Article

Human Activity Recognition Method Based on Edge Computing-Assisted and GRU Deep Learning Network

Xiaocheng Huang [1,2,*]**, Youwei Yuan** [1,2]**, Chaoqi Chang** [1]**, Yiming Gao** [1]**, Chao Zheng** [1] **and Lamei Yan** [3]

1   School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 310018, China
2   Key Laboratory of Brain Machine Collaborative Intelligence of Zhejiang Province, Hangzhou 310018, China
3   School of Media and Design, Hangzhou Dianzi University, Hangzhou 310018, China
*   Correspondence: 221050090@hdu.edu.cn

Abstract: Human Activity Recognition (HAR) has been proven to be effective in various healthcare and telemonitoring applications. Current HAR methods, especially deep learning, are extensively employed owing to their exceptional recognition capabilities. However, in pursuit of enhancing feature expression abilities, deep learning often introduces a trade-off by increasing Time complexity. Moreover, the intricate nature of human activity data poses a challenge as it can lead to a notable decrease in recognition accuracy when affected by additional noise. These aspects will significantly impair recognition performance. To advance this field further, we present a HAR method based on an edge-computing-assisted and GRU deep-learning network. We initially proposed a model for edge computing to optimize the energy consumption and processing time of wearable devices. This model transmits HAR data to edge-computable nodes, deploys analytical models on edge servers for remote training, and returns results to wearable devices for processing. Then, we introduced an initial convolution method to preprocess large amounts of training data more effectively. To this end, an attention mechanism was integrated into the network structure to enhance the analysis of confusing data and improve the accuracy of action classification. Our results demonstrated that the proposed approach achieved an average accuracy of 85.4% on the 200 difficult-to-identify HAR data, which outperforms the Recurrent Neural Network (RNN) method's accuracy of 77.1%. The experimental results showcase the efficacy of the proposed method and offer valuable insights for the future application of HAR.

Keywords: human activity recognition; edge computing; gate recurrent unit; wearable sensor

1. Introduction

Recent years, wearable sensors, such as smart watches and smartphones, have become increasingly important in daily life. These sensors are capable of collecting data on physical activity, health, and other parameters [1,2]. They can accurately identify and classify various human activities, such as jumping, running, and cycling, making them valuable tools for developing personalized healthcare and fitness programs [3,4].

Human activity recognition aims to capture activity data using sensors and utilizes machine learning algorithms to extract meaningful behavior patterns. Advances in sensor technology and machine learning have propelled the progress of human activity recognition. Deep learning methods have gained increased attention in HAR [5,6]. They possess the ability to automatically learn and extract features from data, resulting in highly accurate behavior classification when trained on extensive datasets. Nonetheless, the computational complexity of activity recognition algorithms poses challenges for wearable devices, impacting their endurance and user experience. Collaborating with other computing platforms can enhance the performance of wearable devices and reduce their computational load.

By leveraging edge computing platforms, personalized human activity recognition models can be developed to meet the individual needs of users for small wearable devices,

such as bracelets. This approach can greatly enhance the user experience and prevent long-term damage to the devices caused by the high computational demands of traditional activity recognition methods. Additionally, the edge computing platform can provide real-time analysis and response, meeting the application delay limit requirements on wearable devices [7]. Overall, the use of edge computing in human activity recognition for wearable devices is a promising approach that can significantly improve the performance and usability of wearable devices in various applications.

In addition, deep learning models for human activity recognition still face persistent challenges in time-awareness and training speed. The traditional recurrent neural networks (RNN) and convolutional neural networks (CNN) structures cannot effectively handle time-series data, leading to limitations in processing time-related human activity recognition tasks [8]. Moreover, the traditional CNN and RNN models require a large amount of computation when dealing with large-scale data, resulting in slow training speed. To address these issues, researchers have taken measures such as employing the convolutional recurrent neural network (CRNN) structure, which combines CNN and RNN to better handle time-series data and achieve higher accuracy. In terms of training speed, optimization methods like pretraining and mini-batch training have been adopted, significantly improving the training speed. Although these deep model optimization methods made some progress, there is still room for improvement when applied in HAR.

In this paper, to address the challenges posed by slow training speed and low recognition accuracy in the presence of additional noise, we proposed an innovative approach that uses a new O-Inception Element-Wise-Attention GRU network (OI-EleAttG-GRU) model and integrates it into HAR analysis. In this model, the initial convolution method is used to better preprocess large amounts of training data for analysis [9]. The EleAttG-GRU component, as an improved version of the recurrent neural network, improves its ability to process temporal features by adding an attention mechanism to the network structure [10]. The contributions of this paper are as follows:

- The use of edge computing to human activity recognition provides sufficient computing resources for this task.
- A new O-Inception convolutional network structure is proposed to reduce the time complexity of the original network structure and, to some extent, reduce feature loss.
- A new EleAttG-GRU block to achieve more accurate human activity prediction by adjusting the attention mechanism.
- The use of O-Inception and EleAttG-GRU to human activity recognition, demonstrating the effectiveness of OI-EleAttG-GRU in experiments.

The remainder of the paper is organized as follows. Section 2—Related Works provides a review of human activity recognition studies. Section 3—Methods describes a scheduling model for human activity recognition tasks in edge computing environments and presents the OI-EleAttG-GRU structure that combines O-Inception and EleAttG. Section 4—HAR Datasets and Experiment Settings presents the datasets used in the experiments, including the WISDM dataset, UCI dataset, and a self-designed PHAD dataset, and provides implementation details. In Section 5—Results and Analysis, the experimental results are analyzed, demonstrating that our proposed architecture outperforms other advanced methods for human activity recognition. Finally, Section 6—Conclusions and Future Work provides concluding observations.

2. Related Works

HAR has been researched in many aspects, Despite the different approaches, be it vision-based, sensor-based, or Wi-Fi-based, HAR still faces substantial computational complexity issues. Edge computing, as a computing model that has gradually emerged in recent years, can transmit data to edge nodes for computation, reducing the computational burden on local devices. An early model of edge computing can be considered as Content Delivery Network (CDN) [11], which uses caching grid technology to deploy caching servers at different locations and distribute user access requests to the server closest to the

user's location through the scheduling and load-balancing capabilities of the central cloud platform. This technology reduces network congestion and improves cache hit rate and response speed.

Caceres proposed Cloudlet [12], which was deployed on a host at the edge of the network with more available computing resources. The uplink of Cloudlet is connected to the Internet, and the downlink is connected to mobile devices. Cloudlet provides edge services to reduce bandwidth and latency. Many scholars have conducted research on task offloading problems, which primarily concentrate on three aspects: (i) optimizing the energy consumption of mobile devices [13]; (ii) optimizing the communication and computation overhead between devices and clouds [14]; and (iii) optimizing the task completion time [15,16]. The optimization of these aspects can effectively improve the performance of edge computing in human activity recognition tasks, making it more efficient and accurate.

Wearable sensors are local devices in edge computing environments, usually installed on various parts of the body, such as the wrist and ankle. They obtain human behavioral data by measuring parameters such as acceleration, angular velocity, and direction. This data is then input into deep learning models for training and analysis, enabling automatic recognition of human activity [17]. Modern devices, such as smartphones and smartwatches, have integrated a range of sensors for detecting human behavior, including accelerometers, gyroscopes, and magnetometers. Researchers typically use data collected from these sensors and apply deep learning models to predict corresponding human actions. Jessica Sena proposed a deep CNN approach for human activity recognition by learning features between multiple sensors and analyzing the temporal relationships between sensor data [18]. Lin-Tao Duan designed a motion recorder that uses accelerometers and gyroscopes to record common limb movement data and identifies human behavior actions using three supervised learning algorithms [19].

Deep learning networks have shown excellent performance in processing sensor data to predict human activity. Over the last few years, several deep learning architectures have been proposed, including RNN [20], CNN [21], and a combination of the two (CNN-RNN) [22]. The long short-term memory recurrent neural network (LSTM) method [23], combined with CNN, has demonstrated exceptional performance in analyzing sensor signal data and has become the mainstream for human activity recognition methods [24].

Fan, Y.-C. et al. built a multi-stage deep learning network framework [25], which pre-processes and features data gathered from wearable sensors using variational autoencoders (VAE). Later, they utilized deep convolutional generative adversarial networks (GAN) to process data from VAE processes. Their study achieved an average accuracy of 88.4% on HAR, outperforming other RNN models.

Zhang, P. et al. proposed an LSTM-based evidence generation/combination model [26] that employs several parallel networks instead of one large network as well as a new evidence-theoretic algorithm to enhance the recognition ability of the model. Their approach is considered in terms of time consumption and scale. It achieved higher recognition accuracy than the traditional method on the UCI-HAR Dataset.

Ordonez proposed the DeepConvLSTM, a deep convolutional and LSTM recurrent neural network [27]. This network combines LSTM recursive and convolutional layers, providing a good trade-off between performance and runtime of the recursive architecture [28,29]. The framework they proposed is not only applicable to sensor modalities that are homogeneous, but it also has the capability to merge multi-modal sensors to enhance performance.

Finally, C. Xu proposed a novel deep learning model, InnoHAR [30], which employs separated convolution instead of traditional convolution and classifies activities using an association of inception-based and recurrent neural networks. This model utilizes convolutional neural networks to extract features from waveform data of various sizes and splice them together to extract human behavior features at different durations [31,32]. The experiments conducted demonstrate that their models exhibit superior performance compared to other state-of-the-art models.

The above studies demonstrated the effectiveness of deep learning methods in HAR. However, while addressing the limitations of traditional methods in HAR, deep learning models face challenges, such as increased time complexity and reduced recognition speed when attempting to enhance feature expression ability. Additionally, HAR cannot entirely eliminate the interference caused by erroneous or illegitimate data. Improving the accuracy of models in recognizing difficult-to-identify HAR data has become a pressing challenge in the field.

3. Methods

The proposed O-Inception Element-Wise-Attention GRU network (OI-EleAttG-GRU) architecture is designed for human activity recognition in edge computing environments. It consists of two steps. Initially, a dynamic scheduling algorithm (DTS) is utilized to assign human activity recognition tasks to suitable servers to alleviate local computing pressure. Then, the OI-EleAttG-GRU deep learning architecture, set on the server, executes the human activity recognition tasks and identifies corresponding actions before returning the recognition results to the local device. This architecture combines a novel O-Inception convolutional structure with a GRU block using an element-wise attention gate. The new O-Inception convolutional structure achieves a balance between training speed and recognition accuracy, while the GRU block with EleAttG adopts an effective attention mechanism to improve prediction accuracy. OI-EleAttG-GRU leverages the benefits of both, enabling accurate and fast human activity recognition.

3.1. HAR Task Offload in Edge Computing

The dynamic scheduling model for human activity recognition tasks in edge computing is shown in Figure 1. This model arranges HAR tasks with execution orders into an ordered queue, such as the HAR tasks recognized by smartphones being modeled as a1, a2, ... waiting for a series of subtasks, and the HAR tasks of the smartwatch are divided into b1, b2, waiting for a series of subtasks. Then utilizes a dynamic task scheduling (DTS) algorithm to offload each task in the queue to the corresponding edge server for execution. The proposed deep learning network model is predeployed on each edge server, and the HAR tasks will obtain the corresponding human actions after being identified by the network. Finally, the recognized human activity results will be returned to the wearable sensor.

3.1.1. Calculation Time of HAR Task

Human activity recognition tasks will be assigned to local or edge servers for execution as determined by the offload decision module. When local computing resources are sufficient, the human activity recognition task will be assigned to local execution, and the time for local execution, denoted as $T_{t_i}^{local}$, can be calculated using the following formula:

$$T_{t_i}^{local} = \frac{D_{t_i}}{f^{local}} \tag{1}$$

where f^{local} represents the calculated frequency of the sensing device, and D_{t_i} represents the amount of computation needed for the task.

When the offload decision module schedules human activity recognition tasks to the edge server for execution, its calculation time $T_{t_i}^{server}$ is as follows:

$$T_{t_i}^{server} = \frac{D_{t_i}}{f^{server}} \tag{2}$$

where f^{server} represents the calculation frequency of the server device.
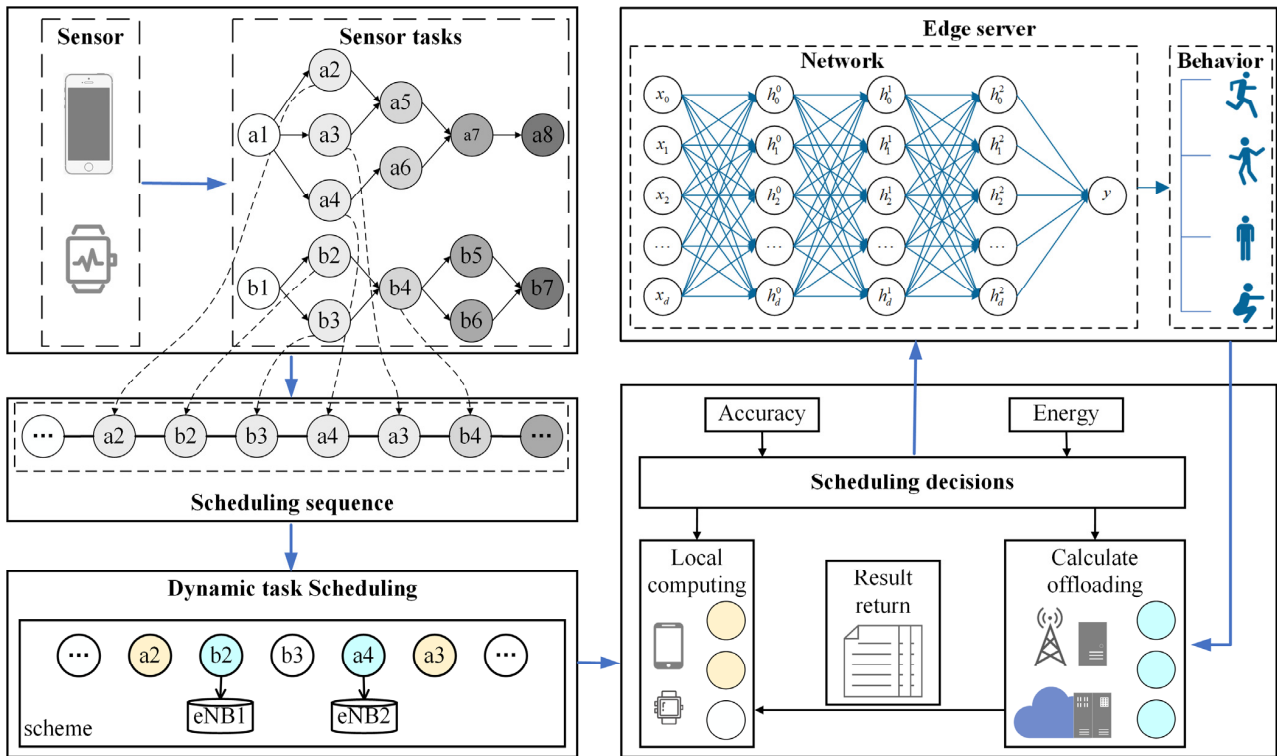
Figure 1. Dynamic scheduling model for HAR tasks in edge computing.

3.1.2. Transmission Time between Wearable Sensors and Servers

The human activity recognition task of wearable devices can be offloaded to the edge server to reduce the local computing load, and the network transmission problem in the mobile edge environment needs to be considered in the process of computing offloading. The transmission rate between the wearable device and the server in the mobile edge environment can be calculated by R_n, representing:

$$R_n = B \log_2(1 + f_{SNR}(d_{i,n})) \tag{3}$$

where B represents the transmission bandwidth between user equipment and servers. $f_{SNR}(d_{i,n})$ represents the signal-to-noise ratio within the channel.

After that, the transmission time for uploading HAR tasks to edge servers is calculated using transmission rate R_n and HAR task's data size DS_{t_i}:

$$T_{t_i}^{up} = \frac{DS_{t_i}}{R_n} \tag{4}$$

3.1.3. Execution Time of HAR on the Edge Server

The time for the user to offload the human activity recognition task to the edge server and acquire the result from the edge server includes the upload time for the uplink, the execution time for the application in the edge server, and the transmission time for the downlink edge server to return the result to the wearer. The total execution time of the HAR task on the server can be calculated by summing up the time for task upload, task processing, and result download, which is denoted as $T_{t_i}^{cloud}$:

$$T_{t_i}^{cloud} = T_{t_i}^{up} + T_{t_i}^{serve} + T_{t_i}^{down} \tag{5}$$

where $T_{t_i}^{down}$ is the task result download time. Due to the small amount of download data, the required download time is ignored.

3.1.4. Energy Consumption of HAR

The energy consumption of wearable devices for acquiring human activity data mainly includes two parts: energy consumption for local execution and energy consumption for task offloading, which includes upload energy consumption and download energy consumption. When computing the energy consumption of devices, the energy consumption of local running tasks should be considered. Local execution of tasks is allowed if the user does not have a server available, or if the task does not need to be offloaded to the server. The energy consumption of wearable devices calculated locally is as follows:

$$E_{t_i}^{local} = T_{t_i}^{local} P_{local} \tag{6}$$

where $T_{t_i}^{local}$ represents the time required by the local computing task and P_{local} is the voltage frequency of the computing task.

Task transmission energy consumption and result download energy consumption are:

$$E_{t_i}^{trans} = T_{t_i}^{trans} P_{trans}, trans = \{up, down\} \tag{7}$$

where $T_{t_i}^{state}$ represents the upload and download time of task transmission, and P_{trans} is the voltage of task transmission.

3.1.5. HAR Task Dynamic Scheduling Algorithm Design

A dynamic scheduling algorithm is proposed for scheduling the HAR tasks on appropriate edge servers for execution. The Dynamic Task Scheduling Method (DTS) begins with an initialization process (lines 1–2), where available servers are encoded, and the offloading position is initialized. If the offloading position is set to −1, the task will be offloaded locally, and the local overhead is denoted by p (line 3–6). Subsequently, the algorithm traverses the servers, calculating the overhead of offloading the task for each server. The values of l and p are updated if the current overhead is lower than the minimum overhead p found so far (lines 7–13); Finally, the task execution position is scheduled according to l (lines 14–16). The pseudocode of the algorithm is shown in Algorithm 1.

Algorithm 1: The HAR Dynamic Scheduling Method in Edge Computing Environment

---

Input: HAR task queue T, Edge server set S;
Output: Task offloading location l;
(1)   Initialization data
(2)   Encode the edge server
(3)   **for** i = 1, 2, 3...T **do**:
(4)   Calculate local HAR task computing time T^{local}
(5)   Calculate local HAR task computing energy consumption E^{local}
(6) $p = a^* E_{local} + T_{local}$;
(7)   **for** i = 1, 2, 3...S **do**:
(8)   Calculate transmission time T^{up}
(9)   Calculate of transmission energy consumption E^{trans}
(10)   Calculate server execution time T^{cloud}
(11)   **If** ($p < a^* E_{trans} + (t_{trans} + T_{excut})$)
(12) $l = i$
(13) $p = a^* E_{trans} + (t_{trans} + T_{excut})$
(14)   **end for**
(15)   Scheduling tasks
(16)   **end for**

3.2. Improved O-Inception Convolution Structure for HAR

As the progenitor of the O-Inception convolution structure, Inception's design principles offer valuable insights for human activity recognition. The core idea of Inception is to substitute a large convolution kernel with several small ones. This approach maximizes the

potential of convolution and pooling, integrates convolutional feature maps, and connects the outputs of each convolutional group. The use of Inception can mitigate the issue of exponentially increasing parameters that occurs when the network is deepened or widened.

Taking the commonly used GoogLeNet structure as an example, the inception Native structure increases the width of the network on the one hand and enhances its adaptability to scaling on the other. However, when faced with the high precision and rapid response speed requirements of human activity recognition, this simple solution has two main drawbacks. Primarily, a larger network usually means more parameters, making the extended network more susceptible to overfitting, particularly when the training set has a limited sample of labels. Subsequently, blindly increasing the size of the network can lead to an exponential increase in computational resource requirements.

To overcome the issues of increased time consumption and decreased recognition speed caused by improving feature expression ability in the original Inception architecture, we propose a new O-Inception convolutional structure. We use dark green to represent the Previous layer and filter Concatenation, this structure achieves higher accuracy and response speed in human activity recognition by replacing the 5×5 convolution in the original structure with 2×2 and 3×3 two-layer convolutions, replacing the original 3×3 pooling layer convolution with a 2×2 structure, and replacing the convolution after the pooling layer with 1×1 to 3×3 convolutions. The replaced convolution is represented in light yellow, and the replaced pooling layer is represented in light red. Three 1×1 convolutions that are not replaced are represented in dark yellow. The improved O-Inception module is shown in Figure 2.
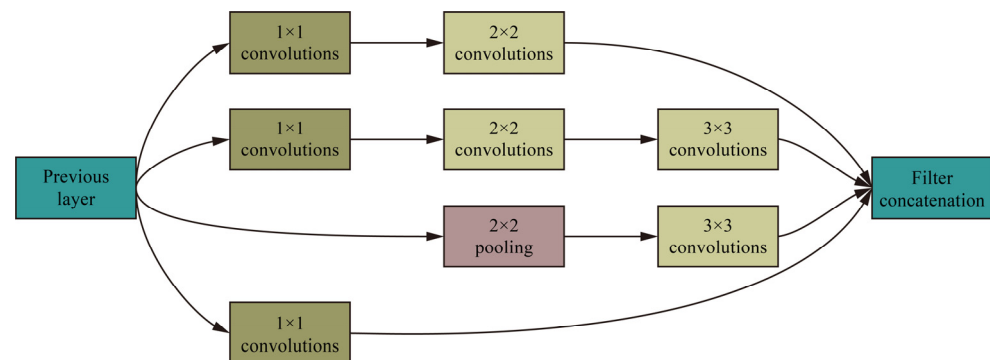

Figure 2. Improved O-Inception module.

The O-Inception structure uses the approach of replacing a large convolution kernel with multiple small convolution kernels to address the issue of excessive time complexity. This transformation significantly reduces the time complexity while keeping the feature map unchanged, thereby increasing the network's recognition and training speed. Moreover, the replacement of the pooling layer can reduce the loss of features to a certain extent. For instance, a 3×3 pooling operation will lose eight pieces of information while a 2×2 pooling operation only loses three pieces of information [33]. Finally, the O-Inception structure employs a 3×3 convolution to replace the 1×1 convolution and enhances the convolution field of view by increasing the field of view window, thereby improving the model's accuracy.

3.3. EleAttG-GRU Block for More Accurate HAR Prediction

Standard RNN (sRNN), LSTM, GRU, etc., have proven to be effective for working on sequence information, such as human activity recognition [5]. Additionally, according to research, the attention mechanism that selectively focuses on different data elements is also effective for human activity recognition. Therefore, combining the gating mechanism and attention, we use an Element-wise-Attention Gate to provide attention to RNN neurons, allowing these RNN neurons to gain the ability to adaptively focus on the key elements of the input. We use a shareable EleAttG with the same dimension of the output attention

vector as the input to apply to all neurons of the RNN block. The original input undergoes modulation by the attention vector, thereby amplifying the significance of crucial information on the analysis results.

Recurrent neural networks can model the spatial correlation and temporal dynamics of time series. To better illustrate the features of EleAttG-GRU, we briefly compare it with the widely used RNN structure. The output response r_t of time step t is calculated from the input x_t of this layer, and the output r_t of the last phase is as follows:

$$r_t = tanh(W_{xr}x_t + W_{rr}r_{t-1} + \varepsilon_r) \tag{8}$$

where W_{ab} represents the weight matrix related to a and b, and ε_c is the bias vector, $a \in \{x, r\}$, $b \in \{r\}$ and $c \in \{r\}$.

We use EleAttG to give RNN neurons attention ability. a_t is a response vector, and its dimension is the same as the previous RNN's input x_t. The calculation formulas are as follows:

$$a_t = \varphi(W_{xa}x_t + W_{ra}r_{t-1} + \varepsilon_a) \tag{9}$$

where φ represents the activation function of Sigmoid. The importance level of the input $\widetilde{x}_t$ is determined by the current input r_{t-1} and the previous hidden state x_t. By means of the attentional response mechanism, the input x is updated, represented as:

$$\widetilde{x}_t = x_t \odot a_t \tag{10}$$

Afterwards, the GRU network will perform a recursive calculation based on updated input x.

The GRU is a type of RNN that was designed to tackle the problems of gradient vanishing and long-term memory. It consists of two types of gates: update gates and reset gates. The former determines how much of the previous state's information should be carried over to the current state, while the latter controls how much of the new input should be incorporated into the current state. When EleAttG is added to GRU blocks, it endows RNN neurons with the ability to selectively attend to key elements in the input sequence. The calculation formulas for EleAttG-GRU blocks are as follows:

$$r_t = \sigma(W_r \odot [h_{t-1}, a_t \odot x_t] + \varepsilon_r) \tag{11}$$

$$z_t = \sigma(W_z \odot [h_{t-1}, a_t \odot x_t] + \varepsilon_z) \tag{12}$$

$$\widetilde{h}_t = tanh(W \odot [r_t \times h_{t-1}, a_t \odot x_t] + \varepsilon_h) \tag{13}$$

$$h_t = h_{t-1} + z_t\left(\widetilde{h}_t - h_{t-1}\right) \tag{14}$$

where h_t refers to the output vector of the hidden state. r_t is the reset gate and z_t is the update gate. a_t is a response vector, W represents the corresponding weight matrix, *tanh*, and σ denotes Activation functions. $\widetilde{h}_t$ is the vector after being activated, and ε is the bias vector.

Figure 3 illustrates the EleAttG-GRU structure with Element-Wise-Attention. The EleAttG is marked in red, the GRU network structure is marked in green, the calculation unit is marked in yellow, and each line represents a vector. We use the response of a_t, the EleAttG, to modulate X_t to $\widetilde{X}_t$, and then replaced X_t with $\widetilde{X}_t$ to perform the follow-up. this can be defined as the EleAttG-GRU.

EleAttG allows for fine-grained adaptive tuning of inputs, which is especially important for tasks that involve complex and varied inputs, such as human activity recognition. With EleAttG, the network can selectively focus on the most relevant features within each input, analyzing different elements with varying levels of attention to achieve more accurate results.
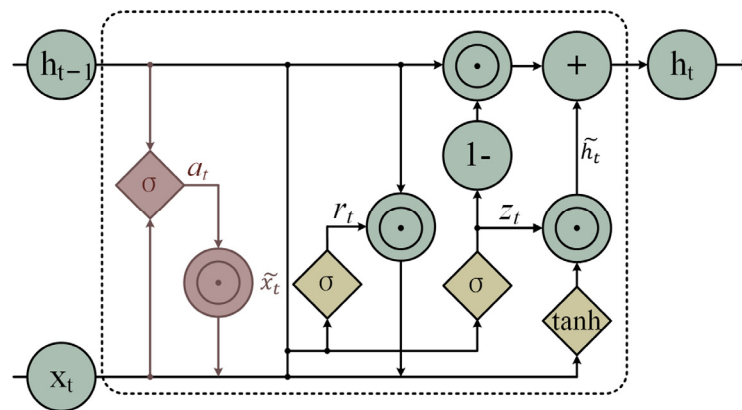
Figure 3. EleAttG-GRU structure with Element-Wise-Attention.

Moreover, EleAttG has fewer parameters than LSTM, yet it leads to a 7–15% improvement in accuracy and performance when recognizing continuous actions with different features. The network overcomes the problem of time-series omission caused by data analysis overlooking the correlation, as well as the issue of diminishing impact from early input data in long-term dependency that is caused by recurrent input. This contributes to an improvement in the recognition accuracy when facing continuous activities.

3.4. Detailed Process of OI-EleattG-GRU in HAR

In the learning process for identifying human activity from sensor signal data, as depicted in Figure 4, there are two analysis methods that can identify human activity using data from various types of sensors (gyroscopes, magnetometers, and accelerometers) obtained at a specified sampling rate. The traditional method is shown in flow chart A. Informative features are extracted in a manually produced manner, followed by applying a sliding window method with variable size on real-time acceleration data. The processed data is then fed into the classifier for training and testing of the classification model (k-means, artificial neural network, naive bayes, random forest, etc.). Flowchart B shown in Figure 4 illustrates a typical process using a deep learning approach where data can be processed from different types of deep network structures, and where results are ultimately obtained using SoftMax classification. The method shown in Flowchart B was used in this study.

The deep network in Flowchart B uses a new OI-EleAttG-GRU structure to achieve accurate human activity recognition. The network structure of proposed OI-EleAttG-GRU combines a new O-Inception convolutional structure with EleAttG-GRU blocks, which can be divided into three modules: sensor data preprocessing, deep learning model training, and prediction result classification output. This creates the OI-EleAttG-GRU Deep Learning Model which is shown in Figure 5.

When extracting features from the waveform data of the sensor, after the convolution and pooling operation was repeated four times, we used the O-Inception module introduced above. O-Inception reduces noise interference and improves network judgment accuracy through recognition rate improvement. After that, flatten and double-layer EleAttG-GRU are used to extract the time features. Finally, this is transferred to the SoftMax classifier to classify the results.

In addition to the high recognition rate, this model also has a good fault tolerance performance because of the EleAttG-GRU network we added after CNN's O-Inception structure. When there are more wrong data or illegal data in the initial data, the recognition rate of the traditional CNN network will drop significantly or even appear to be untrained. On the other hand, EleAttG-GRU has the time-dependent characteristics of GRU, which correspond to multiple feature maps at various times. When the wrong data is input, EleAttG-GRU can predict the wrong data according to the correct data of other feature maps through time-dependent characteristics, thereby weakening the impact of the wrong data on the recognition rate during the training process.
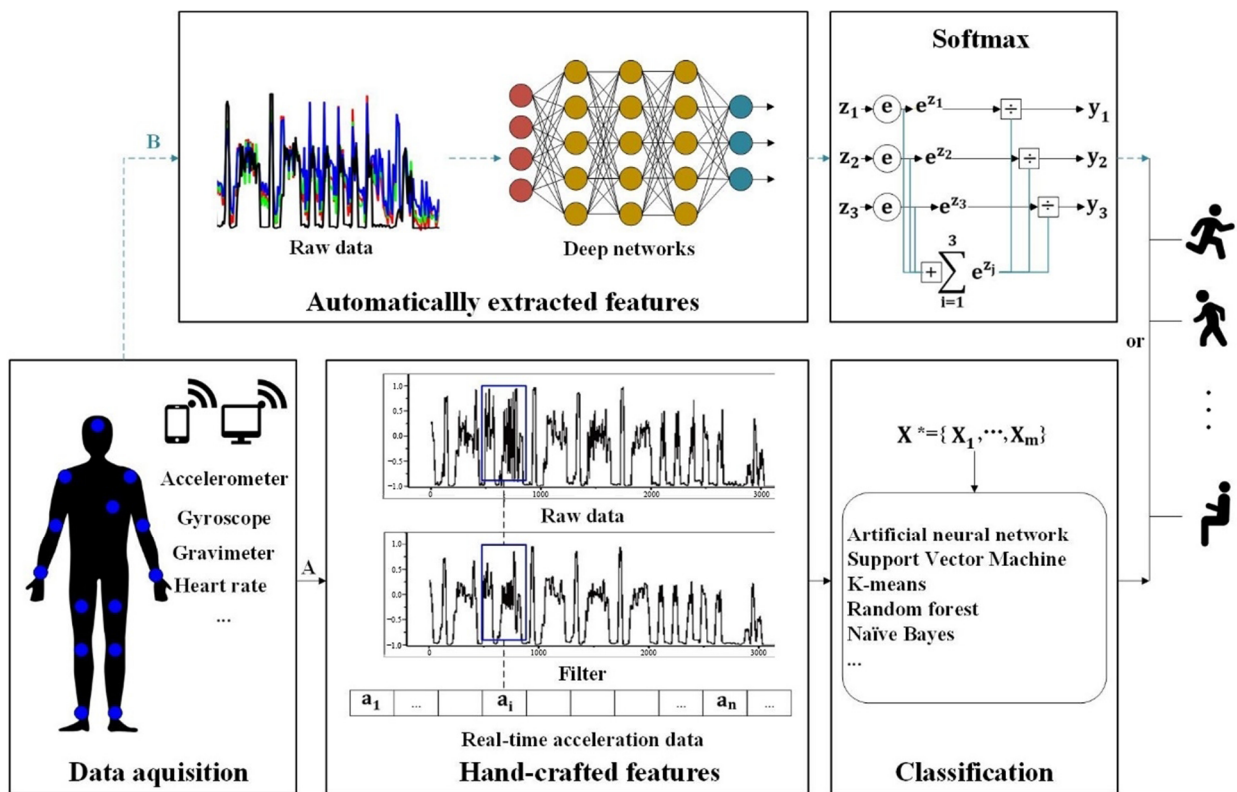
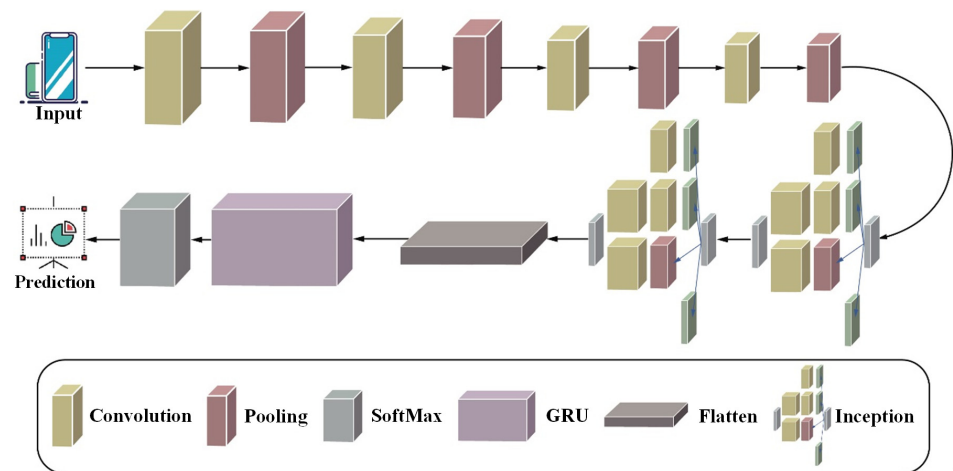**Figure 4.** Learning process for identifying human activity from sensor signal data.

Figure 5. Working of OI-EleAttG-GRU Deep Learning Model.

4. HAR Datasets and Experiment Settings

4.1. HAR Datasets

To demonstrate the effectiveness of the proposed OI-EleAttG-GRU architecture in human activity recognition, we analyzed three datasets with different features. Two of these datasets can be obtained from the UCI Machine Learning Repository, and they have been widely used in human activity recognition research. The third dataset was self-made, containing human daily behavioral actions, which can better reflect the structure of human behavior.

4.1.1. WISDM Datasets

The Wireless Sensor Data Mining (WISDM) dataset is a publicly available dataset of accelerometer data collected from smartphones for studying human activity recognition. The data was collected from 36 participants, each carrying a smartphone and engaging in

six different activities, including walking, running, going up and down stairs, sitting, and standing. This dataset contains 1098 data records, each consisting of three acceleration data axes collected over a fixed time period.

This dataset is widely used in the field of HAR, and many researchers have evaluated the accuracy of activity recognition algorithms that use it. The dataset has been preprocessed to remove noise and resampled to a fixed sampling rate. Additionally, the data is divided into a 70% training set and a 30% testing set, which makes it suitable for testing and comparing the performance of various machine learning and deep learning algorithms.

4.1.2. UCI Datasets

The UCI dataset is a dataset proposed by the University of California, Irvine, specifically designed for machine learning purposes. the UCI HAR dataset includes accelerometer and gyroscope data for six activities performed by 30 participants, namely walking, jogging, climbing stairs, descending stairs, sitting, and standing. Researchers have widely utilized this dataset to evaluate the accuracy of various machine learning and deep learning algorithms in identifying human behavior. Furthermore, the dataset has been divided into training and testing subsets, which makes it a valuable resource for researchers to test and measure the accuracy of different machine learning and deep learning algorithms in the field of human activity recognition.

4.1.3. PHAD Datasets

Phyphox Human Activity Dataset (PHAD) contains the experimental data from a smartphone's magnetometer, gyroscope, and accelerometer, capturing various motion information. We custom-made this dataset to facilitate better data collection for experimental testing. During the preprocessing process, we distinguished between regular and confused data. A group of 30 volunteers was invited to complete six human activities including walking (WK), sitting (ST), standing (SD), running (RN), walking upstairs (WU) and walking downstairs (WD), as shown in Figure 6.
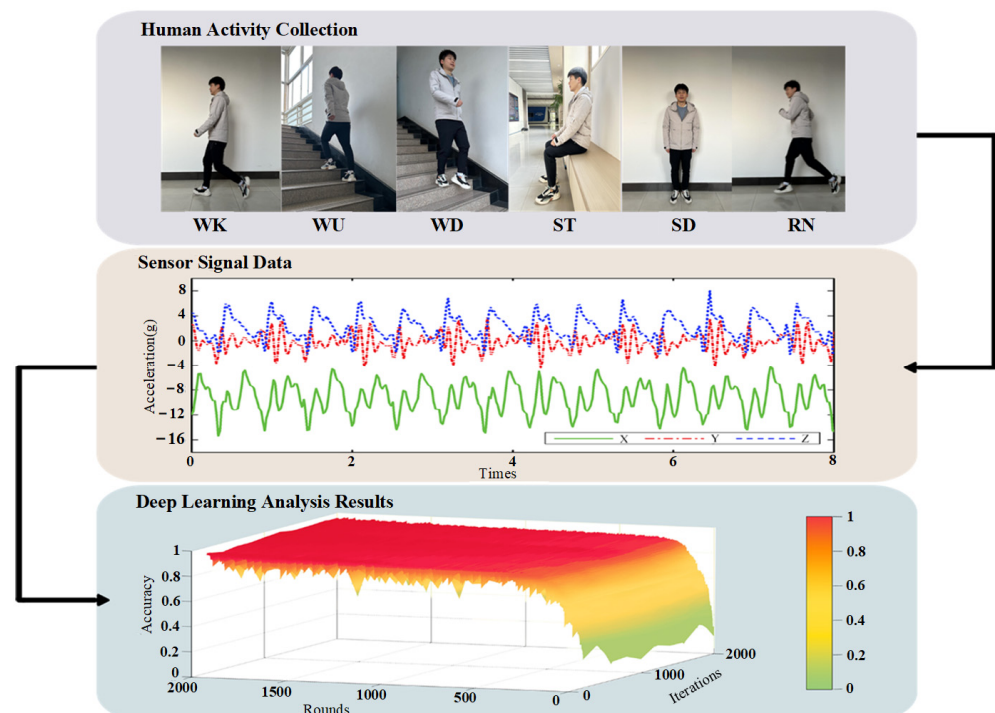

Figure 6. Illustration of process of deep learning, data is collected by sensors on typical daily human behaviors, and the OI-EleAttG-GRU network is used for predictive analysis.

Figure 6 provides a diagram of the six types of human activity and the process of deep learning. The top part of the figure represents a scenario where the human body collects different motion information. The middle part shows the signal data received by the waist and arm sensors, capturing the aforementioned six human behaviors. These time-varying speed data are recorded, processed, and then fed into the deep learning network for model training. The iterative increase in accuracy during training is shown at the bottom of the figure.

These activities are selected because these are common activities that people regularly perform in their daily routines. Analyzing results of these activities provides the best reflection of the performance of our human activity recognition structure. During the experiment, each volunteer wore a smartphone attached to both their waist and arm to collect various behavioral information for a duration of one minute. The various signal data are collected by the sensor including accelerometer, gyroscope and magnetometer. The detailed class distribution of HAR data and the proportion of mixed data is presented in Table 1. The 200 data items with the highest probability of classification error by conventional methods are marked as confused data.

Table 1. The class distribution of HAR data and the proportion of mixed data.

Activity	Distribution	Confused
Walking	36,427 (24.10%)	86 (43.00%)
Sitting	26,815 (17.75%)	3 (1.50%)
Standing	31,764 (21.03%)	8 (4.00%)
Running	29,589 (19.58%)	37 (18.50%)
Walking Upstairs	13,431 (8.89%)	37 (18.50%)
Walking Downstairs	13,067 (8.65%)	29 (14.50%)

4.2. Experiment Settings

Experiments are conducted on TensorFlow's deep learning platform, and various parameters are adjusted for different datasets to achieve better results. For wireless sensor data mining action recognition, due to the different characteristics between different datasets, the initial learning rate is typically set to different constants, including 0.1, 0.01, 0.001, and 0.0001. After multiple accuracy comparisons, it was found that the optimal parameter for the UCI dataset lies between 0.001 and 0.01. Subsequently, using 0.001 as the step size, the final initial learning rate for the UCI dataset was chosen as 0.005. Similarly, we selected 0.002 as the initial learning rates for the model on the PHAD dataset.

The sensor signal data is divided according to each action segment. The three-axis acceleration data for the action behavior of walking is shown in Figure 7a, and gyroscope data is shown in Figure 7b.
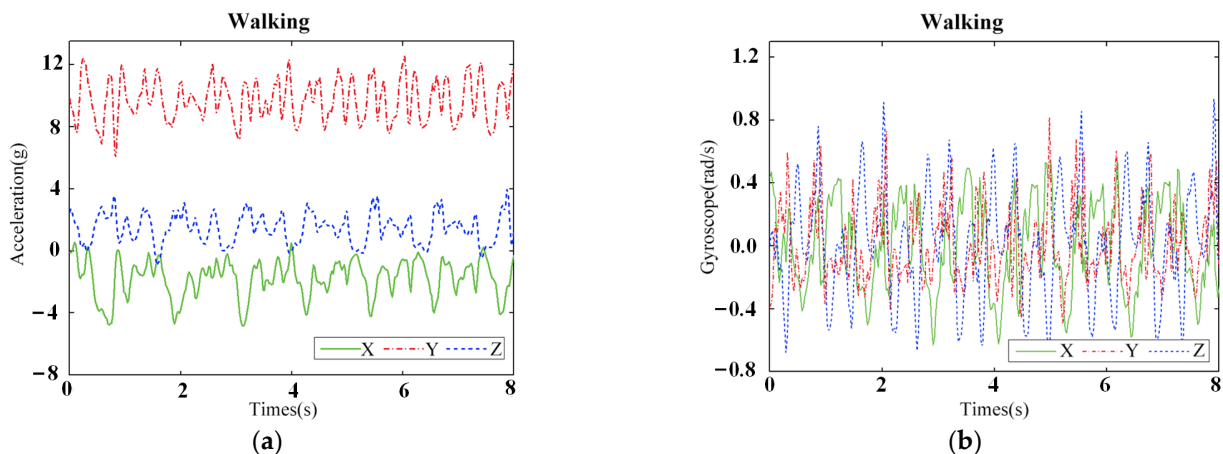

(a)　　　　　　　　　　　　　　　　　　　　　　　　**(b)**

Figure 7. Signals of walking behavior collected by sensors on the arm. (**a**) Acceleration data, (**b**) Gyroscope data.

Optimizers including BGD, SGD, Momentum, Adadelta, RMSprop, and Adam [34] were compared for each dataset. The Adam optimizer yielded the highest accuracy and was chosen for implementation in the model. Additionally, signal-based classification of human activity typically relies on precision and recall metrics to evaluate the performance of the methods that were used in this study to assess the OI-EleAttG-GRU method.

5. Results and Analysis

5.1. Scheduling Performance

In order to analyze and evaluate the performance of the proposed dynamic task scheduling (DTS) algorithm (Algorithm 1), this algorithm together with the OI-EleAttG-GRU is compared with LC (local computing) and OLC (online computing) in terms of energy consumption and accuracy for the task of identifying different numbers of human activity. Figure 8 shows the accuracy versus HAR task number and energy cost versus HAR task number, respectively.
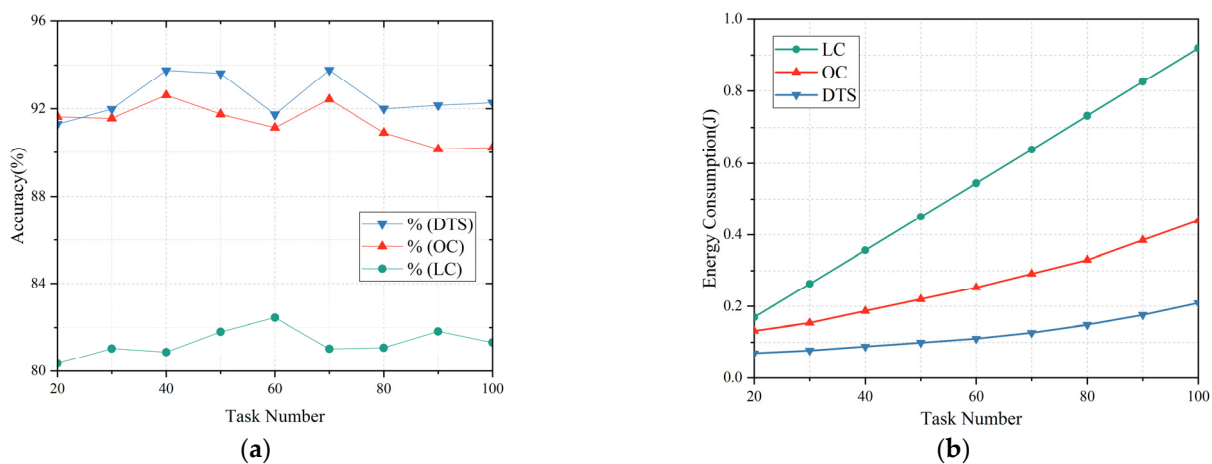

(a) **(b)**

Figure 8. (**a**) Accuracy versus HAR task number. (**b**) Energy cost versus HAR task number.

Figure 8 shows the optimization comparison of the DTS algorithm with the LC and OLC algorithms in terms of accuracy and energy cost, respectively, with the increase of human activity recognition task numbers. From the figure, the optimization effect of accuracy and energy consumption of the DTS algorithm is much better than that of the LC and OLC algorithms. This improvement is ascribed to dynamic scheduling strategy employed by the DTS algorithm, which enables the rational allocation of computing burden while effectively mitigating the impact of changes in the network environment on task scheduling.

5.2. Impact of Dataset Partitioning Ratio

To validate the performance of OI-EleAttG-GRU on different proportions of training and testing sets, we kept all other parameters unchanged and increased the original 70% training set to 80%. Subsequently, a comparison was made based on accuracy and F1-score metrics. The experimental results are presented in Table 2.

Table 2. Comparison of accuracy and F1-score among different partition ratios (Training set: Test set) on the WISDM dataset.

Evaluation	Partition Ratio (Training Set: Test Set)	
	(70%:30%)	**(80%:20%)**
Training_accuracy	98.58%	99.14%
Training_F1-score	98.59%	99.15%
Test_accuracy	97.52%	97.50%
Test_F1-score	97.51%	97.50%

Table 2 indicates that having an 80% training set means there are more samples to learn the features of HAR data. Consequently, under the same number of training epochs, the model can better fit the data, resulting in higher accuracy on the training set. However, the difference in performance between the two sets on the testing set is small. The accuracy and F1-score of the 80% training set are 0.02% and 0.01% lower than those of the 70% training set. For the 70% training set, although there are relatively fewer training samples, the model might be more generalized, enabling better generalization to unseen data.

5.3. Classification Performance

To measure the effectiveness of the proposed OI-EleAttG-GRU (OEG) method, three effective algorithms are used in three human activity datasets, including Inception Neural Network HAR (IN), Support Vector Machine (SVM) and Deep Convolutional and LSTM Recurrent Neural Networks (DCL). The precision comparison in recognizing different human activity in percentages are listed in Table 3.

Table 3. Precision Comparison of OI-EleAttG-GRU (OEG) and existing methods in recognizing different human activity in percentages (%).

Performance Evolution	OEG	SVM	IN	DCL	OI-EleAttG-GRU		
					SVM	IN	DCL
Walking	97.5	81.6	94.8	93.4	15.9	2.7	4.1
Walking Upstairs	94.2	74.6	91.2	87.3	19.6	3.0	6.9
Walking Downstairs	93.6	77.5	88.9	84.4	16.1	4.7	9.2
Sitting	99.1	90.5	96.3	98.6	8.6	2.8	0.5
Standing	97.6	92.4	98.7	97.3	5.2	1.1	0.3
Running	91.8	79.3	86.2	91.4	12.5	5.6	0.4
Overall Precision	95.6	82.7	92.7	92.1	12.9	2.9	3.5

Figure 9 illustrates the comparison of precision and recall on six human activity class sets. Table 3 and Figure 9 show that the proposed OEG is about 12.9%, 2.9%, and 3.5% higher than the other three methods in precision, respectively. In general, the performance of OEG is high, and it has high precision and recall rate.
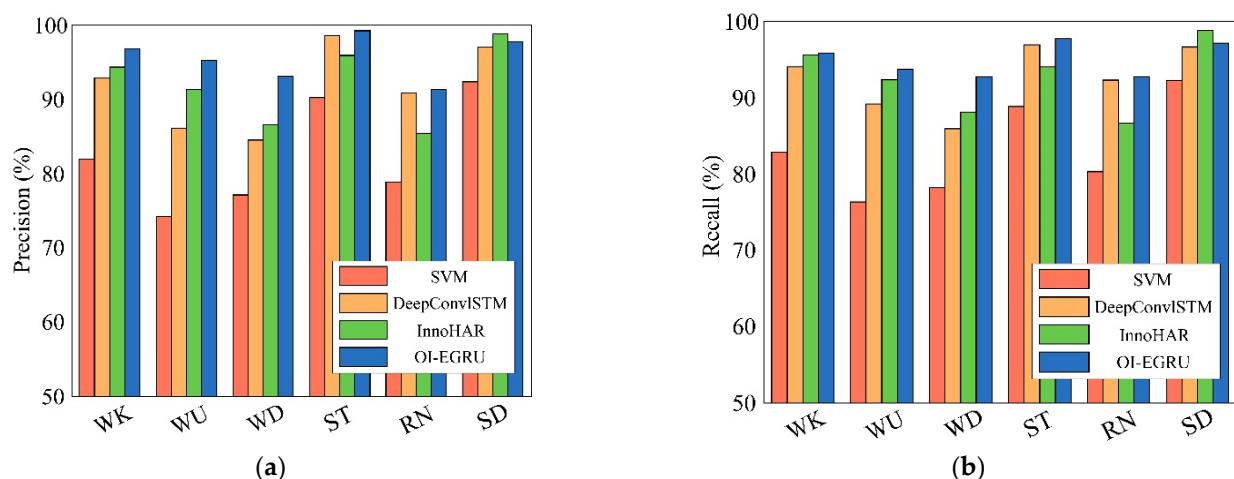

Figure 9. (**a**) Comparison of precision between OI-EleAttG-GRU and three methods on six human activity class sets. (**b**) Comparison of recall between OI-EleAttG-GRU and three methods on six human activity class sets.

The experimental results show that the EleAttG-GRU structure achieves a significant recognition effect in the human activity classification model. For different datasets, we conducted multiple tests to adjust the number of neurons, the batch size, and the EleAttG-GRU layer to achieve better results.

5.4. Comparison with Current Work

Compare OI-EleAttG-GRU with other current advanced models, including LCNN [23], DLFC [35], CNN-GRU [36], BiLSTM [37], as well as other effective models, such as SVM, IN, DCL on the two representative public datasets WISDM and UCI. The comparison is based on two common comprehensive performance metrics: Accuracy and F1-score. The comparison between existing models on the WISDM and UCI dataset is presented in Figure 10.
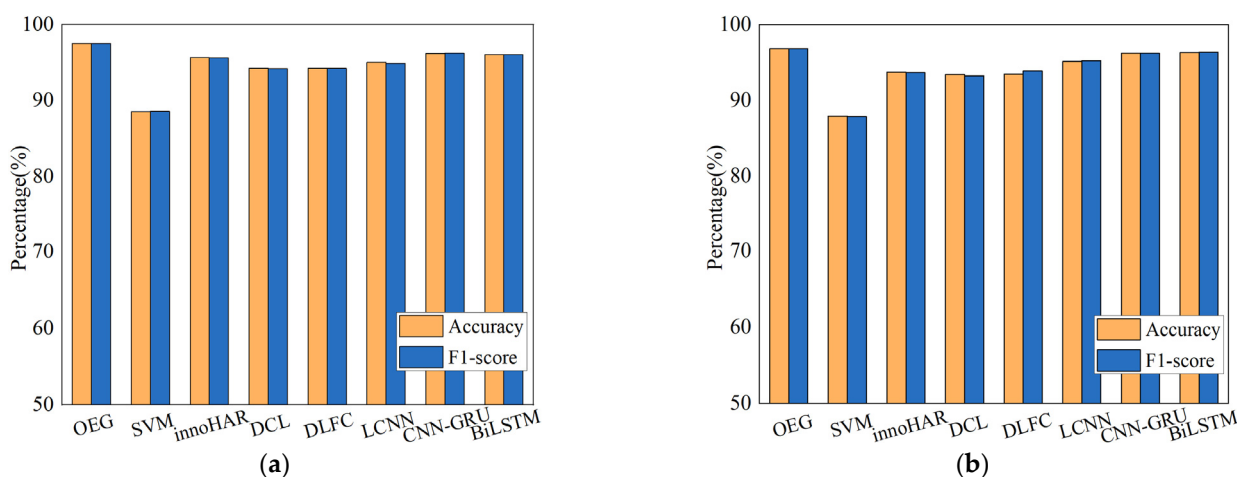

Figure 10. (**a**) Comparison between existing models on the WISDM dataset. (**b**) Comparison between existing models on the UCI dataset.

According to the results shown in Figure 10a,b, the OEG method achieved impressive performance on the WISDM dataset. The accuracy reached 97.52%, and the F1-score reached 97.51%. When compared to the baseline method SVM, OEG exhibited significant improvements of 8.99% and 8.96% in accuracy and F1 score, respectively. Even when compared to current advanced methods, OEG still demonstrated noticeable accuracy improvements of 1.31% to 2.70% and F1-score improvements of 1.29% to 3.25%.

Furthermore, OEG also performed well on the UCI dataset, surpassing the baseline method with an increase of 8.9% in accuracy and 9% in F1-score. Compared to the current advanced methods, OEG achieved improvements of 0.52% to 3.4% in accuracy and 0.48% to 2.96% in F1 score.

Overall, due to its ability to capture the long-term dependencies in sensor data, OI-EleAttG-GRU demonstrates a performance advantage, displaying excellent temporal and spatial feature extraction capabilities, and generalizing well across different datasets when dealing with complex or noisy data.

5.5. Identification Performance of Confused Data

Figure 11 shows the confusion matrix of a state-of-the-arts method on the left and proposed method on the right in one test run. The column shape from dark blue to light yellow represents a classification precision from 1 to 0. The darker the diagonal square is, the higher the accuracy of the classification result. The lighter the other squares, the lower the error rate of the classification result.

Among the 200 confusing data previously screened out, the RNN method has an average accuracy of 77.1%, and the OI-EleAttG-GRU method has an average accuracy of 85.4%, which is a significant increase. These outcomes imply that the proposed method is more advantageous in recognizing human activity recognition compared to RNN. This improvement may be attributed to the combination of inception and EleAttG-GRU with multi-scale and spatiotemporal feature extraction.
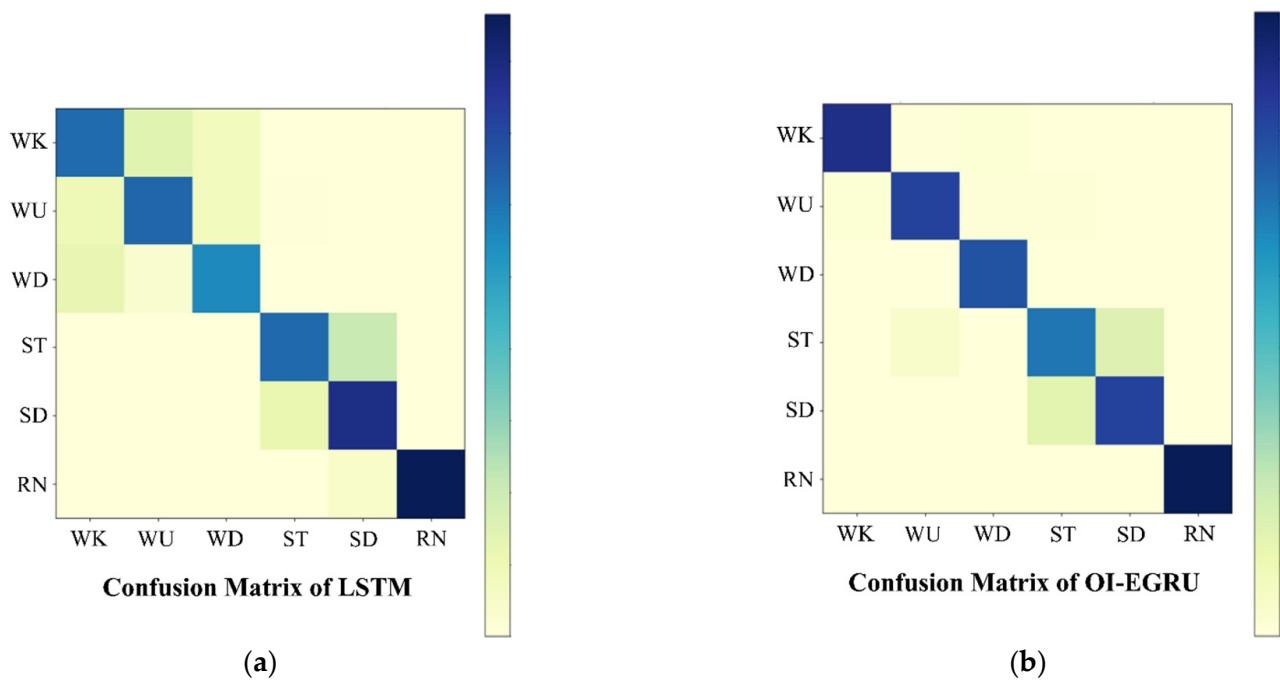
Figure 11. Confusion matrix for two methods in one test. (**a**) The confusion matrix for RNN. (**b**) The confusion matrix for the proposed method.

Despite achieving good recognition results in identifying various complex actions, it needs to be further considered that users can generate different and continuous behavioral actions over a period of time, and the accuracy of segmenting each action can be improved. This will be included in our future work.

6. Conclusions and Future Work

This paper proposed a novel OI-EleAttG-GRU deep learning model for HAR in an edge computing environment. Using a dynamic scheduling algorithm, we were able to offload wearable computing tasks to the edge server and meet HAR computing resource requirements. Moreover, an improved GRU network model is employed at the edge server to improve the recognition accuracy of HAR by using an inception convolution kernel spliced with a GRU network that possesses Element-Wise-Attention properties. With extensive experiments on different datasets, our proposed method demonstrated superior performance compared to both baseline and state-of-the-art methods. Specifically, we dynamically considered the computing frequency and power consumption of wearable devices and servers in task offloading, which significantly reduces the execution time and energy consumption of task offloading. We use inception convolution in data processing to enhance the ability to analyze multi-scale data. In the process of behavior recognition, the network significantly improves the ability to extract features from spatial-temporal features through the EleAttG-GRU's temporal memory capability and attention mechanism.

As the performance for deep learning on HAR in an edge computing environment still has room for improvement, in the future, we will further study the following limitations of edge computing: (i) limited computing resources of some servers; (ii) privacy security protection; and (iii) network and bandwidth delays. In terms of deep learning, improving the recognition accuracy of different continuous actions on larger and more complex datasets is also an important aspect that we need to consider in our future work. We believe that the outcomes of such efforts can benefit various domains, including human health and biomedical applications.

References

1. Hassan, M.M.; Uddin, M.Z.; Mohamed, A.; Almogren, A. A robust human activity recognition system using smartphone sensors and deep learning. *Future Gener. Comp. Syst.* **2018**, *81*, 307–313. [CrossRef]
2. Carter, E.; Adam, P.; Tsakis, D.; Shaw, S.; Watson, R.; Ryan, P. Enhancing pedestrian mobility in Smart Cities using Big Data. *J. Manag. Anal.* **2020**, *7*, 173–188. [CrossRef]
3. Ismail, W.N.; Hassan, M.M.; Alsalamah, H.A. Context-Enriched Regular Human Behavioral Pattern Detection from Body Sensors Data. *IEEE Access* **2019**, *7*, 33834–33850. [CrossRef]
4. Subasi, A.; Khateeb, K.; Brahimi, T.; Sarirete, A. Chapter 5—Human activity recognition using machine learning methods in a smart healthcare environment. In *Innovation in Health Informatics*; Lytras, M.D., Sarirete, A., Eds.; Academic Press: Cambridge, MA, USA, 2020; pp. 123–144.
5. Zhang, S.B.; Li, Y.X.; Zhang, S.; Shahabi, F.; Xia, S.; Deng, Y.; Alshurafa, N. Deep Learning in Human Activity Recognition with Wearable Sensors: A Review on Advances. *Sensors* **2022**, *22*, 1476. [CrossRef]
6. Bianchi, V.; Bassoli, M.; Lombardo, G.; Fornacciari, P.; Mordonini, M.; De Munari, I. IoT Wearable Sensor and Deep Learning: An Integrated Approach for Personalized Human Activity Recognition in a Smart Home Environment. *IEEE Internet Things* **2019**, *6*, 8553–8562. [CrossRef]
7. Nain, G.; Pattanaik, K.K.; Sharma, G.K. Towards edge computing in intelligent manufacturing: Past, present and future. *J. Manuf. Syst.* **2022**, *62*, 588–611. [CrossRef]
8. Kim, J.; Moon, J.; Hwang, E.; Kang, P. Recurrent inception convolution neural network for multi short-term load forecasting. *Energ. Build.* **2019**, *194*, 328–341. [CrossRef]
9. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1800–1807.
10. Zhang, P.; Xue, J.; Lan, C.; Zeng, W.; Gao, Z.; Zheng, N. EleAtt-RNN: Adding Attentiveness to Neurons in Recurrent Neural Networks. *IEEE Trans. Image Process.* **2020**, *29*, 1061–1073. [CrossRef]
11. Pallis, G.; Vakali, A. Insight and perspectives for content delivery networks. *Commun. Acm* **2006**, *49*, 101–106. [CrossRef]
12. Satyanarayanan, M.; Bahl, P.; Caceres, R.; Davies, N. The Case for VM-Based Cloudlets in Mobile Computing. *IEEE Pervas Comput.* **2009**, *8*, 14–23. [CrossRef]
13. Ali, F.A.; Simoens, P.; Verbelen, T.; Demeester, P.; Dhoedt, B. Mobile device power models for energy efficient dynamic offloading at runtime. *J. Syst. Softw.* **2016**, *113*, 173–187. [CrossRef]
14. Yu, Y.F. Mobile Edge Computing towards 5G: Vision, Recent Progress, and Open Challenges. *China Commun.* **2016**, *13*, 89–99. [CrossRef]
15. Shi, W.S.; Cao, J.; Zhang, Q.; Li, Y.H.Z.; Xu, L.Y. Edge Computing: Vision and Challenges. *IEEE Internet Things* **2016**, *3*, 637–646. [CrossRef]
16. Satyanarayanan, M. The Emergence of Edge Computing. *Computer* **2017**, *50*, 30–39. [CrossRef]
17. Wang, X.; Yu, H.; Kold, S.; Rahbek, O.; Bai, S. Wearable sensors for activity monitoring and motion control: A review. *Biomim. Intell. Robot.* **2023**, *3*, 100089. [CrossRef]
18. Sena, J.; Barreto, J.; Caetano, C.; Cramer, G.; Schwartz, W.R. Human activity recognition based on smartphone and wearable sensors using multiscale DCNN ensemble. *Neurocomputing* **2021**, *444*, 226–243. [CrossRef]
19. Duan, L.T.; Lawo, M.; Wang, Z.G.; Wang, H.Y. Human Lower Limb Motion Capture and Recognition Based on Smartphones. *Sensors* **2022**, *22*, 5273. [CrossRef] [PubMed]
20. Pienaar, S.W.; Malekian, R. Human Activity Recognition using LSTM-RNN Deep. Neural Network Architecture. In Proceedings of the 2019 IEEE 2nd Wireless Africa Conference (WAC), Pretoria, South Africa, 18–20 August 2019; pp. 1–5.

21. Münzner, S.; Schmidt, P.; Reiss, A.; Hanselmann, M.; Stiefelhagen, R.; Dürichen, R. CNN-based sensor fusion techniques for multimodal human activity recognition. In Proceedings of the 2017 ACM International Symposium on Wearable Computers; Association for Computing Machinery: Maui, Hawaii, 2017; pp. 158–165.

22. Ma, H.; Li, W.; Zhang, X.; Gao, S.; Lu, S. AttnSense: Multi-level Attention Mechanism For Multimodal Human Activity Recognition. In Proceedings of the International Joint Conferences on Artificial Intelligence Organization, Macao, China, 10–16 August 2019; pp. 3109–3115.

23. Xia, K.; Huang, J.G.; Wang, H.Y. LSTM-CNN Architecture for Human Activity Recognition. *IEEE Access* **2020**, *8*, 56855–56866. [CrossRef]

24. Wang, X.H.; Gao, L.L.; Song, J.K.; Shen, H.T. Beyond Frame-level CNN: Saliency-Aware 3-D CNN With LSTM for Video Action Recognition. *IEEE Signal Proc. Lett.* **2017**, *24*, 510–514. [CrossRef]

25. Fan, Y.C.; Tseng, Y.H.; Wen, C.Y. A Novel Deep Neural Network Method for HAR-Based Team Training Using Body-Worn Inertial Sensors. *Sensors* **2022**, *22*, 8507. [CrossRef]

26. Zhang, P.; Zhang, Z.J.; Chao, H.C. A Stacked Human Activity Recognition Model Based on Parallel Recurrent Network and Time Series Evidence Theory. *Sensors* **2020**, *20*, 4016. [CrossRef] [PubMed]

27. Ordonez, F.J.; Roggen, D. Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition. *Sensors* **2016**, *16*, 115. [CrossRef] [PubMed]

28. Hammerla, N.Y.; Halloran, S.; Ploetz, T. Deep, Convolutional, and Recurrent Models for Human Activity Recogniti on Using Wearables. *arXiv* **2016**, arXiv:1604.08880.

29. Kelotra, A.; Pandey, P. Stock Market Prediction Using Optimized Deep-ConvLSTM Model. *Big Data* **2020**, *8*, 5–24. [CrossRef]

30. Xu, C.; Chai, D.; He, J.; Zhang, X.T.; Duan, S.H. InnoHAR: A Deep Neural Network for Complex Human Activity Recognition. *IEEE Access* **2019**, *7*, 9893–9902. [CrossRef]

31. Ignatov, A. Real-time human activity recognition from accelerometer data using Convolutional Neural Networks. *Appl. Soft Comput.* **2018**, *62*, 915–922. [CrossRef]

32. Ma, C.Y.; Chen, M.H.; Kira, Z.; AlRegib, G. TS-LSTM and temporal-inception: Exploiting spatiotemporal dynamics for activity recognition. *Signal Process-Image* **2019**, *71*, 76–87. [CrossRef]

33. Akhtar, N.; Ragavendran, U. Interpretation of intelligence in CNN-pooling processes: A methodological survey. *Neural Comput. Appl.* **2020**, *32*, 879–898. [CrossRef]

34. Mu, R.H.; Zeng, X.Q. A Review of Deep Learning Research. *Ksii Trans. Internet Inf. Syst.* **2019**, *13*, 1738–1764.

35. Akter, M.; Ansary, S.; Khan, M.A.-M.; Kim, D. Human Activity Recognition Using Attention-Mechanism-Based Deep Learning Feature Combination. *Sensors* **2023**, *23*, 5715. [CrossRef]

36. Dua, N.; Singh, S.N.; Semwal, V.B. Multi-input CNN-GRU based human activity recognition using wearable sensors. *Computing* **2021**, *103*, 1461–1478. [CrossRef]

37. Challa, S.K.; Kumar, A.; Semwal, V.B. A multibranch CNN-BiLSTM model for human activity recognition using wearable sensor data. *Vis. Comput.* **2022**, *38*, 4095–4109. [CrossRef]