep Monitoring System Using Infrared Cameras and Motion Sensor, *IEEE Trans. Instrum. Meas.* 67 (7) (July 2018) 1555–1563, <https://doi.org/10.1109/TIM.2017.2779358>.
- [6] T.H. Kim, S.J. Kwon, H.M. Choi, Y.S. Hong, Determination of lying posture through recognition of multilevel body parts, *Wirel. Commun. Mob. Comput.* (2019).
- [7] J.A. Vitale, G. Banfi, A. Galbiati, L. Ferini-Strambi, A. La Torre, Effect of a night game on actigraphy-based sleep quality and perceived recovery in top-level volleyball athletes, *Int. J. Sports Physiol. Perform.* 14 (2) (2019) 265–269.
- [8] C. Armon, A. Roy and W. Nowack, "Polysomnography: Overview and clinical application", *E-Medicine*, March, 2007.
- [9] Y.-T. Peng, C.-Y. Lin, M.-T. Sun, C.A. Landis, Multimodality Sensor System for Long-Term Sleep Quality Monitoring, *IEEE Trans. Biomed. Circuits Syst.* 1 (3) (Sept. 2007) 217–227, <https://doi.org/10.1109/TBCAS.2007.914481>.
- [10] S. M. Mohammadi, M. Alnowami, S. Khan, D. -J. Dijk, A. Hilton and K. Wells, "Sleep Posture Classification using a Convolutional Neural Network," 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 2018, pp. 1-4, doi: 10.1109/EMBC.2018.8513009.
- [11] H. Sadreazami, M. Bolic, S. Rajan, Contactless Fall Detection Using Time-Frequency Analysis and Convolutional Neural Networks, *IEEE Trans. Ind. Inf.* 17 (10) (Oct. 2021) 6842–6851, <https://doi.org/10.1109/TII.2021.3049342>.
- [12] Y. Tian, G.H. Lee, H. He, C.Y. Hsu, D. Katabi, RF-based fall monitoring using convolutional neural networks, *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.* 2 (3) (2018) 1–24.
- [13] W. Xia, Y. Li, S. Dong, Radar-Based High-Accuracy Cardiac Activity Sensing, *IEEE Trans. Instrum. Meas.* 70 (2021) 1–13.
- [14] V.G. Rizzi Varela, D.V.Q. Rodrigues, L. Zeng, C. Li, Multitarget Physical Activities Monitoring and Classification Using a V-Band FMCW Radar, *IEEE Trans. Instrum. Meas.* 72 (2023) 1–10.
- [15] C. Liu, J. Xiong, L. Cai, L. Feng, X. Chen, D. Fang, Beyond respiration: Contactless sleep sound-activity recognition using RF signals, *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.* 3 (3) (2019) 1–22.
- [16] S. Yue, Y. Yang, H. Wang, H. Rahul, D. Katabi, BodyCompass, *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.* 4 (2) (2020) 1–25.
- [17] C. Hsia, K. Liou, A. Aung, V. Foo, W. Huang, J. Biswas, in: *Analysis and Comparison of Sleeping Posture Classification Methods Using Pressure Sensitive Bed System*, IEEE, 2009, pp. 6131–6134.
- [18] K.M. Chang, S.H. Liu, Wireless portable electrocardiogram and a tri-axis accelerometer implementation and application on sleep activity monitoring, *Telemed. e-Health* 17 (3) (2011) 177–184.
- [19] J.J. Liu, W. Xu, M.C. Huang, N. Alshurafa, M. Sarrafzadeh, N. Raut, B. Yadegar, Sleep posture analysis using a dense pressure sensitive bedsheet, *Pervasive Mob. Comput.* 10 (2014) 34–50.
- [20] X. Li, Y. Gong, X. Jin, and P. Shang, "Sleep posture recognition based on machine learning: A systematic review," *Pervasive Mobile Computing*, vol. 90, p. 101752, 2023.

- [21] S. Liu, S. Ostadabbas, A vision-based system for in-bed posture tracking, in: In Proceedings of the IEEE International Conference on Computer Vision Workshops, 2017, pp. 1373–1382.
- [22] S. Akbarian, G. Delfi, K. Zhu, A. Yadollahi, B. Taati, Automated Non-Contact Detection of Head and Body Positions During Sleep, *IEEE Access* 7 (2019) 72826–72834, <https://doi.org/10.1109/ACCESS.2019.2920025>.
- [23] T. Grimm, M. Martinez, A. Benz, R. Stiefelhagen, in: Sleep Position Classification from a Depth Camera Using Bed Aligned Maps, *IEEE*, 2016, pp. 319–324.
- [24] S.Z. Guruz, M.G. Amin, Radar-Based Human-Motion Recognition With Deep Learning: Promising Applications for Indoor Monitoring, *IEEE Signal Process Mag.* 36 (4) (2019) 16–28.
- [25] H. Hong, et al., Microwave Sensing and Sleep: Noncontact Sleep-Monitoring Technology With Microwave Biomedical Radar, *IEEE Microw. Mag.* 20 (8) (Aug. 2019) 18–29, <https://doi.org/10.1109/MMM.2019.2915469>.
- [26] L. Ren, H. Wang, K. Naishadham, O. Kilic, A.E. Fathy, Phase-Based Methods for Heart Rate Detection Using UWB Impulse Doppler Radar, *IEEE Trans. Microw. Theory Tech.* 64 (10) (Oct. 2016) 3319–3331, <https://doi.org/10.1109/TMTT.2016.2597824>.
- [27] X. Liu, J. Cao, S. Tang, J. Wen, in: Wi-Sleep: Contactless Sleep Monitoring via WiFi Signals, *IEEE*, 2014, pp. 346–355.
- [28] J. Liu, Y. Chen, Y. Wang, X. Chen, J. Cheng, J. Yang, Monitoring Vital Signs and Postures During Sleep Using WiFi Signals, *IEEE Internet of Things Journal* 5 (3) (June 2018) 2071–2084, <https://doi.org/10.1109/JIOT.2018.2822818>.
- [29] M. Piriyajitakonkij, et al., SleepPoseNet: Multi-View Learning for Sleep Postural Transition Recognition Using UWB, *IEEE J Biomed Health Inform* 25 (4) (Apr 2021) 1305–1314.
- [30] S.M.M. Islam, V.M. Lubecke, Sleep Posture Recognition With a Dual-Frequency Microwave Doppler Radar and Machine Learning Classifiers, *IEEE Sens. Lett.* 6 (3) (2022) 1–4.
- [31] J.E. Kiriazi, S.M.M. Islam, O. Boric-Lubecke, V.M. Lubecke, Sleep Posture Recognition With a Dual-Frequency Cardiopulmonary Doppler Radar, *IEEE Access* 9 (2021) 36181–36194.
- [32] D.-K.-H. Lai, et al., Vision Transformers (ViT) for blanket-penetrating sleep posture recognition using a triple ultra-wideband (UWB) radar system, *Sensors* 23 (5) (2023) 2475.
- [33] B. Luo, Z. Yang, P. Chu, J. Zhou, Human sleep posture recognition method based on interactive learning of ultra-long short-term information, *IEEE Sens. J.* 23 (12) (2023) 13399–13410.
- [34] A. Mumuni, F. Mumuni, Data augmentation: A comprehensive survey of modern approaches, *Array* 16 (2022) 100258.
- [35] J. Gallier, Geometric methods and applications: for computer science and engineering, Springer Science & Business Media, 2011.
- [36] D. P. Kingma, “Auto-encoding variational bayes,” *arXiv preprint arXiv:00148*, 2013.
- [37] I. Goodfellow, et al., Generative adversarial networks, *Commun. ACM* 63 (11) (2020) 139–144.
- [38] J. Capon, High-resolution frequency-wavenumber spectrum analysis, *Proc. IEEE* 57 (8) (Aug. 1969) 1408–1418, <https://doi.org/10.1109/PROC.1969.7278>.
- [39] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
- [40] A. Pinkus, Approximation theory of the MLP model in neural networks, *Acta Numer.* 8 (1999) 143–195.
- [41] L. Daojing and Y. Guilong, “2D-OS-CFAR detector for cloud clutter suppression,” in 2001 CIE International Conference on Radar Proceedings (Cat No. 01TH8559), 2001, pp. 350–353: IEEE.
- [42] K. Khan, S. U. Rehman, K. Aziz, S. Fong, and S. Sarasvady, “DBSCAN: Past, present and future,” in The fifth international conference on the applications of digital information and web technologies (ICADIWT 2014), 2014, pp. 232–238: IEEE.
- [43] H. Zhang, et al., Radar-Beat: Contactless beat-by-beat heart rate monitoring for life scenes, *Biomed. Signal Process. Control* 86 (2023) 105360.