Article

Enhancing Human Activity Recognition with LoRa Wireless RF Signal Preprocessing and Deep Learning

Mingxing Nie ![ID], Liwei Zou, Hao Cui, Xinhui Zhou and Yaping Wan *

School of Computer Science, University of South China, Hengyang 421001, China; niemx@usc.edu.cn (M.N.); 20222008110550@stu.usc.edu.cn (L.Z.); cuihao@stu.usc.edu.cn (H.C.); 20200440916@stu.usc.edu.cn (X.Z.)
* Correspondence: ypwan@usc.edu.cn

Abstract: This paper introduces a novel approach for enhancing human activity recognition through the integration of LoRa wireless RF signal preprocessing and deep learning. We tackle the challenge of extracting features from intricate LoRa signals by scrutinizing the unique propagation process of linearly modulated LoRa signals—a critical aspect for effective feature extraction. Our preprocessing technique involves converting intricate data into real numbers, utilizing Short-Time Fourier Transform (STFT) to generate spectrograms, and incorporating differential signal processing (DSP) techniques to augment activity recognition accuracy. Additionally, we employ frequency-to-image conversion for the purpose of intuitive interpretation. In comprehensive experiments covering activity classification, identity recognition, room identification, and presence detection, our carefully selected deep learning models exhibit outstanding accuracy. Notably, ConvNext attains 96.7% accuracy in activity classification, 97.9% in identity recognition, and 97.3% in room identification. The Vision TF model excels with 98.5% accuracy in presence detection. Through leveraging LoRa signal characteristics and sophisticated preprocessing techniques, our transformative approach significantly enhances feature extraction, ensuring heightened accuracy and reliability in human activity recognition.

Keywords: human activity recognition; LoRa wireless RF signal preprocessing; deep learning; feature extraction; differential signal processing

1. Introduction

Human activity recognition (HAR) has emerged as a vital field with wide-ranging applications, including healthcare, surveillance, sports analysis, elderly care, and human–computer interaction [1]. HAR systems aim to identify and categorize human activities by collecting data from various sensors that capture elements such as acceleration, orientation, pressure, and radio frequency signals [2]. Advanced algorithms and machine learning techniques are then applied to analyze these data, providing insights into user behavior and health.

HAR solutions often depend on physical sensors to continuously collect data, applying pattern recognition algorithms to identify user actions collectively [3]. They can be simply divided into two categories: device-based approaches and device-free approaches. Device-based approaches involve attaching physical sensors to humans and use inertia measurement units to collate the subject's activities [4]. In certain settings, wearable sensors may inconvenience users, causing skin irritation and discomfort. Prolonged use of these sensors can negatively impact the user's quality of life, leading to skin pain and other discomforts. Device-free approaches include methods such as video-based and radio frequency sensing-based (RF-based) methods, which break free from device limitations. Vision-based methods have been among the earliest attempts in recording human behaviors using a camera, utilizing various computer vision techniques to analyze the collected data and to identify different activities [5]. Despite their usefulness, vision-based methods have limitations such as dependence on the environment, complexity, privacy concerns, low light conditions, occlusion, etc.

RF-based methods use radio signals and their reflections to capture the surrounding details, and they then apply artificial intelligence algorithms to sense objects [6]. RF signals, which are highly sensitive to environmental changes, can effectively detect alterations in the surroundings caused by human activities. When a target person moves for a certain activity, the human body selectively absorbs, scatters, or reflects RF waves, thus causing signal delay extension and temporal dispersion to fluctuate at the receiver. The amplitude, phase, or frequency of the RF signal will change accordingly in different motion modes. The change in signal quality is harmful to communication, but it may enhance the RF signal characteristic curve and introduce information bursts for activity recognition as a positive side effect. Therefore, the time-varying RF signal is affected by physical movement and thus can be used for monitoring purposes [7]. RF-based HAR uses RF signals to detect human activities, making it a non-invasive approach that does not require any physical contact with the human body. It is capable of providing a continuous monitoring of human activity, which is particularly useful for healthcare applications. Additionally, it is relatively inexpensive compared to other approaches, such as vision-based or sensor-based methods. This makes it an ideal choice for portable designs, and it is particularly suitable for remote applications and field studies.

Although RF-based HAR has excellent characteristics, it still faces challenges such as limited range, interference from other devices, non-line-of-sight propagation, sensitivity to environmental conditions, high cost, and complexity. LoRa (long range radio) is a low-power wireless communication technology that can address many of the challenges faced by traditional RF-based sensing. Firstly, LoRa technology offers a longer range of wireless signal than traditional RF-based HAR, which helps to overcome the limited range issue. It can provide coverage of up to several kilometers, making it well suited for large-scale sensing applications. Secondly, LoRa technology operates using an unlicensed frequency spectrum, which is less prone to interference from other wireless devices. This results in a better reliability and accuracy of sensing, even in congested wireless environments. Thirdly, LoRa-based devices consume less power when transmitting data, making them ideal for continuous monitoring applications where battery life is critical. Finally, LoRa technology is accessible to a wider audience of researchers and practitioners as it is relatively low cost and easy to implement.

However, despite its tremendous promise, LoRa faces specific challenges in the context of HAR, including limited Doppler sensitivity, localization precision, and signal penetration. Previous studies have shown limitations in utilizing LoRa technology for HAR. Recognizing these challenges and gaps in the existing research, this paper presents a novel LoRa-based device-free human activity recognition system that integrates RF sensing with deep learning techniques, thereby addressing these limitations.

Before delving into the details of our proposed system, it is crucial to acknowledge the challenges and gaps that exist in current RF-based HAR methodologies. The limitations include, but are not limited to, the issues of limited range, interference, non-line-of-sight propagation, and sensitivity to environmental conditions. These challenges hinder the widespread adoption of RF-based HAR systems in practical scenarios, particularly in large-scale deployments and real-world applications.

In this paper, we systematically analyze these challenges and present a comprehensive overview of the limitations associated with the use of LoRa technology in human activity recognition. Our work aims to bridge these gaps by introducing novel techniques that enhance the capabilities of LoRa-based systems, thereby making them more robust and applicable in diverse environments.

The main contributions of this paper are summarized as follows:

(1). A Novel LoRa Wireless RF Signal Preprocessing Technique: We introduce a unique preprocessing approach specifically tailored for LoRa wireless RF signals in the context of HAR. This technique enhances the correlation between signals and specific human activities, making the data more suitable for training deep learning models. By addressing the challenge of distinguishing subtle body movements, our preprocessing

step significantly improves the reliability and accuracy of our activity monitoring system, especially when dealing with limited signal diversity.

(2). Pioneer Integration of LoRa Wireless Technology with Deep Learning: we pioneer the integration of LoRa wireless communication technology with advanced deep learning techniques. Our system transforms activity monitoring by eliminating the need for wearable devices or cameras, providing a seamless and unobtrusive solution. The combination of LoRa technology with deep learning enables precise and real-time recognition across various activities, thereby representing a transformative departure from traditional methods. This integration not only addresses the underutilization of specific signal processing techniques, but also significantly advances activity recognition, thus opening new possibilities for practical deployment in remote applications, field studies, and large-scale scenarios.

(3). Scalable Deployment and Real-World Applicability: Our approach offers scalability and practicality for real-world deployment. Through leveraging the benefits of LoRa technology and optimized data preprocessing, our system can be easily deployed in various indoor environments. This contributes to the development of cost-effective and scalable solutions for large-scale activity monitoring. Our research addresses the demand for practical, deployable solutions in the field of human activity recognition, thereby increasing the accessibility and applicability of this technology.

In summary, these contributions collectively establish the distinctiveness and significance of our work, providing valuable insights for researchers and practitioners in the dynamic field of LoRa-based human activity recognition.

The remainder of this paper is structured as follows: Section 2 discusses the related work on LoRa-based HAR systems and RF sensing. Section 3 elaborates on our system's architecture and RF sensing approach. Section 4 presents the experimental results and an evaluation of our system's performance. Finally, in Section 5, we provide a discussion and conclusion for our work.

2. Related Works

2.1. RF Sensing in HAR

Several RF-based methods have been explored for the sensing of human movement in diverse contexts, as illustrated in Table 1. One promising technology for indoor environments is WiFi-based RF sensing, which leverages channel state information (CSI) in WiFi signals to detect the changes caused by human motion and to extract meaningful features for activity identification [8]. However, WiFi-based sensing comes with its own set of limitations, including high computational requirements, limited diversity in signal characteristics, and sensitivity to environmental conditions. Notably, its performance in terms of wall penetration and sensing distance can significantly impact the quality of the data collected.

Table 1. Comparison of various wireless protocols.

Wireless Protocol	Sensing Range	The Effect of through Wall	Frequency Band	Deployment
WiFi [8]	15 m	bad	2.4 GHz	General
FMCW radar [9]	8 m	bad	60 GHz	General
RFID [10]	20 m	General	860 MHz	General
Bluetooth [11]	10 m	bad	2.4 GHz	General
Visible light [12]	6 m	bad	400 THz	bad
6G [13]	10 m	Bad	24 GHz	bad
LoRa [14]	30 m	Great	915 MHz	Easy

In contrast, radar-based RF sensing systems offer the advantage of penetrating walls and physical barriers, making them well suited for indoor monitoring [9]. These systems provide precise readings of an individual's movements with minimal interference from

other objects or signals. They are effective for monitoring a wide range of activities, from simple activities like walking and running to subtle body movements. However, the complexity and cost associated with implementing radar-based HAR systems, including the need for specialized hardware and software, can be a drawback.

Another approach uses non-contact radio–frequency identification (RFID) sensing technology, offering sub-millimeter resolutions for measuring vibration amplitudes [10]. However, RFID's range for device-free sensing is limited to the immediate vicinity of the RFID tag, thereby making it suitable for close monitoring but less practical for applications requiring measurements over greater distances.

Bluetooth wireless communication technology has been explored for passive human sensing applications [11]. Nonetheless, Bluetooth-based HAR faces constraints, including a limited transmission range of approximately 10 m, which can result in data loss if the devices move out of range. Additionally, the reliance on a received signal strength indicator (RSSI) for human body sensing may not effectively handle sudden RSSI changes due to frequency hopping.

Some researchers have investigated the use of visible light communication (VLC) to reconstruct a person's 3D skeletal pose in real time [12]. However, VLC has limitations such as a restricted range, which can pose challenges in large indoor areas with obstacles that block light signals.

Looking ahead, the potential integration of 6G communication networks into HAR applications holds promise, with benefits including high data rates, low latency, and improved coverage compared to previous wireless network generations [13]. Wearables equipped with 6G connectivity could enhance HAR's utility in diverse environments. Nonetheless, this technology is still in its early developmental stages and faces challenges related to availability, compatibility, and cost, which may limit its widespread adoption.

As discussed earlier, various RF-based methods and wireless technologies have demonstrated their potential in the field of HAR. However, it is important to acknowledge that these approaches are often limited by their inherent challenges and limitations, particularly in relation to the sensing range.

These sensing range limitations pose challenges in real-world applications that require comprehensive activity monitoring in large or complex environments. An insufficient sensing range may compromise the effectiveness and accuracy of RF-based HAR systems.

2.2. RF Sensing Methods

In this section, we discuss the latest research and developments in device-free human activity recognition methods, focusing on the novel techniques, architectures, and algorithms proposed by researchers in the field.

Yan et al. presented the WiAct system, a passive WiFi-based approach for human activity recognition. Activity data classification was performed using an extreme learning machine (ELM), which was chosen for its strong generalization ability and fast learning speed. The evaluation results reinforced the system's robustness, highlighting its effectiveness in real-world environments [8]. Li et al. proposed a novel WiFi-based HAR method with the Two-Stream Convolution Augmented Human Activity Transformer (THAT) model. This model utilizes a two-stream structure to capture both time-over-channel and channel-over-time features. It employs a multi-scale convolution augmented transformer to capture range-based patterns [15]. Lin et al. presented a system that utilizes a smartphone and an off-the-shelf WiFi router for human activity recognition at various scales. The system achieved high accuracy in recognizing different scales of motions, demonstrating its usefulness in gesture recognition. Comparisons were made with three machine learning models, including the convolutional neural network (CNNs), decision tree, and long short-term memory [16]. Sakorn Mekruksavanich et al. introduced a versatile framework for human activity recognition using WiFi-based sensing and CSI data. Their proposed CNN-GRU-AttNet hybrid deep learning model effectively extracted spatial–temporal features and outperformed conventional deep learning models [17]. Moghaddam et al. introduced an

innovative approach for device-free fine-grained human activity recognition by combining CSI and RSSI signals. The method involves extracting diverse features and utilizing machine learning classifiers such as SVM, GNB, DT, LR, LDA, KNN, and RF to recognize human interactions. The evaluation results showcase remarkable performance [18]. Saw et al. proposed a hybrid framework called the Convolutional Neural Network–Stochastic Reservoir (CNN-SR) for human activity recognition using WiFi signals. Their method involved computing a subcarrier correlation matrix to represent the CSI signal changes induced by human activities [19]. Shrestha et al. proposed a novel approach for continuous activity monitoring and classification using recurrent LSTM and Bi-LSTM network architectures with radar data. Their approach addressed the challenges of seamless motion with dynamic transitions by utilizing a continuous temporal sequence of micro-Doppler or range–time information [20]. Chen et al. presented a novel approach for radar-based human activity recognition using a Temporal Three-Dimensional Convolutional Neural Network (3DCNN). Their model effectively combined time, range, Doppler, and RCS features to analyze range–Doppler frames. Their experimental results showcased the feasibility and strong performance of the method, which incorporated a temporal attention module to capture sequential context [21]. Rafli et al. proposed an alternative radar-based approach for human activity recognition, which employed novel preprocessing techniques and compared various CNN architectures. Their study focused on transforming raw data signals into images containing target distance information [22]. Huang et al. presented the Au-Id system, which utilizes RFID technology and human motions for non-intrusive user identification and authentication in smart spaces. The system accurately captures physical and behavioral characteristics using multi-modal CNN and one-class SVM classifiers [23]. Liu et al. introduced TransTM, a novel method that leverages the time-streaming multiscale transformer for efficient feature extraction from raw RFID RSSI data without intricate pre-processing steps [24].

In summary, the field of human activity recognition has witnessed significant advancements in both WiFi-based and radar-based approaches. Researchers have explored diverse techniques such as machine learning models, deep learning architectures, and the fusion of multiple sensor signals to achieve an accurate and robust recognition of human activities. These approaches have demonstrated their effectiveness in various real-world environments and showcased superior performance compared to conventional methods. The utilization of WiFi signals, CSI data, RFID technology, and radar data has contributed to the development of non-intrusive, device-free, and passive systems for human activity recognition. These advancements hold great potential for applications in healthcare systems, context-aware applications, security monitoring, and smart spaces.

2.3. LoRa Sensing in HAR

Among low-power wide-area network (LPWAN) technologies, LoRa has garnered significant attention in both industrial and research communities for its unique attributes. LoRa, based on chirp spread spectrum (CSS) modulation, is known for its resilience against interference and noise, along with its high sensitivity, which enables it to receive weak signals with minimal power consumption. This results in a significantly improved link budget, thereby enabling wide coverage over several kilometers and making LoRa an ideal solution for long-range sensing applications [25,26].

Initially, LoRa found its niche in enabling the transmission of real-time activity and location data from wearable devices to cloud servers via LoRa gateways [27]. This capability was especially advantageous in scenarios where monitoring devices with sensors needed to operate in remote or outdoor environments, i.e., where traditional communication technologies have struggled due to limited range [28–30].

One notable innovation in the fusion of LoRa with sensing technologies is LoRadar, a novel technique that seamlessly integrates LoRa and frequency-modulated continuous wave (FMCW) radar systems [31]. LoRadar optimizes spectrum utilization by using FMCW radar to modulate LoRa signals, thereby mitigating potential conflicts between communi-

cation and sensing systems. This integration has led to significant improvements in both communication reliability and sensing accuracy. Zhang et al. explored LoRa's intrinsic potential for long-range, device-free sensing through comprehensive theoretical modeling and empirical research [14]. Their work focused on assessing LoRa's sensing capabilities in terms of range and granularity, shedding light on its suitability for sensing applications.

Moreover, Xie et al. introduced innovative signal processing techniques, including target-induced phase variation enlargement and time-domain beamforming, to enhance LoRa's device-free sensing capabilities [32]. These techniques extend LoRa's range for device-free sensing of human activities, reaching distances of 50–120 m for human walking and 75 m for fine-grained respiration sensing. While these advancements have made LoRa-based sensing systems more versatile and user-friendly, they often require custom-designed LoRa gateways. Zhang et al. addressed this challenge by introducing a multi-antenna-based beamforming technique, enhancing signal reception for long-range multi-target respiration sensing with LoRa [33]. However, this approach still necessitates specialized signal processing within LoRa gateways, thus presenting challenges for real-world deployment. In another promising development, Xie et al. integrated LoRa sensing with mobile robots to extend the sensing range, thereby enabling precise respiration monitoring and human walking sensing, even in scenarios involving receiver motion [34]. Nonetheless, the system's reliance on custom signal processing limits its applicability to off-the-shelf commercial LoRa devices.

The pioneering efforts to combine LoRa technology with HAR applications have significantly expanded the possibilities for non-invasive, long-range, and low-power human activity monitoring. However, the need for customized gateways and specialized signal processing remains a challenge, thus setting the stage for our novel LoRa-based device-free human activity recognition system.

2.4. Gaps and Limitations

While LoRa technology holds great promise for device-free human activity recognition, it is crucial to recognize the specific challenges it encounters in this context. These challenges include the following: (1) Weak signal changes—LoRa technology is primarily designed for long-range wireless communication, and its sensitivity to subtle signal changes caused by human activities might be limited. (2) Recognizing multiple activities—previous studies in LoRa-based HAR have predominantly focused on detecting and classifying a few specific activities. However, it remains uncertain whether it can effectively distinguish and recognize a broader range of activities, especially when dealing with similar activities that exhibit high similarity or overlapping characteristics. (3) RF signal preprocessing—the effective preprocessing of RF signals is crucial to assist deep learning models in extracting the relevant features from the weak signals generated by human activities, particularly in the presence of strong signal backgrounds. This limitation stems from the inherent noise and variability in RF signal data.

To bridge this gap and overcome the limitations, our research aims to enhance HAR with LoRa wireless RF signal preprocessing and deep learning. By developing specialized signal processing mechanisms that accentuate signal variations and texture features, we can effectively extract and analyze the features from LoRa RF signals. Additionally, by leveraging recent advancements in deep learning, we can train models to automatically extract relevant features, thus reducing the need for manual feature engineering.

The objective of our study is to optimize activity recognition accuracy and efficiency by exploring the potential of deep learning techniques in combination with LoRa technology. By addressing the challenges of preprocessing LoRa RF signals and harnessing the power of deep learning algorithms, we strive to unlock the full potential of LoRa-based device-free HAR systems.

3. Methodology

3.1. System Architecture

The proposed system aims to advance human activity recognition (HAR) by leveraging LoRa wireless RF signal preprocessing and deep learning techniques. The system architecture is illustrated in Figure 1.
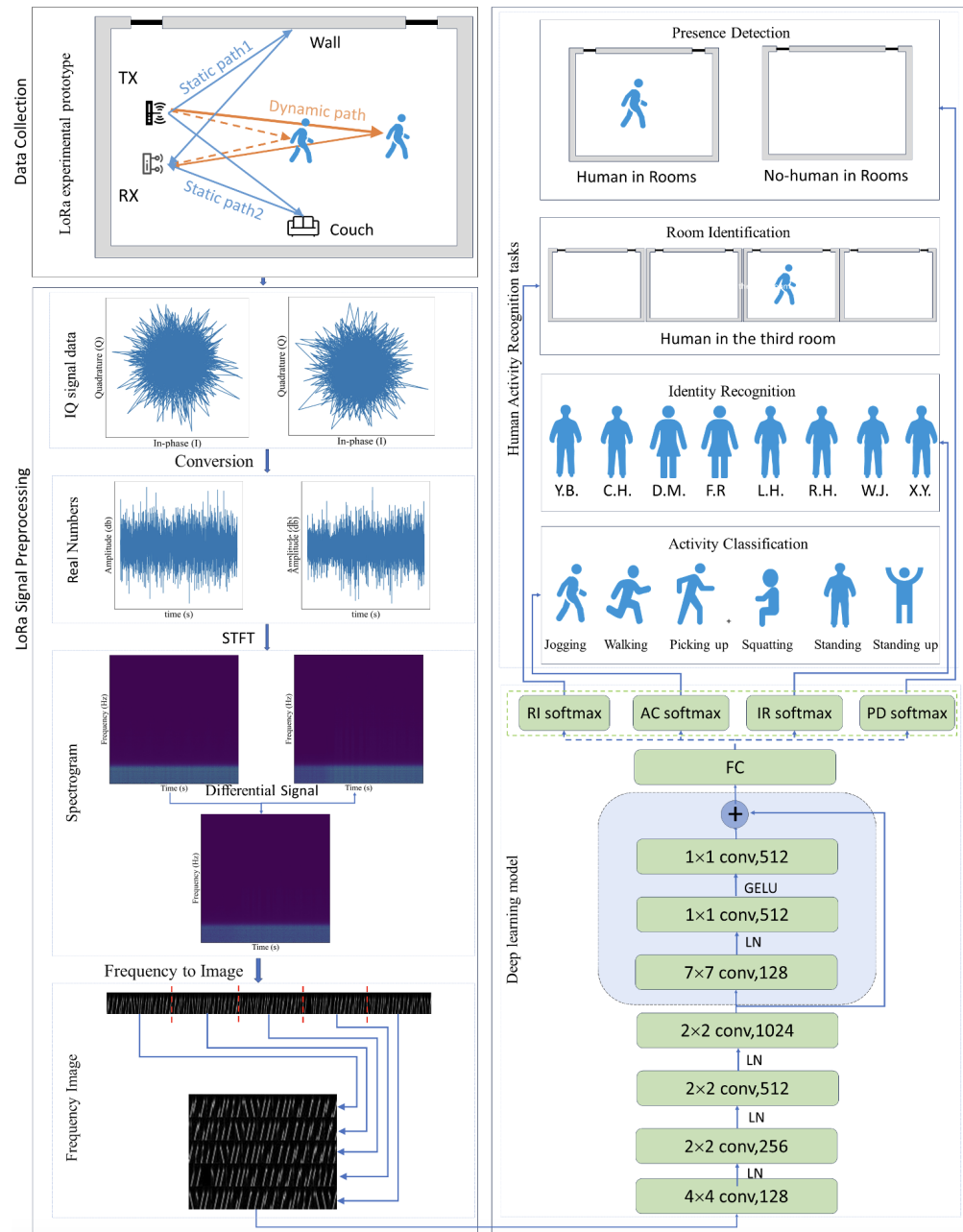

Figure 1. The proposed system architecture.

The system employs off-the-shelf LoRa transceivers for primary data collection, which are strategically positioned within the monitoring environment to capture the RF signals influenced by human activities. These transceivers, capable of transmitting and receiving LoRa signals, serve as primary data collection devices without requiring any custom modifications. To ensure a sufficient coverage of the monitoring area—based on specific requirements such as environment size and layout, as well as desired activity recognition granularity—the LoRa transceivers are strategically positioned. This approach makes the system cost-effective and readily deployable without the need for customization.

On the signal reception end, the system includes a combination of USRP B210 and two directional receiving antennas, thereby forming the signal reception system. This system is connected to a backend processing computer, where the received data is efficiently processed using GNU Radio 3.8.2.0 software. The integration of USRP and directional antennas ensures precise signal reception, thus playing a critical role in the system's data acquisition and processing.

In addition to data collection, the system incorporates enhanced preprocessing techniques to improve the correlation between LoRa wireless RF signals and specific human activities. This preprocessing stage utilizes a combination of methods such as transforming in-phase and quadrature-phase (I/Q) signals into a real number domain, thus generating a continuous time-domain signal sequence. The Short-Time Fourier Transform (STFT) is then applied to analyze the frequency content of the signal over short time intervals, thereby providing valuable insights into the temporal characteristics of the RF signals.

To further enhance the accuracy of activity recognition, the system employs differential signal processing techniques. By computing the difference between consecutive time intervals of the continuous time-domain signal sequence, the system can capture and extract the dynamic changes caused by human activities. This differential processing helps to eliminate static background noise and emphasize the distinctive features associated with human movements. The processed RF signal data is transformed into a spectrogram representation, thereby converting the frequency axis to the image domain. This conversion enables the application of image processing techniques and deep learning algorithms, which are well established in the field of computer vision. By treating the RF signal data as images, the system can leverage the power of convolutional neural networks (CNNs) to extract high-level features from the spectrogram and to perform accurate activity recognition.

By incorporating these specific preprocessing methods, including transforming I/Q signals, employing STFT, applying differential signal processing, and conducting frequency-to-image conversion, the system effectively addresses the challenge of discerning subtle body movements. Additionally, it ensures reliable monitoring even in scenarios with limited signal diversity. These refined preprocessing techniques filter out the impact of strong background signals, enhance the magnitude of signal changes caused by human activities, and establish a robust foundation for accurate activity recognition.

Moreover, to determine the feasibility of directly applying well established deep learning architectures from the field of visual processing to HAR, the system aims to avoid requiring any specific modifications. This approach seamlessly integrates LoRa wireless technology with classic deep learning models, enabling a direct and efficient combination of hardware and software. By exploring this direct application, the system strives to offer a viable solution for activity monitoring that does not rely on wearable devices or cameras. It provides a convenient and non-intrusive means of capturing and identifying human activities.

Furthermore, the system's design places strong emphasis on scalability and adaptability to various indoor environments. It can be deployed cost effectively for large-scale activity monitoring, addressing the need for practical and deployable solutions in the field of human activity recognition. This scalability ensures that the system can cater to a wide array of monitoring scenarios, ranging from small-scale deployments to extensive, complex environments.

In summary, this multifaceted system, which combines LoRa wireless RF signal preprocessing with deep learning, is poised to revolutionize human activity recognition. Its goal is to establish a non-intrusive, cost-effective, and scalable solution that overcomes the limitations of traditional approaches, thereby providing a practical and efficient system for monitoring and identifying human activities in real-world scenarios.

3.2. LoRa Signal Processing

The primary challenge in wireless sensing lies in extracting human activity features from complex LoRa signals and determining specific features to extract. Unlike radar

signals, LoRa signals lack the mixing process of a reflected signal with a transmitted signal, as seen in FMCW radar. This absence of a direct formulaic approach significantly complicates the analysis and extraction of the changing features influenced by human activities. To address this challenge, this chapter includes a detailed analysis of the specific propagation process of linearly modulated LoRa signals, as well as explores the variations caused by human activity so as to extract the relevant features.

In summary, the unique nature of LoRa signals, which is distinct from radar signals, presents a distinct challenge in extracting human activity features due to the absence of a direct formulaic approach. The subsequent section delves into the specific propagation process of linearly modulated LoRa signals by analyzing the variations induced by human activities with the aim of extracting relevant features. This exploration lays the groundwork for a comprehensive understanding of LoRa signal processing, seamlessly guiding us into the subsequent discussion on signal analysis techniques.

3.2.1. Conversion to Real Numbers

In the initial preprocessing phase of the LoRa RF signal for human activity recognition (HAR), the raw input data comprise a complex set of in-phase and quadrature-phase (I/Q) dual channels [35], which are conventionally expressed as complex numbers. To simplify the signal representation and make it more amenable to various signal processing techniques, we transformed these complex numbers into a real number domain, thus generating a continuous time-domain signal sequence.

This transformation from complex to real numbers serves the fundamental purpose of simplifying the signal representation. In this context, complex numbers, consisting of both real and imaginary components, are converted into a real number format, thus effectively generating a continuous time-domain signal sequence.

3.2.2. Short-Time Fourier Transform (STFT)

Building on the conversion to real numbers, the next crucial technique for LoRa RF signal processing is the Short-Time Fourier Transform (STFT) [36]. This technique involves dividing the signal into smaller, overlapping frames and transforming them into a time-frequency representation known as a spectrogram. The resulting spectrogram is a matrix, with the rows representing time and the columns representing frequencies. It accurately captures the signal's spectral composition and its temporal changes, thus providing the foundation for the subsequent feature extraction that is necessary for accurate human activity recognition.

$$STFT(t, f) = \int_{-\infty}^{\infty} x(\tau)w(\tau - t)e^{-j2\pi f\tau}d\tau, \tag{1}$$

where t represents the temporal offset or moment in time for the Short-Time Fourier Transform, f denotes frequency, and $e^{-j2\pi f\tau}$ represents a complex exponential factor used for the frequency domain transformation. $w(\tau - t)$ is a window function, which is typically a function limited within the range $\tau - t$ and is used to truncate the signal for localized spectral analysis.

Figure 2a illustrates a spectrogram displaying the frequency content over time for six different activities. Figure 2b presents a spectrogram that depicts the activity scenario in multiple contiguous rooms. In Figure 2c, the displayed spectrogram represents the LoRa spectrum generated by different individuals for human identity. Lastly, Figure 2d showcases a spectrogram representing scenarios with a human presence in rooms, as well as scenarios without humans.

Each spectrogram visually represents the time–frequency characteristics of the respective data, thus offering valuable insights into the RF signal's composition and behavior in different contexts.
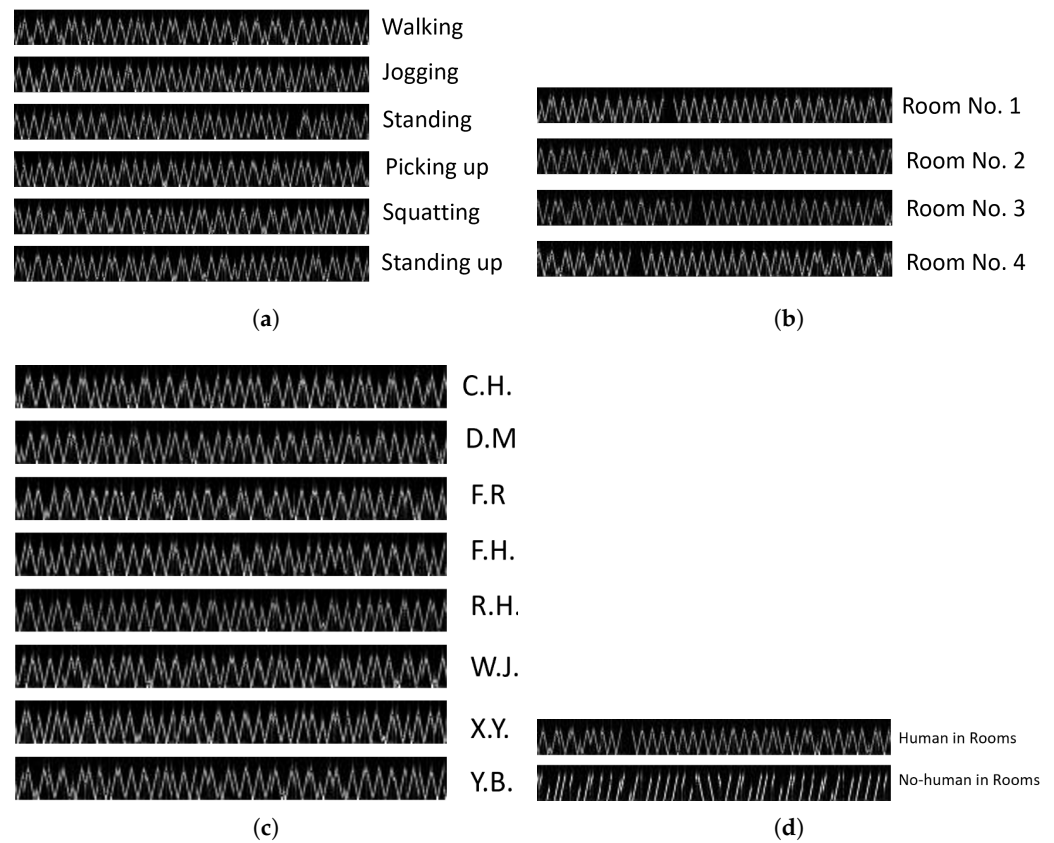
Figure 2. LoRa signal spectrum diagrams in various scenarios. (**a**) Spectrogram for six different activities. (**b**) Spectrogram of the activity scenario in multiple contiguous rooms. (**c**) Spectrogram for human identity. (**d**) Spectrogram for human in rooms and no-human in room scenarios.

3.2.3. Differential Signal Processing (DSP)

To enhance activity recognition accuracy, this study employed differential signal processing (DSP) [14]. This process involves comparing the activity LoRa data with the non-activity LoRa data collected from the same antenna. By subtracting the activity LoRa data from the non-activity LoRa data for each antenna, the DSP technique highlights the distinct signal variations caused by human activities.

If there are no active targets in the environment, the signal received by RX is relatively stable, as shown in Figure 3a. The LoRa signal is primarily obtained through the reflection of static objects (such as walls and sofas) in the sensing area, thus forming a static vector combination. When the objects are active in the sensing area, the signal received by RX consists of the static vector of static objects and the dynamic vector of the target's active reflection. As shown in Figure 3b, the Lora signal is not only reflected by static objects in the sensing area, but the activity of the target also affects the reflection of the Lora signal, thus leading to different changes in the dynamic vector. Therefore, when there is human activity, the propagation path of the LoRa signal in the entire sensing space can be divided into two categories: those unaffected by human activity (Sv) and those affected by human activity (Dv). The signal received at the receiving end in the presence of human activity can be represented as follows:

$$Rx_1(t) = Sv_1 + Sv_2 + Dv_1. \tag{2}$$

On the other hand, in the absence of human movement, the signal received at the receiving end can be represented as follows:

$$Rx_2(t) = Sv_1 + Sv_2 + Sv_3. \tag{3}$$

By comparing the signals received by the two groups of receivers, it is evident that, regardless of whether there is human activity, there will always be a portion of static components Sv_1 and Sv_2 in all propagation paths. This is a significant error term for the dynamic component Dv_1, which is relatively smaller in coverage due to human activity. Although the changes between each group of data were not identical, they theoretically reflect similar characteristics. When finely divided into the absence of activity, the change characteristics of the data on each propagation path were also similar. Therefore, the data collected multiple times in the absence of activity should be interchangeable to some extent, and the static components Sv_1 and Sv_2 that they contain are also interchangeable. At this point, we can obtain, via DSP, the data of human activity from the two sets of data with and without human activity as follows:

$$Rx = Dv_1 - Sv_3. \qquad (4)$$

At this moment, although there is still an error in the static component Sv_3 for the relative dynamic component Dv_1, the influence of the remaining static component Sv_3 on the dynamic component Dv_1 significantly decreases after removing the static components Sv_1 and Sv_2. Due to the additivity of the short-time Fourier transform, the addition and subtraction of signal values can be transformed into the addition and subtraction of instantaneous frequencies without affecting the final representation of instantaneous frequencies. In this way, through a simple subtraction operation, the influence of certain static components can be eliminated, thereby highlighting the impact of human activity on the frequency changes of a LoRa signal. Furthermore, as the change patterns among the same type of activity data are similar, the method of subtracting the data of the same activity from different data without human activity can also be considered as a data augmentation strategy through which to enhance the robustness of data to subtle environmental changes, CFO, and other errors.

Through analyzing the differences between activity and non-activity LoRa data, the DSP technique effectively enhances the signal-to-noise ratio, thus amplifying the activity-related signal patterns. This improvement leads to enhanced accuracy in activity detection and classification within the overall activity recognition system.

The output of the DSP process provides valuable insights into the modulation, amplitude, and other characteristics of LoRa wireless RF signals when associated with specific activities, thereby facilitating the development of robust feature representations and enhancing the effectiveness of subsequent deep learning models in achieving accurate activity recognition.

In summary, the employed differential signal processing technique enhances the discriminative features of human activities, thus improving accuracy in activity recognition by capturing and amplifying the unique signal characteristics associated with different activities.
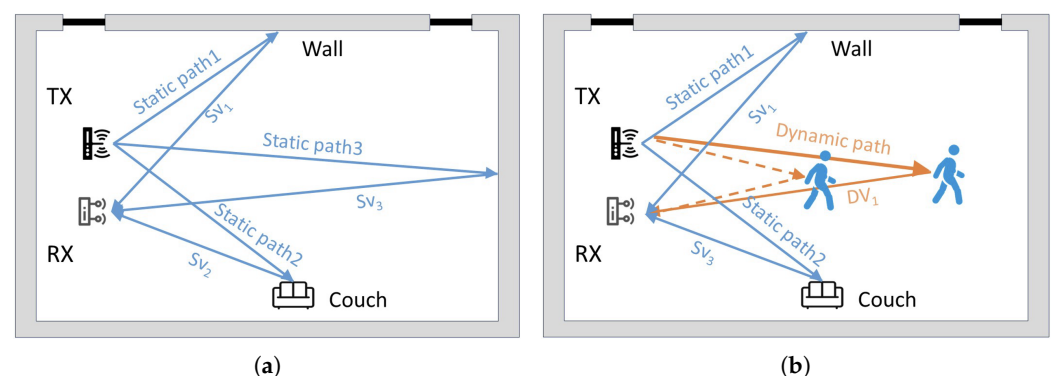

Figure 3. The received signal with/without an activity target in the environment. (**a**) Without activity target. (**b**) With activity target.

3.2.4. Frequency-to-Image Conversion

Frequency-to-image conversion plays a pivotal role in signal processing and deep learning [37], offering advantages such as improved visualization of frequency data and efficient pattern recognition. This transformative process enhances the interpretability of signals for subsequent deep learning tasks. In our approach, frequency-to-image conversion follows the acquisition of the frequency domain representation, which involves mapping the frequency data to a pixel grid. Each pixel's intensity or color corresponds to the magnitude of the associated frequency component.

To address the sequence's imbalance, which is stored as a two-dimensional matrix, we propose a reverse approach inspired by the vision transformer network. Consecutive frames in the sequence were concatenated to form consistent-length matrices, which were treated as patches, as illustrated in Figure 4. These matrices were then vertically concatenated to yield a balanced two-dimensional matrix, and this was subsequently saved in image format.
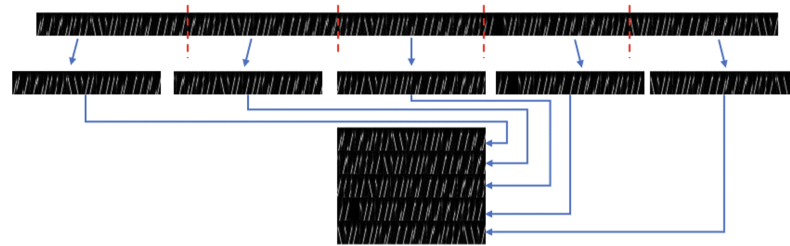

Figure 4. Frequency-to-image conversion.

This innovative approach aligns with the principles of the vision transformer network, thus allowing for a more balanced and interpretable representation of the frequency data. The sequential concatenation of frames into patches and their vertical integration into a unified matrix contribute to mitigating imbalances in the original sequence, thus preparing the data effectively for downstream processing.

By adopting this reverse approach, we address the challenge posed by the unevenness in the two-dimensional matrix, thereby ensuring that the subsequent image-based processing benefits from a more balanced and coherent input. The resulting images serve as a comprehensive and interpretable representation of the temporal evolution of frequency components, thus further enhancing the effectiveness of deep learning models in tasks such as human activity recognition.

3.3. Deep Learning Model

After completing the aforementioned signal processing, the creation of activity recognition datasets based on LoRa becomes feasible. To validate the proposed method's viability, several state-of-the-art (SOTA) deep learning network architectures—including CNN-LSTM, Swin Transformer, ConvNext, and Vision TF—can be employed. In our proposed system, the inputs to the neural networks are the spectrogram representations of the pre-processed RF signal data. These spectrograms, which are generated through the application of Short-Time Fourier Transform (STFT), capture the frequency content of the RF signals over short time intervals. Each spectrogram represents a specific time segment of the signal data and provides a visual representation in the form of an image. These spectrograms serve as the input data to the neural networks. The outputs of the neural networks are the predictions or classifications made by the models. Specifically, the neural networks are trained to classify the input spectrograms into different activity categories, such as walking, running, etc., according to the specific application.

The choice of these SOTA models is strategic, whereby the challenges and requirements of activity recognition when using LoRa signals via the leveraging of their respective strengths are addressed, such as spatial feature extraction, temporal modeling, long-range dependency capturing, and global context understanding. Their adoption in the evaluation

process allows for a comprehensive analysis of the proposed approach's effectiveness in accurately recognizing activities from LoRa signals.

3.3.1. CNN-LSTM

The CNN-LSTM model, as depicted in Figure 5, combines convolutional neural networks (CNNs) and long short-term memory networks (LSTMs), and it is a potent approach for image classification [38]. It inherits CNNs' ability to extract spatial features hierarchically, thus making it adept at recognizing patterns and objects. LSTMs complement this by capturing temporal dependencies, which are crucial when dealing with image sequences or video data. The model effectively fuses spatial and sequential information, thus adapting to varying sequence lengths and providing contextual reasoning. LSTM regularization helps mitigate overfitting, and attention mechanisms allow for fine-grained analysis. End-to-end training simplifies the process, thus making the CNN-LSTM model valuable for image classification, particularly in tasks involving dynamic content or temporal contexts. By leveraging both spatial and temporal information, CNN-LSTM effectively captures complex relationships within LoRa signals, thus making it suitable for activity recognition.
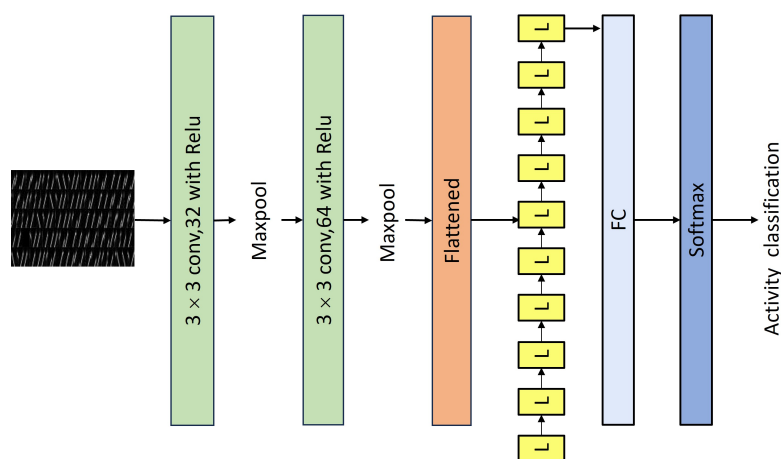

Figure 5. The CNN-LSTM model.

3.3.2. Swin Transformer

The Swin transformer is a deep learning architecture that combines the principles of transformers and convolutional neural networks (CNNs) to address various computer vision tasks, such as image classification, object detection, and segmentation [39]. Figure 6 illustrates the architecture of the Swin transformer model. The Swin transformer has achieved competitive results on various benchmarks and is considered a state-of-the-art architecture for a variety of computer vision tasks. Its hybrid design, which combines elements of transformers and CNNs, allows it to capture long-range dependencies in images effectively, thus making it a promising tool in the field of computer vision. It provides a robust framework for capturing long-range dependencies, which are crucial for understanding the temporal dynamics of activities in LoRa signals. Its modular structure with hierarchical processing enables effective feature extraction and representation learning, thus making it suitable for activity recognition tasks.

In Figure 6, the W-MSA (windowed multi-head self-attention) is a conventional self-attention mechanism utilized to capture global contextual information. Conversely, SW-MSA (sliding window multi-head self-attention) introduces local attention relationships into the self-attention mechanism by incorporating a sliding window. By utilizing a sliding window, the SW-MSA enables the model to effectively capture dependencies among nearby positions in the sequence without the requirement of computing global self-attention. Additionally, LN denotes layer normalization.

3.3.3. ConvNext

ConvNext is a purely convolutional neural network model that is constructed entirely from standard convolutional neural network modules [40]. Figure 7 illustrates the architecture of the Swin transformer model. It is designed to be simple, accurate, efficient, and scalable. Starting from ResNet-50 or ResNet-200, it borrows ideas from Swin transformer in five aspects: macro design, depthwise separable convolution (ResNext), inverse bottleneck layers (MobileNet), large convolutional kernels, and detailed design. While maintaining the structure of a convolutional neural network, ConvNext achieves similar or better results than the Swin transformer in classification, detection, and segmentation downstream tasks, which is achieved through parameter tuning techniques borrowed from methods like Swin transformer. Its deep convolutional layers can spatially learn the discriminative features present in LoRa signals. ConvNext models have demonstrated strong capabilities in extracting complex patterns, which are essential for accurately recognizing activities from LoRa signals.
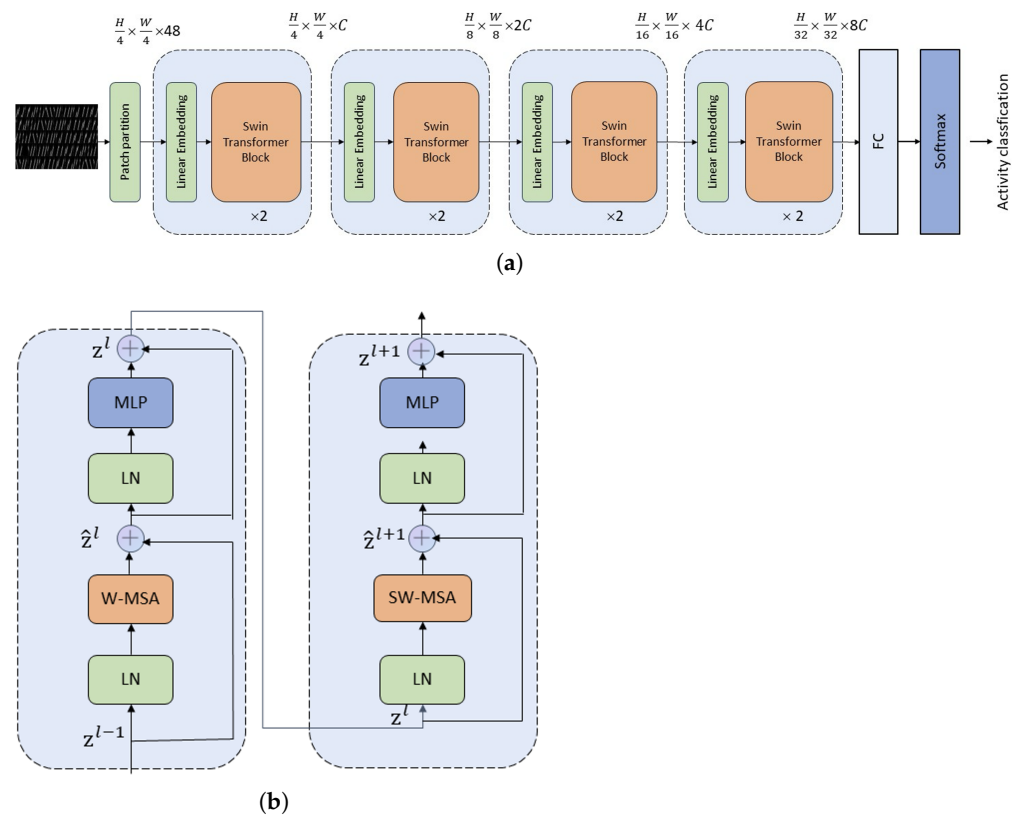
Figure 6. The Swin transformer model. (**a**) Architecture. (**b**) Two successive Swin transformer blocks.
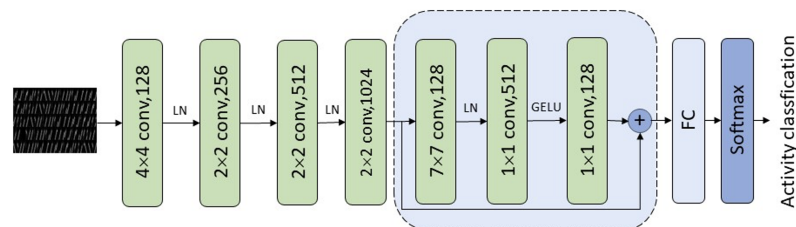
Figure 7. The ConvNext model.

3.3.4. Vision TF

Vision TF is a model based on the standard transformer architecture [41]. Figure 8 illustrates the architecture of the Vision TF model. It splits the entire image into small image patches, linearly embeds these patches into sequences, and feeds them as inputs into

the transformer network. It then uses supervised learning for image classification. Unlike traditional CNN algorithms, Vision TF attempts to directly apply the standard transformer structure to images with minimal modifications to the entire image classification process. To meet the input requirements of the transformer structure, it divides the entire image into small image patches and linearly embeds these patches into sequences for input into the network. It excels in capturing global dependencies across the entire signal, thereby enabling a more comprehensive understanding of the activity patterns in LoRa signals. Vision TF's attention mechanism attends to the signal's spatial and positional information, making it uniquely suitable for analyzing the complex characteristics of LoRa signals.
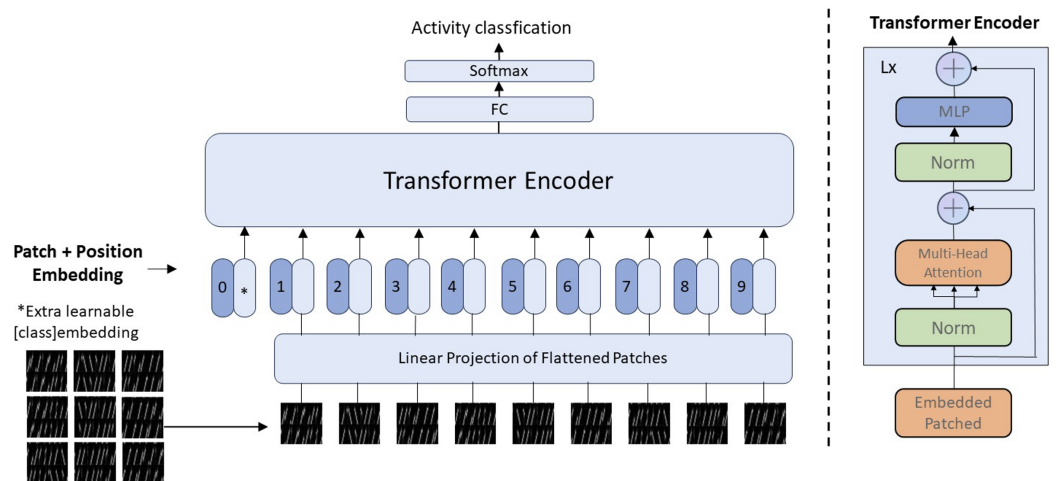

Figure 8. The Vision TF model.

3.4. Experimental Design

3.4.1. Experimental Setup

The LoRa experimental prototype system consists of a LoRa transmitter and a LoRa receiver, as depicted in Figure 9. The transmitter comprises an Arduino Uno development board equipped with a Semtech SX1276 LoRa node, thus enabling signal transmission. To enhance the transmission range and reliability, a directional antenna was employed for the effective propagation of the LoRa signals.
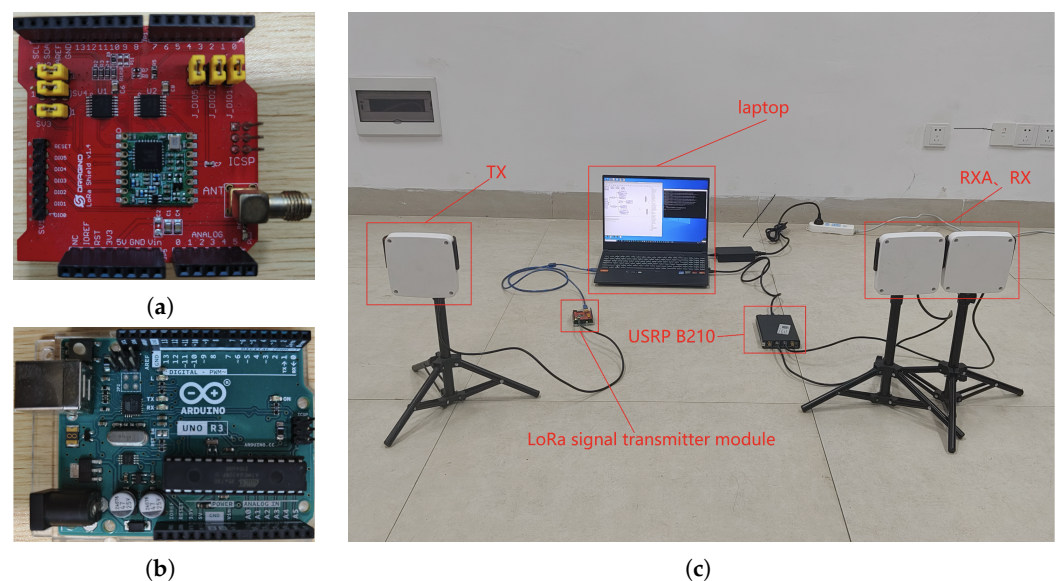

Figure 9. LoRa experimental prototype system. (**a**) Semtech SX1276 LoRa node. (**b**) Arduino Uno development board. (**c**) Deployment of the transmitting and receiving antennas.

On the receiver side, a USRP B210 was utilized in conjunction with a GNU Radio. The receiver setup incorporated two separate antennas, each receiving data independently. This configuration enabled improved reception performance and diversity in capturing the transmitted signals.

Specifically, the Semtech SX1276 LoRa node (Figure 9a) was connected to the Arduino Uno development board (Figure 9b), which served as the control unit for the transmitter. The deployment of transmitting and receiving antennas (Figure 9c) ensured optimal signal propagation and reception.

The LoRa transmitter operates within a 915 MHz frequency band and transmits signals with a channel bandwidth of 125 kHz. On the receiver end, within the GNU radio, the sampling rate was set to 900 kHz, and the gain type attribute was configured as 'Normalized'. The gain value was gradually adjusted to select the appropriate gain setting, thereby ensuring optimal signal reception and processing.

To facilitate data collection and processing, the receiver was connected to a laptop via USB. This connection enables the transfer of received LoRa signals from the receiver to the laptop for further analysis, interpretation, and processing.

Overall, the LoRa experimental prototype system involves the integration of the Semtech SX1276 LoRa node, Arduino Uno development board, USRP B210, GNU radio, and other appropriate antennas. These components work together to create a comprehensive experimental setup for LoRa signal transmission and reception, thereby supporting data collection and subsequent analyses.

3.4.2. Experimental Scenario Design and Datasets

To validate the feasibility of the designed LoRa activity recognition system, two experimental scenarios were designed.

A. The Design of a Fine-Grained Activity Scenario in a Single Room

In a spacious 10 m × 12 m room (see Figure 10), the gain value was adjusted to 0.3. Eight participants performed six different activities, and data were collected for each activity. Two data collectors monitored the received data, thus ensuring accurate annotations. The collected data were processed into two datasets: one for human activity recognition (Table 2) and another for identity recognition (Table 3).
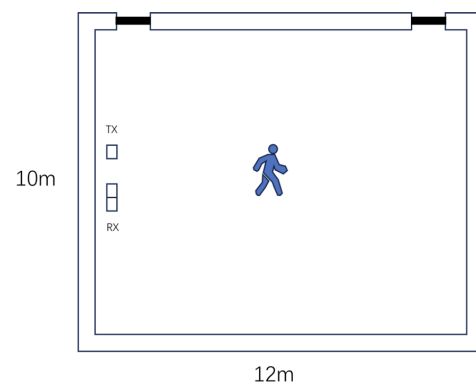

Figure 10. Fine-grained activity scenario in a single room.

We processed the collected data into two distinct datasets within the framework of our study. The first dataset was created by labeling the samples with activity names, which resulted in a human activity recognition dataset with six activity categories, as shown in Table 2. The second dataset was formed by labeling the samples with the names of the volunteers, and this was achieved by treating the activities performed by each individual (including all six types) as samples of the same class. This resulted in an identity recognition dataset consisting of eight individuals, as presented in Table 3.

Table 2. Activity dataset in the fine-grained activity scenario.

Activity Categories	Number of Samples	Percentage
Jogging	1600	14.9%
Walking	1600	14.9%
Picking up	1604	14.9%
Squatting	1584	14.8%
Standing	2756	25.7%
Standing up	1584	14.8%
Sum	10,728	100.0%

The dataset in Table 2 was constructed by assigning activity names as sample labels, which also represent the six distinct human activities. This dataset offers valuable insights into accurately recognizing and classifying these specific activities.

**Table 3.** Human identity dataset in the fine-grained activity scenario.

Name Abbreviation	Number of Samples	Percentage
Y.B.	1204	11.2%
C.H.	1772	16.5%
D.M.	1200	11.2%
F.R.	1196	11.1%
L.H.	1200	11.2%
R.H.	1200	11.2%
W.J.	1164	10.9%
X.Y.	1792	16.7%
Sum	10,728	100.0%

In Table 3, samples were labeled based on the names of the volunteers, and this was achieved by considering each individual's activities as samples of the same class. This dataset allowed us to explore the recognition of different individuals based on their distinct activity patterns.

The two datasets were derived from the same experimental environment to evaluate the effectiveness of the LoRa activity recognition system. The availability of these datasets allowed for the development and evaluation of activity recognition and identity recognition algorithms, thereby contributing to advancements in context-aware applications and personalized services.

B. The Design of the Coarse-Grained Activity Scenario in Multiple Rooms

We selected four contiguous rooms, as depicted in Figure 11, as the experimental setting. The gain value was adjusted to 0.83, and five participants were invited to freely move within these four rooms without any restrictions on speed, location, or direction.
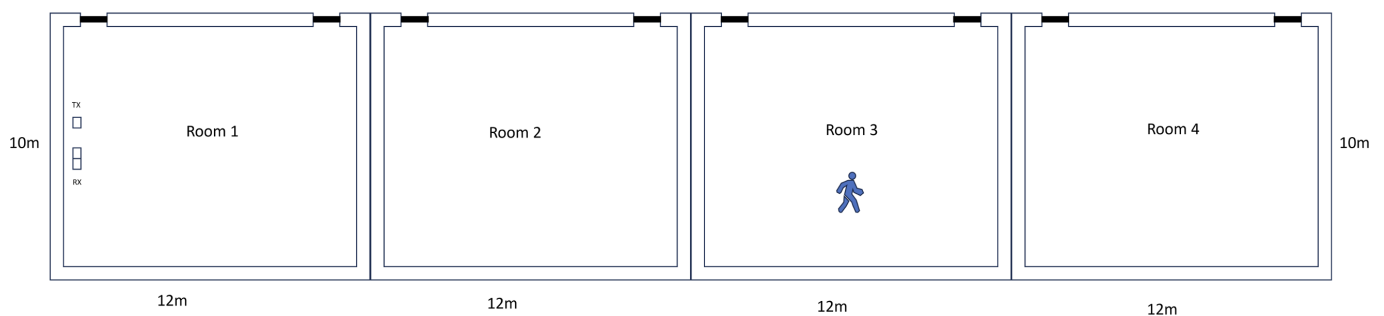
Figure 11. Coarse-grained activity scenario in multiple contiguous rooms.

Participants voluntarily engaged in activities in four different rooms. Data samples were collected every 5 seconds and labeled with the corresponding room number. By

meticulously recording the activities of each participant in each room and assigning them the corresponding room numbers as labels, we were able to create a comprehensive dataset that accurately reflected the coarse-grained nature of activities occurring in different room environments, as shown in Table 4. This dataset aimed to capture activities across multiple rooms, as well as to facilitate the recognition and classification of activities based on room locations.

Table 4. Dataset for the coarse-grained activity scenario in multiple rooms.

Room No.	Number of Samples	Percentage
1	1280	24.6%
2	1280	24.6%
3	1280	24.6%
4	1360	26.2%
Sum	5200	100.0%

In addition, we also collected a dataset of 3240 samples for non-humans in the rooms, which specifically captures instances where there is no activity or when the room is in an empty state. When combined with the previously collected dataset of 5200 samples for human activities in four rooms, we created a comprehensive dataset for presence detection in multiple rooms. This dataset was designed to evaluate the system's capability to accurately detect and differentiate between "Human in Rooms" activities and "Empty room state" instances in various rooms. The dataset details are provided in Table 5.

Table 5. Dataset for presence detection in multiple rooms.

State	Number of Samples	Percentage
Human in Rooms	5200	61.6%
No-human in Rooms	3240	38.4%
Sum	8440	100.0%

3.4.3. Experimental Configuration

The experimental configuration in this paper aimed to ensure the adoption of common and consistent settings to evaluate the compatibility of our signal preprocessing methods and data collection processes with general-purpose models. The intention was to make the models as simple to use as possible. For the deep learning models utilized in this study, namely CNN-LSTM, Swin transformer, Vision TF, and ConvNext, the selection was based on specific considerations. CNN-LSTM is known for its ability to capture spatial and temporal dependencies, Swin transformer is recognized for its attention mechanism, Vision TF excels in vision tasks, and ConvNext has shown superior performance in previous activity recognition studies. This diverse set of models was chosen to assess their individual strengths and weaknesses in the context of our research objectives. To ensure a fair and comprehensive comparison, the training process involved 400 epochs for each model, maintaining a fixed learning rate of 0.001 and a batch size of 8. These parameter settings were selected to strike a balance between model performance and computational efficiency. By using these standardized parameter settings, we aimed to assess the effectiveness of our signal preprocessing techniques and the generalizability of the models across different datasets and tasks.

With respect to hardware, an Intel Core i9-12900k CPU, an NVIDIA GeForce RTX 4090 GPU, and 64 GB of RAM were utilized. This setup provided ample computational power to handle the training and evaluation of the models. The experiment was conducted in a Python environment with Python version 3.10.9. The models were developed and executed using the PyTorch library, specifically version 2.0.1. Furthermore, the CUDA toolkit version 11.8 was employed to leverage GPU acceleration, thus enabling faster training and inferences.

4. Experimental Evaluation

4.1. Experimental Description

In this section, we delineate the experimental designs that were aimed at evaluating the efficacy of our proposed approach in human activity recognition. Our focus encompasses five key facets: activity classification, identity recognition, room identification, presence detection, as well as a comparative study between differential signal processing (DSP) and non-differential signal processing (NDSP) in activity classification.

Activity Classification (AC): Our objective here was to classify the diverse activities using deep learning models. We utilized a dataset featuring a spectrum of activities performed by eight participants, including jogging, walking, picking up, squatting, standing, and standing up (refer to Table 2 for details).

Identity Recognition (IR): The aim here was to recognize individuals based on their activity patterns. Through leveraging a labeled dataset containing volunteer names, we trained neural network models on separate training and testing sets, as outlined in Table 3.

Room Identification (RI): This aspect involved determining the specific room where an activity occurs. We segregated our dataset into training and testing sets based on room information and train models to identify room-specific patterns, as detailed in Table 4.

Presence Detection (PD): The focus here was on distinguishing between human and non-human activities. The models were trained on activities that were performed in designated rooms, along with a separate dataset representing no activity (see Table 5). The evaluation was conducted on a dedicated testing set.

Differential Signal Processing (DSP) vs. Non-Differential Signal Processing (NDSP): Here, we delved into a thorough assessment of the effectiveness of differential signal processing (DSP) and non-differential signal processing (NDSP) techniques in the domain of activity classification. We conducted a comprehensive evaluation to measure and compare the accuracy achieved by each approach. This analysis allowed us to gain valuable insights into the performance and suitability of the DSP and NDSP methods for accurately classifying various human activities.

4.2. Evaluation Metrics

In selecting our evaluation metrics, we deliberately chose accuracy (ACC) and the confusion matrix due to their relevance and contributions toward a comprehensive assessment of our approach in human activity recognition.

Accuracy (ACC): We utilized ACC as it offers a straightforward measure of the overall correctness in activity recognition. By comparing the predicted labels with ground truth labels, ACC provides a high-level understanding of the model's effectiveness in capturing various activities within the dataset.

Confusion Matrix: This matrix, which presents a detailed breakdown of correct and misclassified instances for each activity class, offers a nuanced view of model performance. It goes beyond a singular accuracy score, thereby allowing us to identify specific areas of strength and potential improvement. The confusion matrix is particularly valuable for understanding how well the model distinguishes between different activities.

Together, ACC and the confusion matrix form a robust set of metrics. ACC provides a global assessment, while the confusion matrix offers a fine-grained analysis, thus allowing us to interpret the model's behavior across diverse activity classes. This combination ensures a thorough and insightful evaluation, thereby aligning with the primary focus of our work on signal processing in the context of human activity recognition.

4.3. Experimental Results and Analysis

In this section, we present a comprehensive analysis and evaluation of the experimental results obtained from various deep learning models in different experimental designs, including activity classification, identity recognition, room identification, and presence detection, as well as a comparison between differential signal processing (DSP) and non-differential signal processing (NDSP) in activity classification.

The results from Table 6 highlight the performance of different models in various recognition tasks, thereby providing important insights into our findings. Let us delve into the analysis.

Table 6. The ACC of different models in various recognition tasks in DSP and NDSP scenarios.

Model	AC (DSP)	AC (NDSP)	IR (DSP)	IR (NDSP)	RI (DSP)	RI (NDSP)	PD (DSP)	PD (NDSP)
CNN-LSTM	0.802	0.257	0.835	0.167	0.774	0.262	0.834	0.627
Swin transformer	0.913	0.883	0.165	0.165	0.838	0.262	0.652	0.627
Vision TF	0.946	0.883	0.92	0.605	0.836	0.740	**0.975**	0.961
ConvNext	**0.967**	0.918	**0.938**	0.167	**0.979**	0.771	0.922	0.876

Activity Classification (DSP): The ConvNext model achieved the highest accuracy of 0.967, which was closely followed by Vision TF at 0.946. While Swin transformer and CNN-LSTM showed slightly a lower accuracy (0.913 and 0.802, respectively), the overall results underlined the effectiveness of deep learning models in accurately classifying diverse human activities.

Identity Recognition (DSP): Once again, ConvNext demonstrated superiority with an accuracy of 0.938, followed by Vision TF at 0.920. It is worth noting that Swin transformer lagged behind with an accuracy of 0.165, while CNN-LSTM performed moderately well at 0.835. These results emphasized ConvNext's capability in precisely identifying individuals based on their activity patterns.

Room Identification (DSP): The ConvNext model exceled in room identification with an accuracy of 0.979. The other models of CNN-LSTM, Swin transformer, and Vision TF exhibited commendable accuracies of 0.774, 0.838, and 0.836, respectively. These outcomes underscored ConvNext's proficiency in accurately identifying different rooms.

Presence Detection (DSP): The Vision TF model displayed an outstanding accuracy at 0.975 in accurately detecting the presence of individuals. ConvNext follows closely with an accuracy of 0.922, while Swin transformer and CNN-LSTM showed moderate performances at 0.652 and 0.834, respectively. These results highlighted the robust capabilities of Vision TF and ConvNext in detecting the presence of human activities.

The comparison between the models trained with DSP and NDSP approaches revealed significant differences in the accuracy across all evaluated tasks (AC, IR, RI, and PD). For example, ConvNext achieved an accuracy of 0.967 in the DSP scenario in activity classification, while the accuracy dropped to 0.918 in the NDSP scenario. Similarly, Vision TF showed a higher accuracy of 0.946 in the DSP scenario compared to 0.883 in the NDSP scenario. These results confirmed the superiority of incorporating differential signal processing techniques in enhancing the accuracy of activity classification, identity recognition, room identification, and presence detection.

As demonstrated in Figure 12, the comparison between the ACC values for human activity recognition in the DSP and NDSP scenarios further reinforced the superiority of DSP over NDSP. The results unequivocally revealed that employing DSP technology yielded a considerably higher accuracy in the dataset validation process compared to NDSP.

In Figure 13, we present the confusion matrices detailing the performance of the ConvNext network in activity classification, identity recognition, and room identification, alongside the Vision TF network in presence detection in the DSP scenario. These networks emerged as standout performers as they yielded the most promising results in their respective tasks.

Activity Classification: As shown in Figure 13a, the ConvNext network demonstrated exceptional performance by achieving an accuracy of 96.7%. The associated confusion matrix revealed the network's precision in classifying various activities, with minimal instances of misclassification. Notably, jogging and walking activities were identified with high precision, and this was evident from the dominant diagonal entries in the matrix.
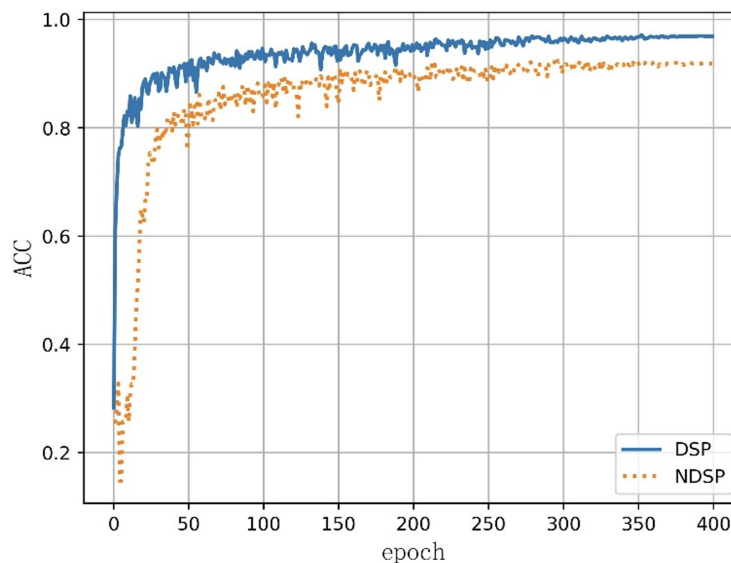
Figure 12. Comparison of the ACC results in AC for the NDSP and DSP scenarios.

Identity Recognition: As shown in Figure 13b, the ConvNext network continued to impress with an accuracy of 97.9% in the identity recognition task. The confusion matrix highlighted the network's proficiency in correctly identifying individuals, thereby showcasing minimal confusion between different identities. Furthermore, the network's high accuracy was noteworthy in recognizing the identity labeled "Y.B.", as evidenced by a strong diagonal entry in the corresponding row.

Room Identification: As show in in Figure 13c, the ConvNext network showcased its effectiveness with an accuracy of 97.3% in room identification. The associated confusion matrix illustrated the network's capability to accurately identify the presence of individuals in various rooms. While Room 1 experienced a slightly higher misclassification rate of 2.7%, the network demonstrated outstanding performance in differentiating individuals in Rooms 2, 3, and 4, in which accuracy rates exceeding 98% were achieved.

Presence Detection: As shown in Figure 13d, the Vision TF network delivered an exceptional accuracy of 98.5% in the presence detection task. The corresponding confusion matrix emphasized the network's ability to accurately detect the presence of individuals, exhibiting minimal false positives or false negatives. The Vision TF network consistently showcased remarkable performance across all tested scenarios, underscoring its efficacy in presence detection tasks.

These results, visually represented through confusion matrices, provide a comprehensive understanding of the networks' capabilities in each recognition task, thereby further validating their excellence in diverse aspects of human activity recognition.

In conclusion, the experimental findings strongly support the superior performance of the ConvNext model in human activity recognition when trained and validated using the DSP approach. The incorporation of differential signal processing techniques enhances the model's ability to accurately classify and recognize various human activities by improving the signal-to-noise ratio, amplifying activity-related signal patterns, and providing valuable insights into the characteristics of LoRa wireless RF signals. The comparison between models trained with DSP and NDSP further confirmed the effectiveness of DSP in enhancing accuracy. The results from the confusion matrices highlighted the precise classification and recognition capabilities of the ConvNext and Vision TF networks in different recognition tasks.
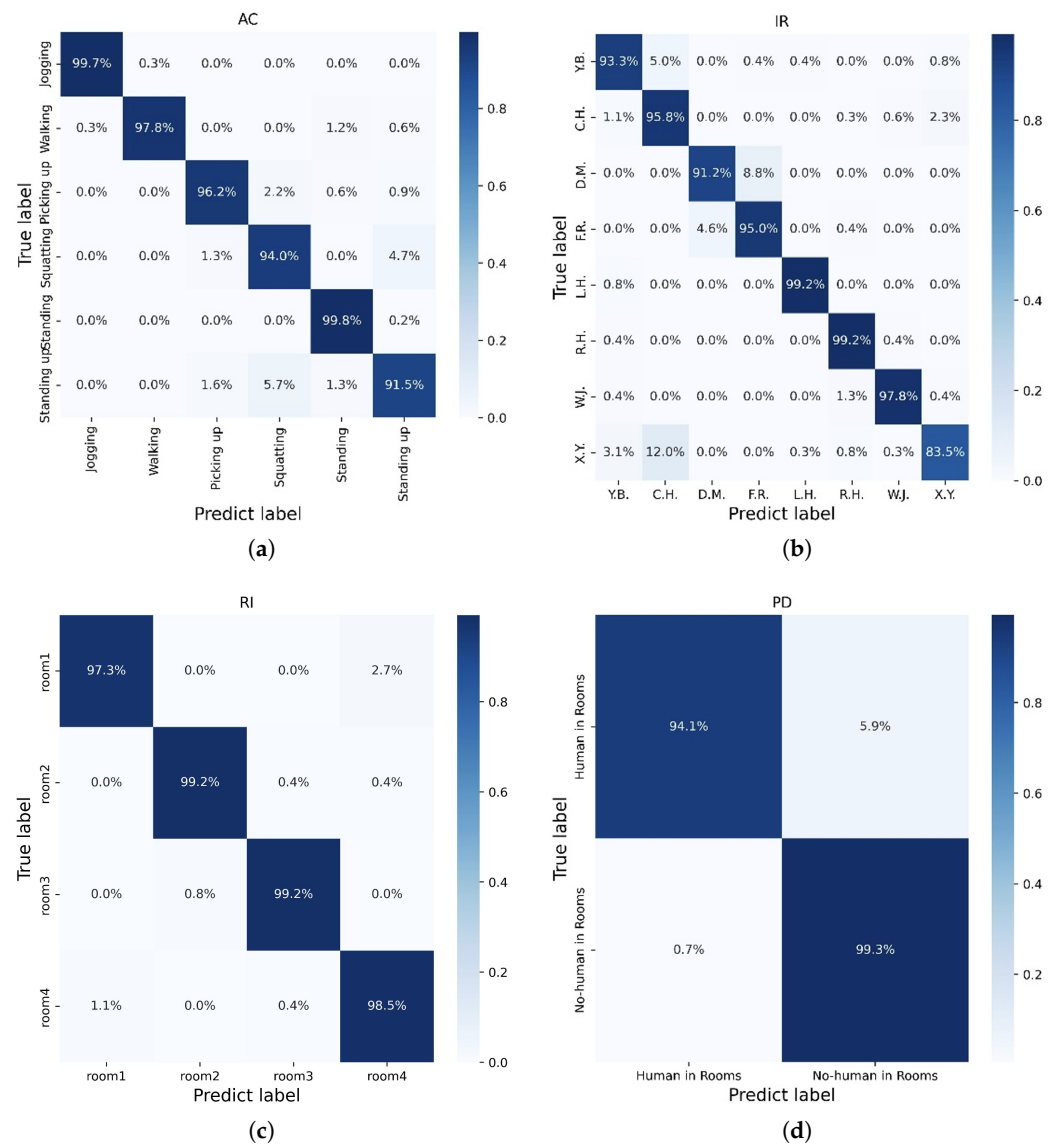
Figure 13. The confusion matrices of AC, IR, RI, and PD. (**a**) The confusion matrix of AC. (**b**) The confusion matrix of IR. (**c**) The confusion matrix of RI. (**d**) The confusion matrix of PD.

5. Conclusions and Future Work

In this paper, we presented a comprehensive approach through which to improve human activity recognition when using LoRa wireless RF signal preprocessing and deep learning techniques. Our focus was on addressing the challenge of extracting relevant features from complex LoRa signals through various signal processing methods.

To enhance the representation and analysis of signals, we employed preprocessing techniques, such as converting complex LoRa RF signal data into real numbers and utilizing the Short-Time Fourier Transform (STFT) for spectral analysis. Additionally, we leveraged differential signal processing (DSP) techniques to effectively capture unique signal patterns associated with human activities. The adoption of frequency-to-image conversion techniques allowed for visualizing the frequency data, thereby significantly improving the accuracy of activity recognition.

The experiments conducted with our approach demonstrated its effectiveness in various aspects of activity recognition. Deep learning models that were trained on our provided dataset successfully classified different activity types, whereby they achieved the accurate recognition of individuals based on their activity patterns and developed models

capable of accurately identifying the specific room in which an activity occurs, as well as were able to differentiate between human and non-human activities.

Looking ahead, our future work will focus on exploring advancements in our approach. This includes investigating different deep learning architectures, feature engineering techniques, and signal processing algorithms to further enhance the accuracy and robustness of our system. We also plan to collect more diverse datasets to evaluate the generalizability of our approach. Additionally, we will explore the deployment of our system in real-world scenarios to assess its performance and feasibility in practical applications.

Author Contributions: Conceptualization, L.Z. and M.N.; methodology, H.C.; software, H.C.; validation, L.Z. and M.N.; formal analysis, L.Z. and M.N.; investigation, H.C. and X.Z.; resources, M.N. and Y.W.; data curation, H.C. and X.Z.; writing—original draft preparation, M.N.; writing—review and editing M.N. and Y.W.; visualization, H.C.; supervision, M.N. and Y.W.; funding acquisition, M.N. and Y.W. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The article contains the data that are also available from the corresponding authors upon reasonable request.

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

References

1. Arshad, M.H.; Bilal, M.; Gani, A. Human activity recognition: Review, taxonomy and open challenges. *Sensors* **2022**, *22*, 6463. [CrossRef] [PubMed]
2. Yadav, S.K.; Tiwari, K.; Pandey, H.M.; Akbar, S.A. A review of multimodal human activity recognition with special emphasis on classification, applications, challenges and future directions. *Knowl.-Based Syst.* **2021**, *223*, 106970. [CrossRef]
3. Wang, S.; Zhou, G. A review on radio based activity recognition. *Digit. Commun. Netw.* **2015**, *1*, 20–29. [CrossRef]
4. Ige, A.O.; Noor, M.H.M. A survey on unsupervised learning for wearable sensor-based activity recognition. *Appl. Soft Comput.* **2022**, *127*, 109363. [CrossRef]
5. Beddiar, D.R.; Nini, B.; Sabokrou, M.; Hadid, A. Vision-based human activity recognition: A survey. *Multimed. Tools Appl.* **2020**, *79*, 30509–30555. [CrossRef]
6. Khunteta, S.; Saikrishna, P.; Agrawal, A.; Kumar, A.; Chavva, A.K.R. RF-Sensing: A New Way to Observe Surroundings. *IEEE Access* **2022**, *10*, 129653–129665. [CrossRef]
7. Li, W.; Vishwakarma, S.; Tang, C.; Woodbridge, K.; Piechocki, R.J.; Chetty, K. Using RF transmissions from IoT devices for occupancy detection and activity Recognition. *IEEE Sens. J.* **2021**, *22*, 2484–2495. [CrossRef]
8. Yan, H.; Zhang, Y.; Wang, Y.; Xu, K. WiAct: A passive WiFi-based human activity recognition system. *IEEE Sens. J.* **2019**, *20*, 296–305. [CrossRef]
9. Pramudita, A.A.; Suratman, F.Y. Low-power radar system for noncontact human respiration sensor. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–15. [CrossRef]
10. Xie, B.; Xiong, J.; Chen, X.; Fang, D. Exploring commodity rfid for contactless sub-millimeter vibration sensing. In Proceedings of the 18th Conference on Embedded Networked Sensor Systems, Virtual Event, Japan, 16–19 November 2020; pp. 15–27.
11. Iannizzotto, G.; Milici, M.; Nucita, A.; Lo Bello, L. A perspective on passive human sensing with bluetooth. *Sensors* **2022**, *22*, 3523. [CrossRef]
12. Li, T.; An, C.; Tian, Z.; Campbell, A.T.; Zhou, X. Human sensing using visible light communication. In Proceedings of the 21st Annual International Conference on Mobile Computing and Networking, Paris, France, 7–11 September 2015; pp. 331–344.
13. Tan, D.K.P.; He, J.; Li, Y.; Bayesteh, A.; Chen, Y.; Zhu, P.; Tong, W. Integrated sensing and communication in 6G: Motivations, use cases, requirements, challenges and future directions. In Proceedings of the 2021 1st IEEE International Online Symposium on Joint Communications & Sensing (JC&S), Dresden, Germany, 23–24 February 2021; pp. 1–6.
14. Zhang, F.; Chang, Z.; Niu, K.; Xiong, J.; Jin, B.; Lv, Q.; Zhang, D. Exploring lora for long-range through-wall sensing. In *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*; ACM Digital Library: New York, NY, USA, 2020; Volume 4, pp. 1–27.

15. Li, B.; Cui, W.; Wang, W.; Zhang, L.; Chen, Z.; Wu, M. Two-stream convolution augmented transformer for human activity recognition. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtually, 2–9 February 2021; Volume 35, pp. 286–293.

16. Lin, G.; Jiang, W.; Xu, S.; Zhou, X.; Guo, X.; Zhu, Y.; He, X. Human activity recognition using smartphones with WiFi signals. *IEEE Trans.-Hum.-Mach. Syst.* **2022**, *53*, 142–153. [CrossRef]

17. Mekruksavanich, S.; Phaphan, W.; Hnoohom, N.; Jitpattanakul, A. Attention-based hybrid deep learning network for human activity recognition using WiFi channel state information. *Appl. Sci.* **2023**, *13*, 8884. [CrossRef]

18. Moghaddam, M.G.; Shirehjini, A.A.N.; Shirmohammadi, S. A WiFi-based system for recognizing fine-grained multiple-subject human activities. In Proceedings of the 2022 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Ottawa, ON, Canada, 16–19 May 2022; pp. 1–6.

19. Saw, C.Y.; Wong, Y.C. Neuromorphic computing with hybrid CNN–Stochastic Reservoir for time series WiFi based human activity recognition. *Comput. Electr. Eng.* **2023**, *111*, 108917. [CrossRef]

20. Shrestha, A.; Li, H.; Le Kernec, J.; Fioranelli, F. Continuous human activity classification from FMCW radar with Bi-LSTM networks. *IEEE Sens. J.* **2020**, *20*, 13607–13619. [CrossRef]

21. Chen, H.; Ding, C.; Zhang, L.; Hong, H.; Zhu, X. Human Activity Recognition using Temporal 3DCNN based on FMCW Radar. In Proceedings of the 2022 IEEE MTT-S International Microwave Biomedical Conference (IMBioC), Suzhou, China, 16–18 May 2022; pp. 245–247.

22. Rafli, R.; Suratman, F.Y.; Istiqomah. FMCW Radar Signal Processing for Human Activity Recognition with Convolutional Neural Network. In Proceedings of the 3rd International Conference on Electronics, Biomedical Engineering, and Health Informatics: ICEBEHI 2022, Surabaya, Indonesia, 5–6 October 2023; pp. 429–445.

23. Huang, A.; Wang, D.; Zhao, R.; Zhang, Q. Au-id: Automatic user identification and authentication through the motions captured from sequential human activities using rfid. In *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*; ACM Digital Library: New York, NY, USA, 2019; Volume 3, pp. 1–26.

24. Liu, Y.; Huang, W.; Jiang, S.; Zhao, B.; Wang, S.; Wang, S.; Zhang, Y. TransTM: A device-free method based on time-streaming multiscale transformer for human activity recognition. *Def. Technol.* 2023, *in press* .

25. Sun, Z.; Yang, H.; Liu, K.; Yin, Z.; Li, Z.; Xu, W. Recent advances in LoRa: A comprehensive survey. *ACM Trans. Sens. Netw.* **2022**, *18*, 1–44. [CrossRef]

26. Zhang, F.; Chang, Z.; Xiong, J.; Zhang, D. Exploring LoRa for Sensing. *Getmobile Mob. Comput. Commun.* **2021**, *25*, 33–37. [CrossRef]

27. Lin, S.; Ying, Z.; Zheng, K. Design and implementation of location and activity monitoring system based on LoRa. *arXiv* **2019**, arXiv:1902.01947.

28. Shi, L.; Xu, H.; Ji, W.; Zhang, B.; Sun, X.; Li, J. Real-time human activity recognition system based on capsule and LoRa. *IEEE Sens. J.* **2020**, *21*, 667–677. [CrossRef]

29. Dos Reis, B.R.; Easton, Z.; White, R.R.; Fuka, D. A LoRa sensor network for monitoring pastured livestock location and activity. *Transl. Anim. Sci.* **2021**, *5*, txab010. [CrossRef]

30. O'kennedy, M.; Niesler, T.; Wolhuter, R.; Mitton, N. Practical evaluation of carrier sensing for a LoRa wildlife monitoring network. In Proceedings of the 2020 IFIP Networking Conference (Networking), Paris, France, 2–26 June 2020; pp. 614–618.

31. Huang, Q.; Luo, Z.; Zhang, J.; Wang, W.; Zhang, Q. LoRadar: Enabling concurrent radar sensing and LoRa communication. *IEEE Trans. Mob. Comput.* **2020**, *21*, 2045–2057. [CrossRef]

32. Binbin, X.; Yuqing, Y.; Jie, X. Pushing the Limits of Long Range Wireless Sensing with LoRa. In *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*; ACM Digital Library: New York, NY, USA, 2021; Volume 5, pp. 1–21.

33. Zhang, F.; Chang, Z.; Xiong, J.; Zheng, R.; Ma, J.; Niu, K.; Jin, B.; Zhang, D. Unlocking the beamforming potential of LoRa for long-range multi-target respiration sensing. In *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*; ACM Digital Library: New York, NY, USA, 2021; Volume 5, pp. 1–25.

34. Xie, B.; Ganesan, D.; Xiong, J. Embracing lora sensing with device mobility. In Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems, Boston, MA, USA, 6–9 November 2022; pp. 349–361.

35. Lin, Y.; Yang, F. IQ-Data-Based WiFi Signal Classification Algorithm Using the Choi-Williams and Margenau-Hill-Spectrogram Features: A Case in Human Activity Recognition. *Electronics* **2021**, *10*, 2368. [CrossRef]

36. Ding, C.; Zhang, L.; Chen, H.; Hong, H.; Zhu, X.; Fioranelli, F. Sparsity-based Human Activity Recognition with PointNet using a Portable FMCW Radar. *IEEE Internet Things J.* **2023**, *10*, 10024–10037. [CrossRef]

37. Baldini, G.; Gentile, C.; Giuliani, R.; Steri, G. Comparison of techniques for radiometric identification based on deep convolutional neural networks. *Electron. Lett.* **2019**, *55*, 90–92. [CrossRef]

38. Yao, H.; Tang, X.; Wei, H.; Zheng, G.; Yu, Y.; Li, Z. Modeling spatial-temporal dynamics for traffic prediction. *arXiv* **2018**, arXiv:1803.01254.

39. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 10012–10022.

40.   Liu, Z.; Mao, H.; Wu, C.Y.; Feichtenhofer, C.; Darrell, T.; Xie, S. A convnet for the 2020s. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 11976–11986.

41.   Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16×16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.