applied sciences

MDPI

Article

Human Activity Recognition by the Image Type Encoding Method of 3-Axial Sensor Data

Changmin Kim [1] and Woobeom Lee [2,*]

[1] AI Software Education Institute, Soonchunhyang University, Asan 31538, Republic of Korea; changingstart@gmail.com

[2] Department of Information Communication Software Engineering, Sangji University, Wonju 26339, Republic of Korea

* Correspondence: beomlee@sangji.ac.kr

Abstract: HAR technology uses computer and machine vision to analyze human activity and gestures by processing sensor data. The 3-axis acceleration and gyro sensor data are particularly effective in measuring human activity as they can calculate movement speed, direction, and angle. Our paper emphasizes the importance of developing a method to expand the recognition range of human activity due to the many types of activities and similar movements that can result in misrecognition. The proposed method uses 3-axis acceleration and gyro sensor data to visually define human activity patterns and improve recognition accuracy, particularly for similar activities. The method involves converting the sensor data into an image format, removing noise using time series features, generating visual patterns of waveforms, and standardizing geometric patterns. The resulting data (1D, 2D, and 3D) can simultaneously process each type by extracting pattern features using parallel convolution layers and performing classification by applying two fully connected layers in parallel to the merged data from the output data of three convolution layers. The proposed neural network model achieved 98.1% accuracy and recognized 18 types of activities, three times more than previous studies, with a shallower layer structure due to the enhanced input data features.

Keywords: human activity recognition (HAR); 3-axial sensor; image type encoding method; WISDM dataset; CNN

check for updates

Citation: Kim, C.; Lee, W. Human Activity Recognition by the Image Type Encoding Method of 3-Axial Sensor Data. *Appl. Sci.* **2023**, *13*, 4961. https://doi.org/10.3390/app13084961

Academic Editors: Luigi Bibbò and Marley M.B.R. Vellasco

Received: 10 March 2023
Revised: 7 April 2023
Accepted: 10 April 2023
Published: 14 April 2023

1. Introduction

Currently, smartphones are one of the essential items in daily life [1]. Smartphones integrate various sensors such as accelerometers, gyroscopes, light sensors, and temperature sensors, making them versatile for a wide range of services such as device control and monitoring. They are also used as wearable devices for analyzing physical activity [2–5]. For this analysis, data from 3-axis accelerometers and gyroscopes are commonly used, as they provide useful information on speed, direction, and angles of human movement. This data is crucial for human activity recognition (HAR), a technology that learns and infers advanced knowledge necessary for physical activity recognition based on raw sensor data. HAR can be effectively utilized in everyday life [6].

HAR is being pursued through various measurement methods and related services and research. Tian et al. [7] attempted HAR using a single-band wearable accelerometer and proposed an ensemble-based filter feature selection method that enhanced the strength of a single accelerometer and improved accuracy by removing overlap and unnecessary attributes. Kang et al. [8] proposed a hybrid deep learning model that uses both sensor data from accelerometers and skeleton data from images. Anguita et al. [9] collected sensor data by attaching smartphones to people's waists to differentiate various human activities and performed activity recognition using support vector machines. Sengul et al. [10] distinguished four common activities in daily life using accelerometer and gyroscope data to predict injuries caused by falls in the elderly. Moreover, many previous studies have focused on segmentation algorithms for accelerometer time series data [11], random undersampling, random oversampling, ensemble learning methods [12], and so on.

However, human physical activities can be divided into various types (walking, running, hiking, drinking water, sitting, etc.), and they also include similar activities (drinking water vs. eating, etc.) as well as types with clear differences (lying down vs. climbing stairs, etc.). Additionally, 3-axial sensor data can be prone to errors due to noise and uncertainty (sensor shaking, functional impairment, etc.), and the data size is smaller than that of video data, making it difficult to train. Therefore, various explored to obtain stable 3-axial sensor data, and there is considerable interest in visualization research for encoding sensor data into images without loss [13–16].

Therefore, this paper proposes a method to improve the accuracy of HAR by utilizing the 3-axial data (accelerometer and gyroscope sensor data) of a smartphone to visualize 2D and 3D. In addition, it recognizes 18 human physical activities through a single device (smartphone) instead of attaching multiple devices. Partial activity patterns of a single body movement were obtained through time series data grouped at regular intervals, and they were visualized in 2D and 3D image streaming formats. By clearly differentiating between similar human physical activities through this process, an improved HAR is proposed.

Section 2 of this study describes the body activity recognition technology using sensors. Section 3 introduces the proposed method of encoding the raw sensor data into an image form. Section 4 comparatively analyzes the performances of the previously studied neural network learning model and the proposed model. In the final section, we present our conclusions.

2. Related Research

Defining human actions as a single motion or external form is difficult because even if two motions may appear identical, they may result in different outcomes depending on subsequent movements. Therefore, time series data that captures the changes in data over time is used more frequently than a single data point for recognizing human actions [17,18]. Sensors are the most effective devices for gathering such data [12,19,20]. Currently, deep-learning-based models associated with sensor data can automatically extract and classify the characteristics of time series data, enabling accurate behavioral recognition.

In [21], a CNN with local loss was proposed for HAR. The experimental results showed that the local loss performed better than the global loss for the baseline architecture, and various human activities could be identified despite the low number of parameters. However, this study only showed high performance in recognizing six activities (walking, jogging, walking upstairs, walking downstairs, sitting, and standing) with 98.6% accuracy. The present study proposes a method to recognize 18 different types of actions, enabling more diverse biometrics.

A lightweight deep learning model for HAR was proposed in [22]. This model was developed using long short-term memory (LSTM) and recurrent neural network (RNN) and showed high performance, achieving an accuracy of 95.78% for recognizing 18 types of activities on the WISDM dataset. However, due to their recurrent structures, LSTM and RNN models require longer training and inference times compared to general CNN-based models. To address this issue, we utilized only convolutional layers (1D, 2D, and 3D convolutional layers) in a parallel structure, allowing us to analyze and observe a small dataset from various perspectives.

Ignatov et al. [23] studied an independent deep-learning-based approach for the classification of human actions. In addition to the simple statistical feature of preserving the global shape of time series data, they proposed a CNN model for extracting local characteristics. This study segmented the collected accelerometer sensor data into various sizes to determine the most effective segmentation size and evaluated the performance of each segmentation. In our study, we used the duration of the actions to set the size of the segmented data and performed activity classification using this configuration.

In [24], to capture various activities for HAR, mobile devices with built-in perceptual extraction networks were attached to users, and the data collected from these devices were used for the initial training. The trained weight values were transferred to the server through the communication network. The transferred data were compared with the trained weight values from other devices to determine the optimal weight value, and the final weight value was delivered to each device for re-training. The method proposed by [24] allows for the simultaneous collection and training of multiple activity data, and strong performance can be achieved by comprehensively determining the weights of individually trained models. However, comprehensive weight determination can emphasize strong performance, but it may also reduce accuracy when classifying similar activities for precise analysis. In consideration of this, we proposed a method to enhance the original sensor data, which enabled the classification of 18 distinct activities.

Previous studies on HAR have utilized various methods such as using RNN-based models to learn temporal changes or hybrid models that mix CNNs. Although HAR using CNN models has also been studied, it does not perform as well as RNN or hybrid models (refer to Chapter 6). However, RNN-based models can be limited in real-life usage due to long training and inference times, and hybrid models have complex structures that make it difficult to understand the learning process. Additionally, since human physical activity is diverse and there are many similar movements, there is a high possibility that features may be lost during the operation process of the layers in deep model structures, and it is difficult to wear many wearable devices due to discomfort. To address these issues, we propose a HAR method based on a wearable device using a single smartphone.

To effectively collect human physical activity from wearable devices, we expand (encode) high-dimensional 3-axis sensor data. This generates new features of human physical activity that could not be detected in one-dimensional data and removes fine noise from the sensor. In other words, by defining high-dimensional features such as directionality and spatiality in one-dimensional data, we propose new information about features of human physical activity. These new features enable the recognition of more diverse types of human physical activity and the discovery of unique features among similar types of human physical activity. Additionally, to effectively learn from the increased input data, we connect convolutional layers in parallel to enable parallel computation and complement the missing information in the encoding and learning processes using various dimensional data (1D data (3-axis sensor), 2D data (image), and 3D data (video image)). The encoding process is described in detail in Section 3.

3. Image Type Encoding Method of the 3-Axial Sensor Data

Accelerometer and gyroscope sensors that measure the velocity, momentum change, etc., of an object can detect the active state of an object, due to which both these devices are used extensively. The (x, y, and z) 3-axial data values from these sensors are arranged into a time series structure to recognize human activities using the properties of data changes according to time. However, in the case of similar human activities, the recognition accuracy decreases due to the small data dimension, which limits the expression of the characteristics. Therefore, the 3-axial raw data gathered through the accelerometer and gyroscope from this study were encoded into 2D and 3D images that express time properties. The image data were trained together with the 1D raw data to increase the precision and accuracy in order to perform high-dimensional HAR.

3.1. Three-Axial Acceleration and Gyroscope Data Analysis of the WISDM Dataset

The 3-axial accelerometer and gyroscope sensor data used in this study are from the "WISDM smartphone and smartwatch activity and biometrics" database published by Weiss [25]. This database consists of data gathered at 50 ms intervals for 18 daily activities from smartphones placed in the pockets of 51 subjects for three minutes. Table 1 summarizes the 18 measured activities, which are largely distinguished into basic activities related to walking (A), hand-based activities (B), and dining activities (C).

Table 1. Smartphone acceleration and gyroscope data from the WISDM database.

Label	Activity	No. of Columns		No. of Merged Columns	No. of Data	Grouping Type
		Accel	Gyro			
0	Walking	279,817	203,919	152,114	51	A
1	Jogging	268,409	200,252	154,020	49	A
2	Stairs	255,645	197,857	160,430	50	A
3	Sitting	264,592	202,370	180,315	51	A
4	Standing	269,604	202,351	165,068	51	A
5	Typing	246,356	194,540	166,646	49	B
6	Brushing teeth	269,609	202,622	168,771	51	B
7	Eating soup	270,756	202,408	164,177	51	C
8	Eating chips	261,360	197,905	160,237	50	C
9	Eating pasta	249,793	197,844	170,598	50	C
10	Drinking	285,190	202,395	149,138	51	C
11	Eating sandwich	265,781	197,915	164,635	51	C
12	Kicking	278,766	202,625	150,651	51	A
13	Catching	272,219	198,756	146,675	50	B
14	Dribbling	272,730	202,331	150,333	51	B
15	Writing	260,497	197,894	175,638	51	B
16	Clapping	268,065	202,330	165,304	51	B
17	Folding clothes	265,214	202,321	164,006	51	B

In Table 1, activity A is based on lower body movements, while most activities in B involve both lower and upper body movements, and C includes activities such as eating or drinking. Each activity's data includes a minimum of 194,540 raw data points or more, and the accelerometer and gyroscope data were merged based on the measurement time (Table 1, no. of merged columns). Since the WISDM database comprises similar activity groups and a small amount of data from 49 to 51 (number of subjects), in this study, we augmented the training dataset by segmenting the data into time units.

3.2. Walking-Activity-Based Data Argumentation

Among the 18 activities of the WISDM dataset, the "Walking" activity in the given time unit was the easiest to analyze. "Walking" is among the most common human activities, and a healthy person can normally walk 4.5 km/h, and approximately 8 km can be covered in 10,000 steps [26–30]. This shows that about 800 ms is required for a movement of 1 m. In addition, it can be inferred that about 6,400,000 ms (=1 h 46 m 40 s) is required for an 8 km walk, which amounts to 10,000 steps. The time required for one step, denoted as T_w, corresponds to about 640 ms of time. Therefore, this study sets the data segment size (DSS) as shown in Equation (1) for the raw sensor data of the WISDM generated at 50 ms intervals based on T_w, which equals one human step.

$$DSS = \frac{T_w}{T_R} + bias \tag{1}$$

where, T_w: one step time; T_R: sampling time of the WISDM dataset.

In Equation (1), $T_R = 50$ ms indicates the interval of data collection of the WISDM dataset, and $T_w = 640$ ms indicates the time consumed per step taken. The DSS was set to 15 with a bias value of 2.2. One input pattern for neural network training corresponds to 15 raw sensor data points, and the raw WISDM dataset segments the data repetitively by moving by one each. Ultimately, 910 physical activity data points were increased to 2,896,476 as a result of using the data segmentation method proposed in this study. These data were divided into training data and test data in a ratio of 8:2 (2,317,180 data points in the training set and 579,296 in the test set).

3.3. Brightness Intensity Distribution Pattern Transformation

Image type expansion was performed for the increased data obtained through raw sensor data segmentation. The accelerometer and gyroscope sensor data segmented into identical sizes can express 2D image patterns using the raw values that correspond to the amplitude of the continuous data, and these pattern data can be used to analyze physical activities.

Figure 1 shows an example of the raw accelerometer sensor data expressed as a brightness value. The raw time series data in Figure 1a are mapped to a brightness value and visualized according to that value. In the case of transforming each point A–E of the time series graph into a brightness value, the brightness intensity distribution pattern (BIDP) for each physical activity data can be obtained, as shown in Figure 2b.

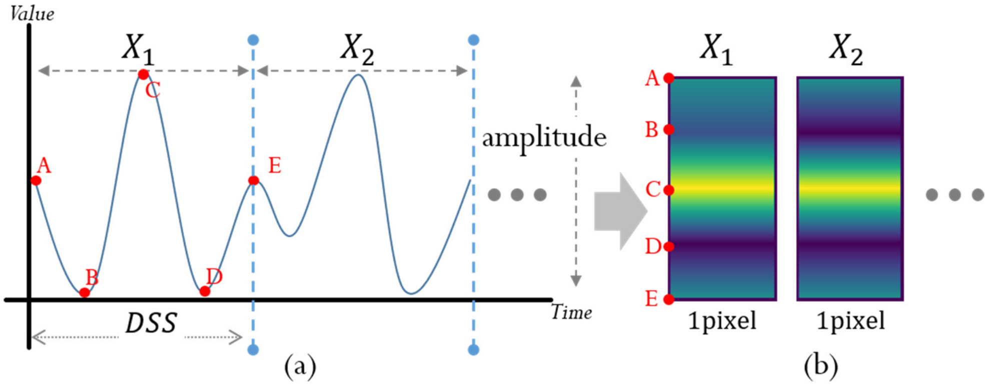
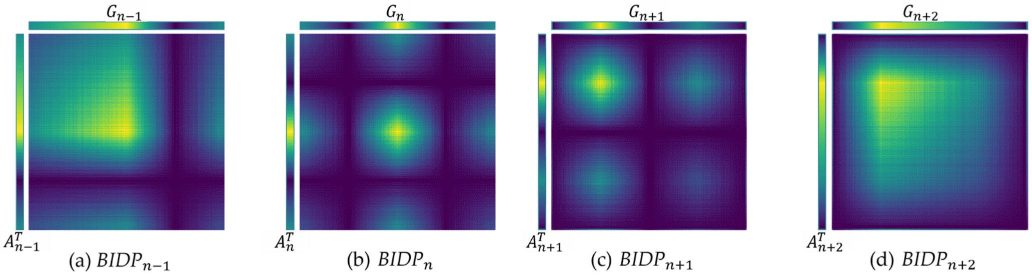


Figure 1. Brightness distribution transformation of raw sensor data: (**a**) raw data graph, and (**b**) brightness intensity distribution pattern ($BIDP$).



(a) $BIDP_{n-1}$ (b) $BIDP_n$ (c) $BIDP_{n+1}$ (d) $BIDP_{n+2}$

Figure 2. Example of $BIDP$ visualization.

Each point is expressed as a distinct brightness value according to the measured value. In the case of transformation into a 256-grayscale image, a brightness value of 128 is assigned to point A as it is located at the center between the maximum and minimum amplitudes. Point B, which has the minimum amplitude, is assigned a brightness value of 0, while point C, which has the maximum amplitude, is assigned a brightness value of 255. Points D and E are assigned brightness values of 0 and 128, respectively.

First, to represent the consistent pattern of physical activity in an image format, the BIDP was transformed into a $DSS \times DSS$ matrix by applying Equation (2) after expressing the raw data from the accelerometer and gyroscope sensors as brightness values in a 1 × DSS matrix.

$$BIDP = A^T G = \begin{bmatrix} a_1 \\ \vdots \\ a_{DSS} \end{bmatrix} \begin{bmatrix} g_1 & \ldots & g_{DSS} \end{bmatrix} = \begin{bmatrix} a_1 g_1 & \cdots & a_1 g_{DSS} \\ \vdots & \ddots & \vdots \\ a_{DSS} g_1 & \cdots & a_{DSS} g_{DSS} \end{bmatrix} \quad (2)$$

where, $A = \begin{bmatrix} a_1 & a_2 & a_3 & \ldots & a_{DSS} \end{bmatrix}$, $G = \begin{bmatrix} g_1 & g_2 & g_3 & \ldots & g_{DSS} \end{bmatrix}$.

In Equation (2), A and G represent the $1 \times DSS$ size $BDIP$ matrices of the accelerometer and gyroscope sensors that correspond to one DSS, respectively. They are transformed into images of $DSS \times DSS$ size by taking the dot product with the transposed matrix of matrix A, denoted as A^T.

Figure 2 shows the results of the $BIDP$ dimension expansion over time. The spatial characteristics of physical activities can be obtained by discerning brightness intensity within patterns, which can be observed based on the raw sensor values. Figure 2a shows a strong area of brightness distributed at the beginning of the $BIDP$, and the resulting image is characterized by an emphasized space at the upper left corner. Figure 2b shows a dot pattern with a strong brightness area distributed between light brightness intensities, which is emphasized at the center. Figure 2c also shows a dot pattern, but the strong brightness area is emphasized at the upper left corner instead of the center. Figure 2d is similar to Figure 2a, but the location of the brightness area differs. In short, distinct spatial characteristics can be obtained depending on the location of the strong brightness intensities, which can be used to emphasize the properties of the sensor data.

However, Figure 2 shows experimental results that did not consider the 3-axial nature of the raw data. The raw accelerometer data, which comprises three axes, does not exhibit a standardized form as shown in Figure 3. Therefore, this study serializes the 3-axial sensor data to apply Equation (2) above and express the spatial characteristics of physical activities in a more accurate form.

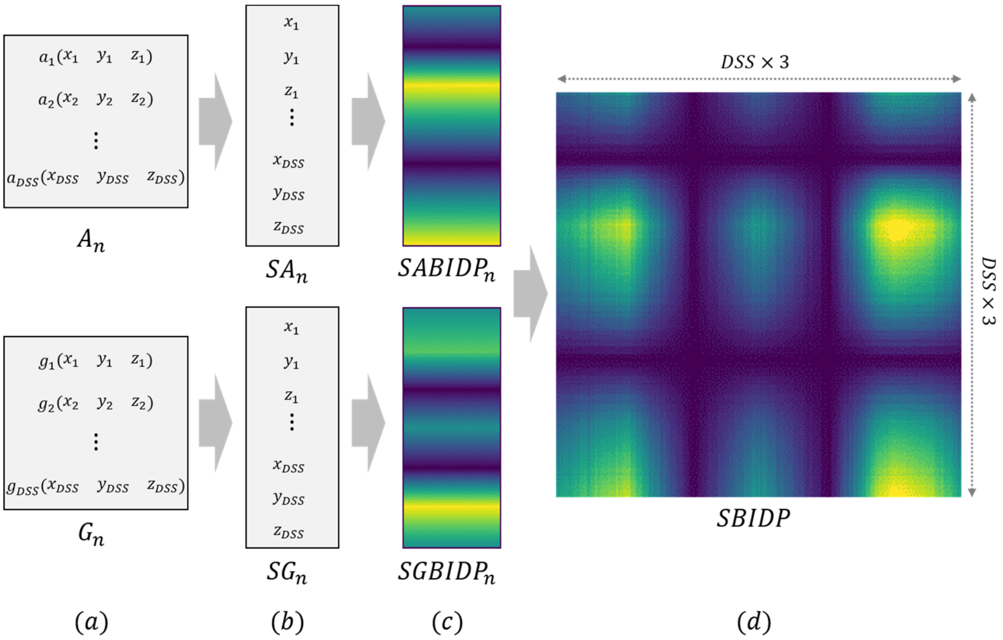


Figure 3. Example of BIDP visualization by 3-axial raw data serialization: (**a**) A$_n$: acceleration dataset; G$_n$: gyro dataset; (**b**) SA$_n$: serialized acceleration data; SG$_n$: serialized gyro data; (**c**) $SABIDP_n$: $BIDP$ of serialized acceleration data; $SGBDP_n$: $BIDP$ of serialized gyro data; (**d**) $SBIDP_n$: serialized $BIDP$.

Serializing the 3-axial sensor data, as shown in Figure 3, differentiates the sensor values for each axis and expresses a more complex geometric spatial pattern. A_n and G_n in Figure 3 represent the 3-axial dataset of the accelerometer and gyroscope sensors with DSS size, respectively, while SA_n and SG_n represent each component of the 3-axial dataset serialized into linear form. Applying Equation (2) generates a $SBIDP_n$ of size $(DSS \times 3) \times (DSS \times 3)$. The generated $SBIDP_n$ exhibits greater geometric spatial patterns than Figure 2, which uses 1-axial data. This is clearly evident in the dimensional expansion using actual 3-axial data. The generated $SBIDP_n$ exhibits greater geometric spatial patterns than Figure 2, which uses 1-axial data. This is clearly evident in the dimensional expansion using actual 3-axial data.

Figure 4 shows an example of $SBIDP$ for the 18 types of physical activities presented by the WISDM dataset using the actual 3-axial raw accelerometer and gyroscope data. All physical activities show dot patterns while some linear patterns can be seen in the inner part due to the effect of the color space caused by the different ranges of brightness. The line patterns inside the image represent information expressed from the different strength values of each axis, which can be recognized as the spatial characteristics of the physical activities. These characteristics are emphasized to a greater extent depending on the magnitude of the differences in the strength values.

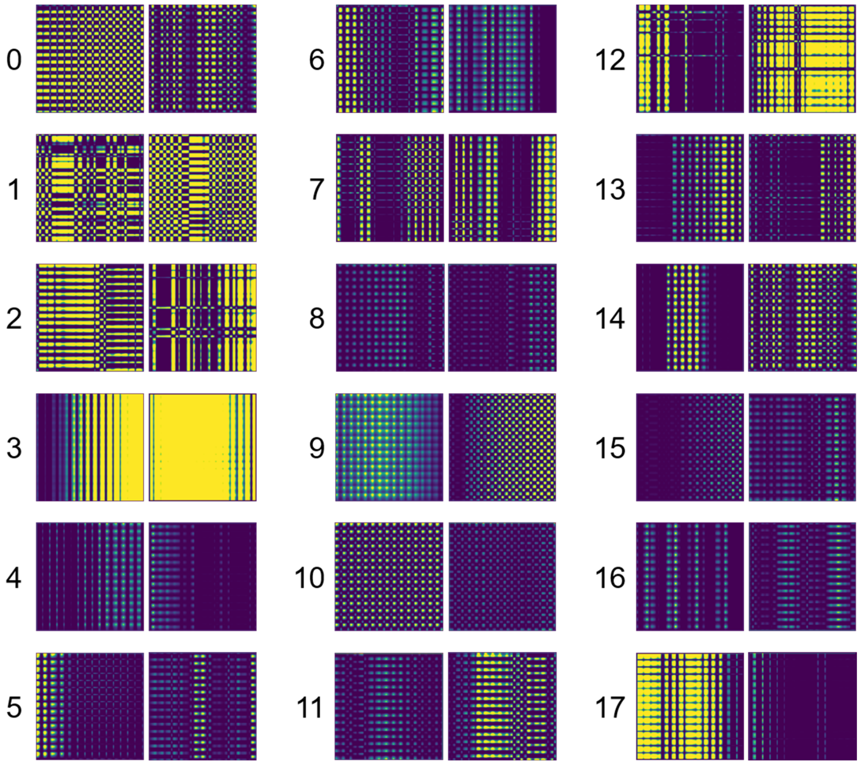


Figure 4. $SBIDP$ example of 18 activities in the WISDM dataset.

Figure 5 shows the magnified $SBIDP$ results for physical activity labels 3, 7, and 17 from Figure 4. While all patterns may appear rectangular or magnified, different patterns are expressed based on brightness. Therefore, these patterns are used as classification features for physical activities.

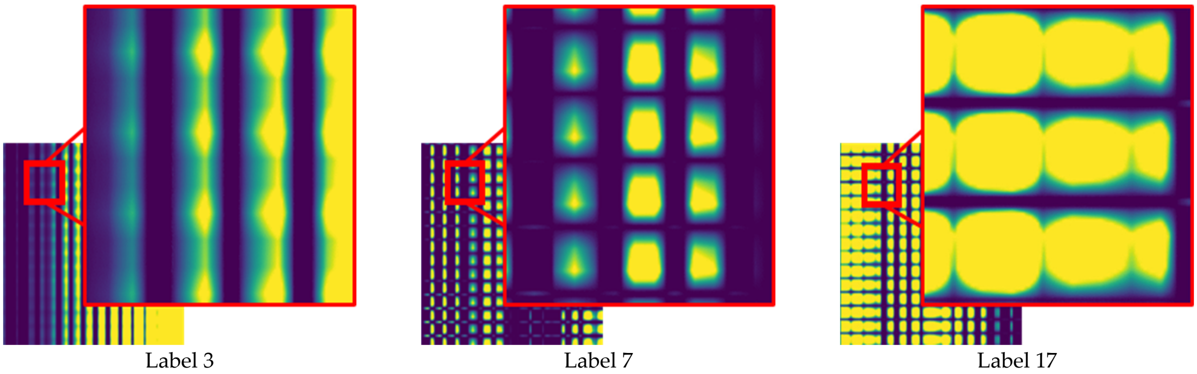


Figure 5. Example of the magnified $SBIDP$ of some samples in Figure 4.

3.4. 2 Step SBIDP Enhancement Method

The $SBIDP$ generated through image encoding of raw sensor data expresses physical activity characteristics as spatially diverse brightness, patterns, and shapes, as shown in Figure 6. Figure 6 shows the change in the continuous $BIDP$ images for three physical activity data points according to the change in T. Labels 3, 7, and 17 have overall strong brightness and take the form of vertical grid patterns, but their detailed characteristics differ, as seen in their magnifications in Figure 5. In addition, while the physical activity performed in label 3 of Figure 6 remains the same, the vertical pattern gradually becomes stronger over time. However, the detailed pattern of label 3 in Figure 5 does not change. Similar results were obtained for physical activity in labels 7 and 17. They exhibited stronger grid patterns compared to label 3, as illustrated in Figure 5, which shows the varying levels of brightness in different areas upon magnification. However, utilizing these robust grid pattern features without any modifications as input for training a neural network model may negatively impact its ability to accurately recognize physical activities.

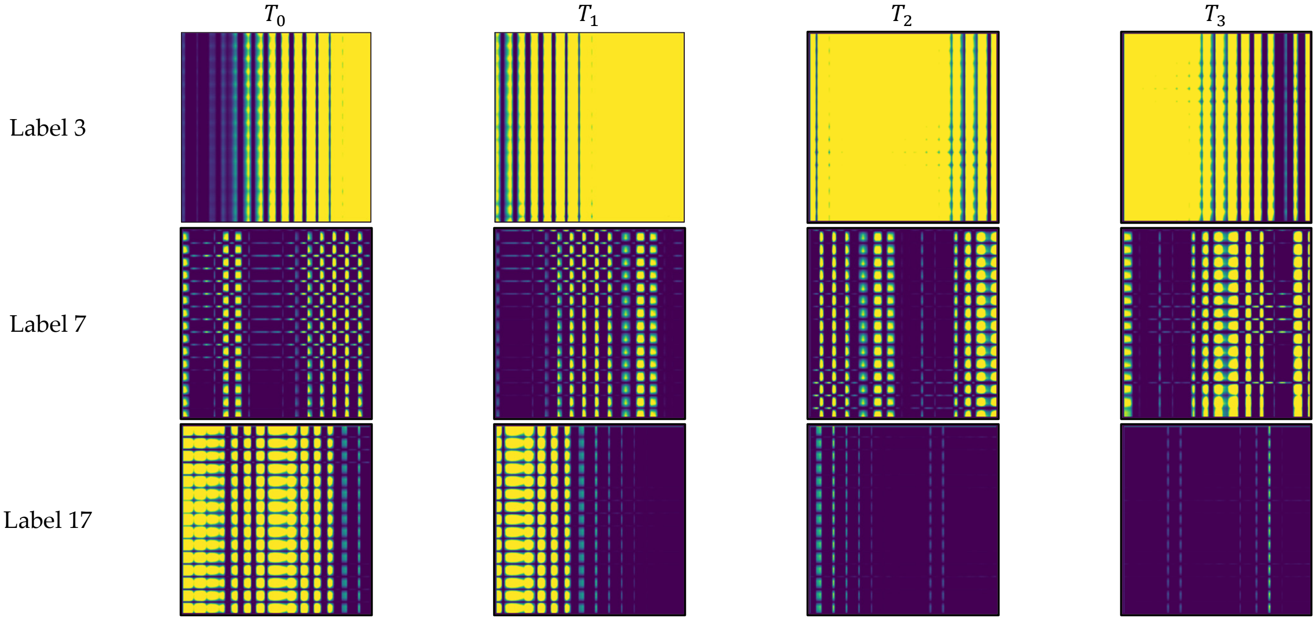


Figure 6. Examples of sequential $SBIDP$.

Therefore, in this study, to transform the detailed characteristics of the grid pattern into one large pattern, component values of the raw accelerometer sensor value matrix $A[\cdot]$ and of the gyroscopic sensor value matrix $G[\cdot]$ were arranged to generate $SBIDP$. This, as a primary pre-processing step for neural network input, generates $BIDP_{E1}$ with strengthened spatial characteristic information. Figure 7 is an example of $BIDP_{E1}$ generated using the arranged sensor data matrix, and as can be seen, there are clearer and more defined gradation spatial characteristics compared with the grid pattern of each label in Figure 6. However, due to varying brightness values, the resulting pattern took on a curved shape. The angular features of this curve were utilized to represent changes in physical activity data, and Equation (3) was employed to generate $BIDP_{E2}$ with further improved spatial characteristics as a secondary step.

$$BIDP_{E2} = \sum_{i=0}^{3} BIDP_{E1}\left(\frac{\pi}{2} \times i\right) \tag{3}$$

where $BDIP_{E1}(\theta)\ :\ \theta\ rotated\ BIDP_{E1}$.

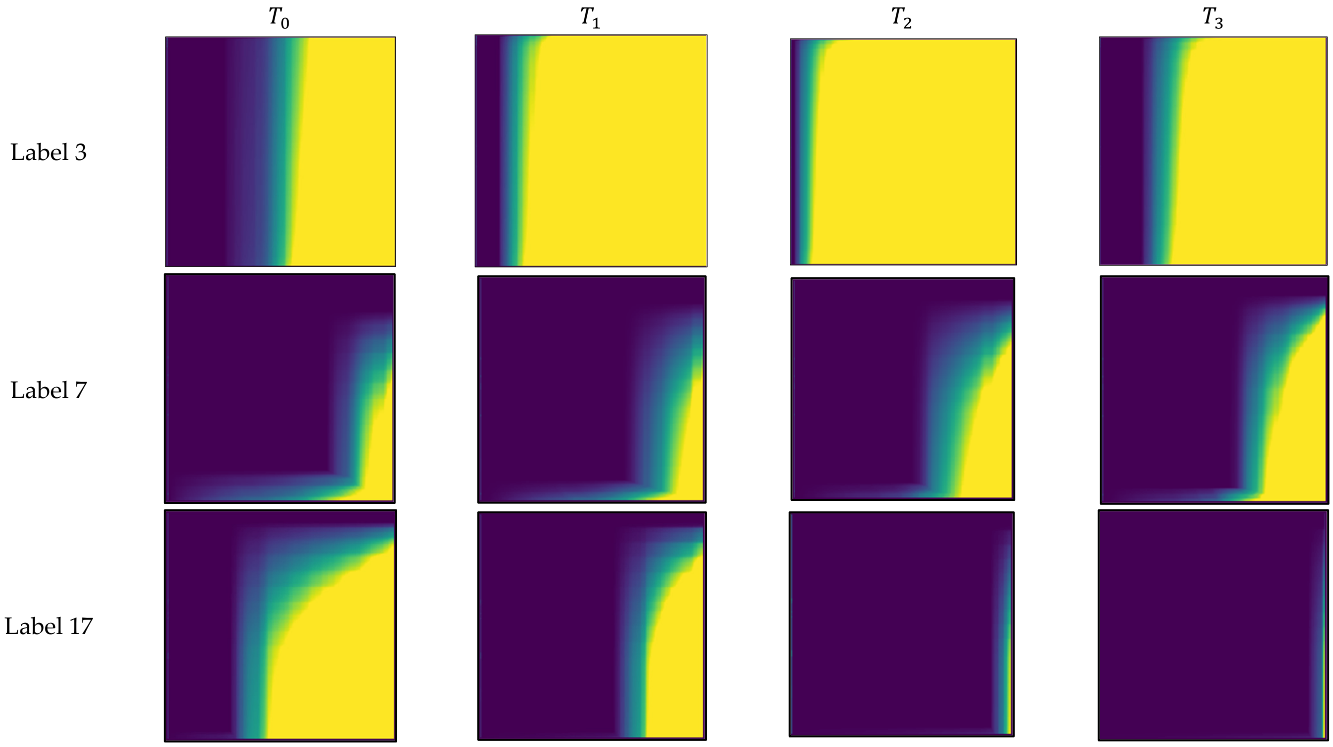


Figure 7. $BIDP_{E1}$ examples of 1st enhanced $SBIDP$ by sorting elements of the $A[\cdot]$ and $G[\cdot]$ matrices.

$\sum$ in Equation (3) denotes the image sum (OR) operation, and it refers to the image OR of the arranged $BIDP_{E1}$ rotated by 90°, 180°, and 280°. Figure 8 illustrates the outcomes of $BIDP_{E2}$ after the secondary enhancement of spatial characteristics, wherein label 3 is represented as an angled propeller and label 17 as a curved propeller. Figure 9 displays the resulting $BIDP_{E2}$ for all 18 physical activities in the WISDM dataset.

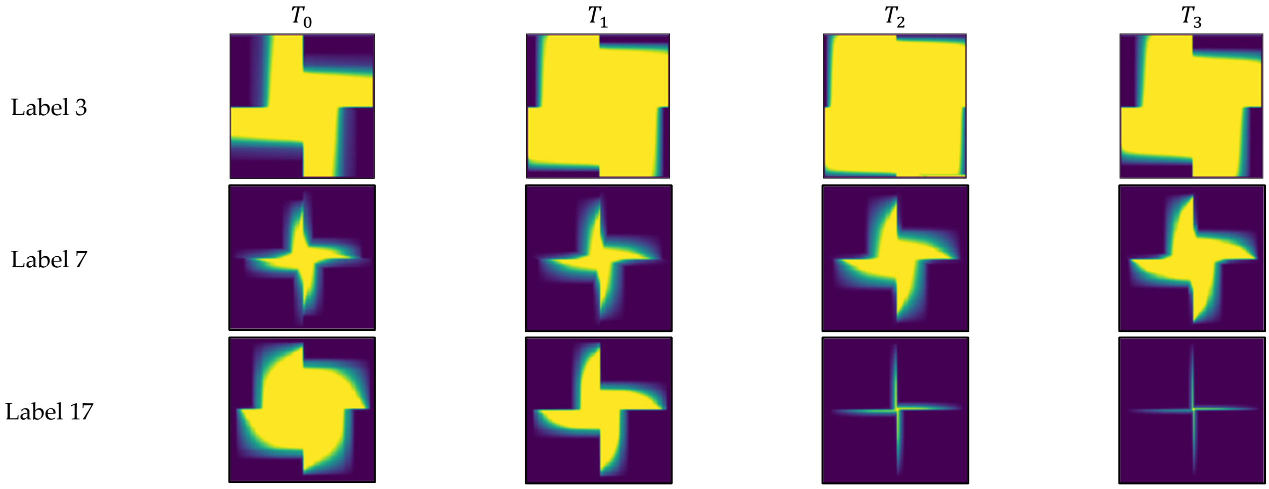


Figure 8. $BIDP_{E2}$ examples of 2nd enhanced $BIDP_{E1}$.

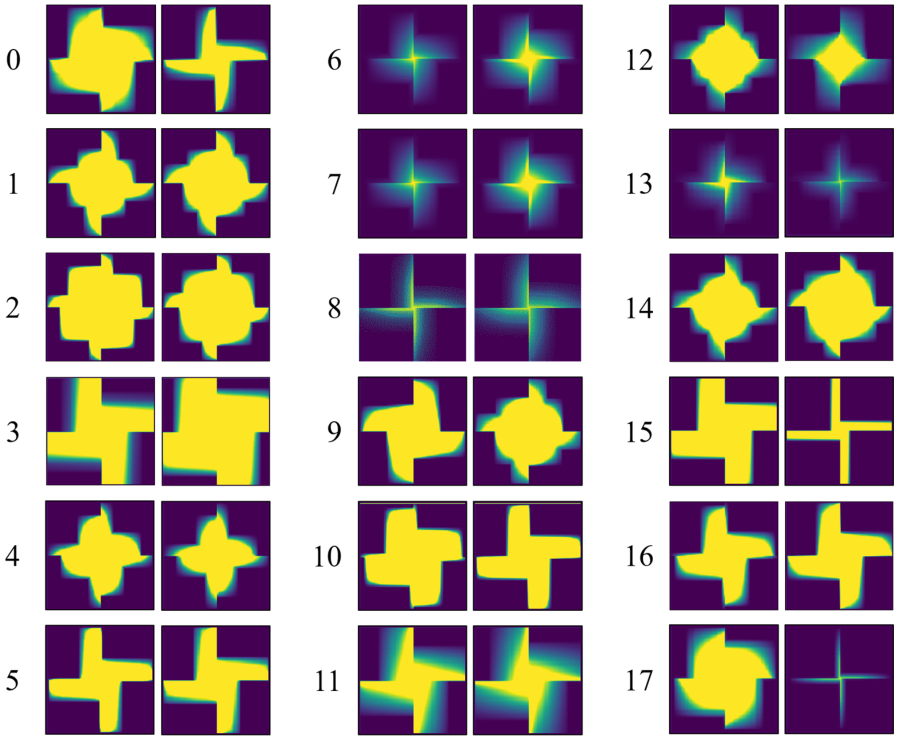


Figure 9. $BIDP_{E2}$ example of 18 activities in the WISDM dataset.

4. Three-Dimensional Visualization Method of BIDP

The $BIDP_{E2}$ produced by the secondary process of enhancing spatial characteristics is transformed into an image with various shapes based on the finely expressed brightness value. To express this characteristic in detail, this section visualizes this image into a 3D image with depth information, as shown in Figure 10. In general, raw sensor data as time series data contain the recognition of the features of the physical activity of humans according to time. In this study, to spatially express the time series feature of these raw data, one physical activity record segmented as DSS was divided into three equal parts for encoding into the form of a 3D image, as shown in Figure 10.

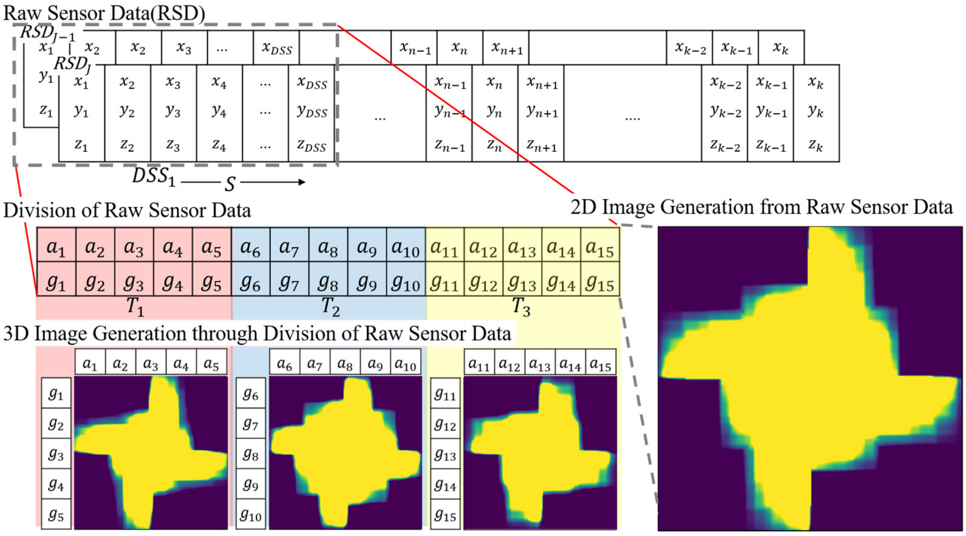

Raw Sensor Data(RSD)
RSD_{J-1}
RSD_J
x_1 x_2 x_3 x_4 ... x_{DSS}
y_1 y_2 y_3 y_4 ... y_{DSS}
z_1 z_2 z_3 z_4 ... z_{DSS}
x_{n-1} x_n x_{n+1}
y_{n-1} y_n y_{n+1}
z_{n-1} z_n z_{n+1}
x_{k-2} x_{k-1} x_k
y_{k-2} y_{k-1} y_k
z_{k-2} z_{k-1} z_k
DSS_1
S
Division of Raw Sensor Data
a_1 a_2 a_3 a_4 a_5 a_6 a_7 a_8 a_9 a_{10} a_{11} a_{12} a_{13} a_{14} a_{15}
g_1 g_2 g_3 g_4 g_5 g_6 g_7 g_8 g_9 g_{10} g_{11} g_{12} g_{13} g_{14} g_{15}
T_1
T_2
T_3
2D Image Generation from Raw Sensor Data
3D Image Generation through Division of Raw Sensor Data

Figure 10. 3D Visualization processing concept from $BIDP$ of raw sensor data (J is no. of datasets).

Figure 11 presents an example of $BIDP_{3D}$ for the 18 physical activities of the WISDM dataset, which were generated using the processing steps shown in Figure 10. $BIDP$ refers to the 2D image, and $BIDP_{3D}(t_1)$, $BIDP_{3D}(t_2)$, and $BIDP_{3D}(t_3)$ each represent one of the three even parts of a segmented physical activity, as a set of continuous 2D images, showing 3-channel spatial characteristics with time properties.

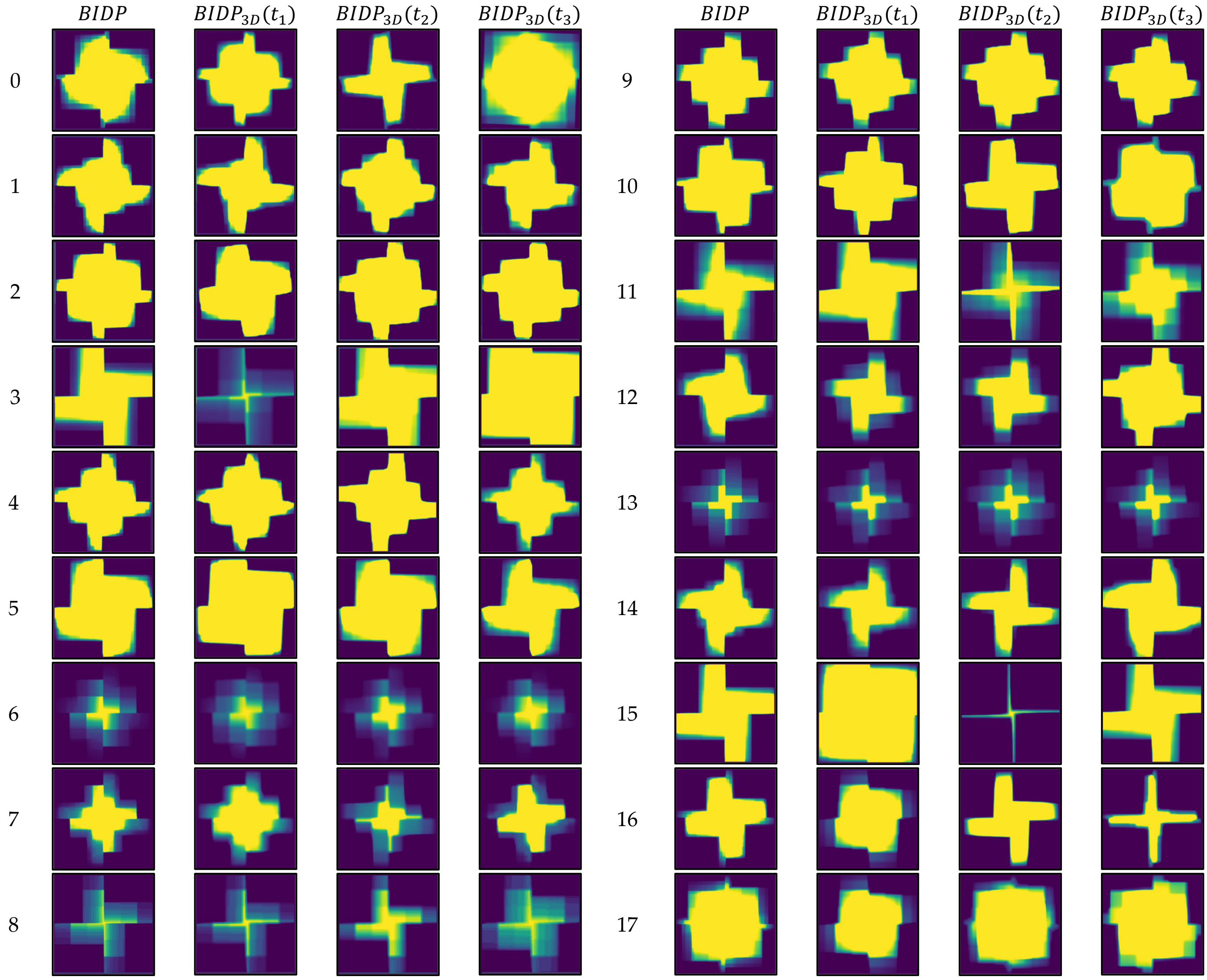


Figure 11. $BIDP_{3D}$ examples of 18 activities in the WISDM Dataset.

5. Proposed CNN Architecture for Learning Activity Data

For the simultaneous training of 1D raw sensor data (RSD), 2D $BIDP$, and 3D $BIDP_{3D}$ data, 1D, 2D, and 3D convolutional layers are used. The 1D convolutional layer convolves the sequence data and is well-suited for training long sequences, such as text. The 2D convolutional layer can extract the feature map for the spatial and directional information of image data, while the 3D convolutional layer extracts the feature map for the spatial and directional changes over time. Figure 12 shows the CNN model structure for training 1D RSD and the expanded 2D $BIDP$ and 3D $BIDP_{3D}$ data.

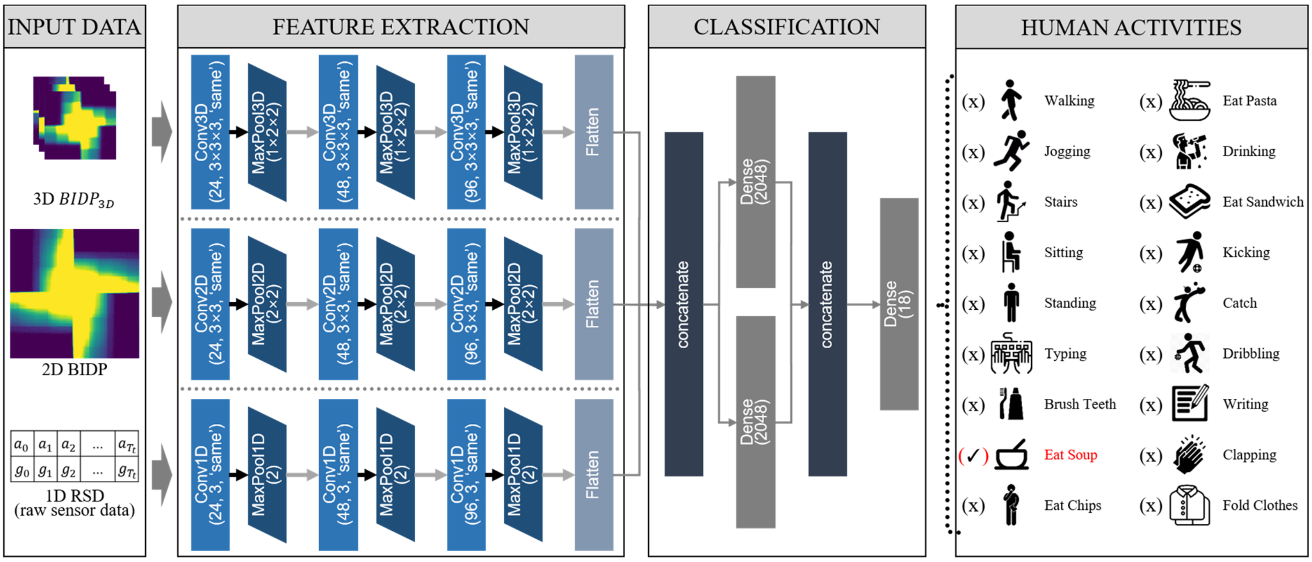


Figure 12. A Proposed multi-dimensional convolutional neural network.

As shown in Figure 12, the feature extraction part consists of three convolutional layers, and the max-pooling layer (1D, 2D, and 3D) is used for subsampling. The number of kernels in each convolutional layer is 24, 48, and 96, and the filter size is (3), (3 × 3), and (3 × 3 × 3). All layers have "same" padding, and "ReLU" is used as the activation function. The max-pooling layer was set to (2), (2 × 2), and (1 × 2 × 2) to reduce the feature map size by 50%, and the resulting feature map was flattened into 1D. Classification using two dense layers was performed in parallel, after merging the feature maps extracted through the convolutional layer of each dimension. Each dense layer has 2,048 nodes, and "ReLU" is used as the activation function. The results from the dense layers were merged again using the concatenate function and used as the input to the output layer.

The model parameters mentioned in this study were set using the "keras_tuner" of the open-source Keras library. The system used for the experiments was a Windows 10 64-bit environment with an i7-6700 CPU, 48 GB of RAM, and two NVIDIA GeForce RTX 3060 GPUs with 12 GB of memory each.

6. Performance Evaluation

6.1. Training Result

The training results for the images generated using RSD and the original data with the learning model shown in Figure 12 demonstrate identical accuracy and loss, as shown in Table 2 and Figure 13. The accuracy of the training data in Table 2 was 99.6%, and the loss was approximately 0.0134. The model completed training at 90 epochs because there was no significant difference in loss after the 73rd epoch, as shown in Figure 13.

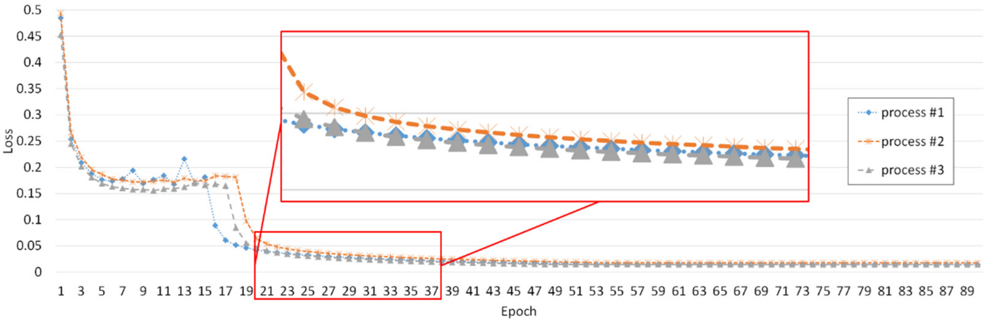


Figure 13. Loss of training data.

Table 2. Accuracy and loss from training model.

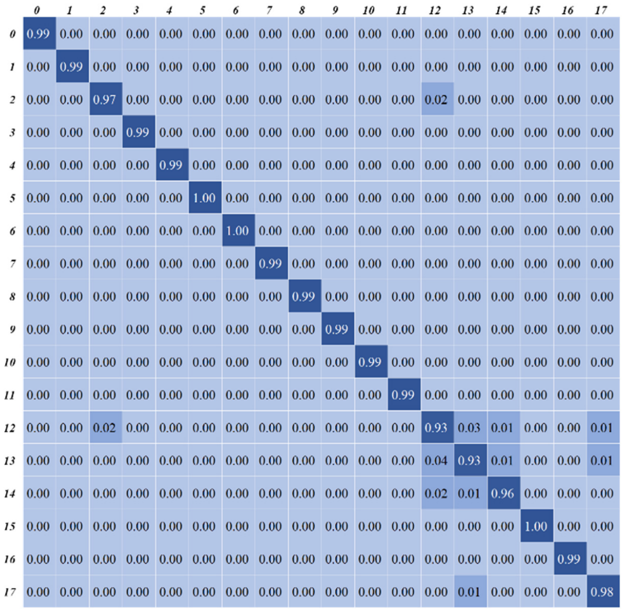
	Our Model
Accuracy	99.52%
Loss	0.0134

6.2. Performance Evaluation Result

Table 3 shows the precision, recall, and F1 scores of the proposed model for each class using validation data. The model achieved a performance of 90% or above for all 18 physical activity classes, but classes 12, 13, 14, and 17 showed low performance. This was due to the fact that physical activities such as climbing stairs, kicking, catching, dribbling, and folding clothes, which are expressed as slight left and right or up and down movements, and have similar activity data, were included in classes 2, 12, 13, 14, and 17, as presented in the confusion matrix in Figure 14.

Table 3. Performance evaluation of expanded data.

Class	**Precision**	**Recall**	**F1 Score**
0	99.0%	99.0%	99.0%
1	99.0%	99.0%	99.0%
2	97.0%	97.0%	97.0%
3	99.0%	99.0%	99.0%
4	99.0%	99.0%	99.0%
5	100.0%	100.0%	100.0%
6	100.0%	100.0%	100.0%
7	99.0%	99.0%	99.0%
8	99.0%	99.0%	99.0%
9	99.0%	99.0%	99.0%
10	99.0%	99.0%	99.0%
11	99.0%	99.0%	99.0%
12	91.0%	93.0%	92.0%
13	94.0%	93.0%	94.0%
14	96.0%	96.0%	96.0%
15	100.0%	100.0%	100.0%
16	99.0%	99.0%	99.0%
17	97.0%	98.0%	97.0%



	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
0	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	0.00	0.97	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00
11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.00	0.00	0.00	0.00	0.00	0.00
12	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.93	0.03	0.01	0.00	0.00	0.01
13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.93	0.01	0.00	0.00	0.01
14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.01	0.96	0.00	0.00	0.00
15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.99	0.00
17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.98

Figure 14. Confusion matrix of validation data (%).

6.3. Performance Evaluation Comparison by Using the WISDM Dataset

This section compares the performance of the proposed model and that of a well-known neural network model with those of previous studies. Table 4 compares the HAR performance with that of the previous RNN-based model using the WISDM dataset. The proposed method shows a value of 98.15%, which is higher than the corresponding results of previous studies. However, the proposed method showed a slightly lower performance compared with the structure that serially connected numerous models (CNN-GRU-LSTM); however, the proposed algorithm has a relatively simple and shallow layer structure as a parallel convolutional layer, as shown in Figure 12.

Table 4. Evaluation of the proposed model compared with models based on RNN.

Ref.	Model	F1 Score (%)	Accuracy (%)
[31]	Tri-PSRNN	96.62	94.76
[31]	PSDRNN	94.01	93.06
[32]	LSTM-CNN	-	95.85
[33]	LSTM-RNN	95.40	96.40
[34]	Single-input CNN-GRU model A	92.42	92.03
[34]	Single-input CNN-GRU model B	94.50	94.71
[34]	Single-input CNN-GRU model C	92.55	92.37
[34]	Multi-input CNN-LSTM	95.55	95.45
[34]	Multi-input CNN-GRU	97.22	97.21
[35]	CNN-GRU-LSTM	98.52	98.51
-	Proposed model	98.00	98.15

The use of RNN-based models for HAR can lead to performance degradation due to the issues of exploding and vanishing gradients in back-propagation. Although LSTM and GRU techniques have been introduced to address these issues, the sequential nature of vector inputs allows for the processing of only one sequential data at a time, making it difficult to take advantage of the parallel processing capabilities of GPUs. As a result, training and inference models may experience somewhat slower speeds. However, the algorithm proposed in this paper uses CNN-based methods to overcome these shortcomings. With a relatively simple image encoding method, it can perform HAR with dimensional concepts (such as space and direction) in the CNN model, allowing for the extraction of features that were not previously detectable in time series data.

Table 5 compares the proposed method with CNN-based models, including CNN models that use input data that have been expanded into multidimensional data. The proposed method achieved higher performance than previous CNN-based models. In addition, the HAR data were composed in the form of a time series. Thus, the RNN model that used the data change according to time showed a higher performance than the CNN-based models. However, the method proposed in this study uses only a convolutional layer and shows results similar to those of the RNN-based models. This implies that 18 physical activities can be classified even with a relatively simpler eight-layer model.

When examining the structure of the comparison models in Table 5, the large-scale models (Inception-V3 with 313 layers, EfficientNet B0 with 233 layers, and Xception with 126 layers) showed an accuracy of 90.27%, while the small-scale models (Multichannel CNN-GRU with 9 layers, CNN with an attention mechanism with 6 layers, CNN with 6 layers) showed a higher accuracy of 95.38% compared to the large-scale models. We attribute this performance to the loss of feature points between classes due to deep-layer operations on the input data. When visualizing time series data in a typical way, such as generating waveform-based visual data such as graphs or histograms, the feature information that can be obtained from the waveform information is limited, and all features will eventually be integrated unless there are clear feature points. This is because the entire waveform can contain similar features. To prove this, we designed a shallow-layer neural model and chose a parallel input structure and method of expanding the dimension of input data to mimic deep feature information even in shallow layers. Through this, we were able to recognize many categories of classes with a shallow structure compared to the comparison model.

Table 5. Evaluation of the proposed model in comparison with models based on CNN.

Ref.	Model	No. of Activities	Layer	F1 Score (%)	Accuracy (%)
[36]	Baseline	6	10	-	89.55
[37]	VGG16	6	23	-	89.32
[38]	Inception-V3	6	313	-	91.54
[39]	Xception	6	126	-	90.17
[40]	EfficientNet B0	6	233	-	89.11
[23]	CNN	6	6	-	93.32
[41]	Multichannel CNN-GRU	6	9	96.39	96.41
[42]	U-Net	6	11	96.50	96.40
[43]	CNN with an attention mechanism	6	6	-	96.40
-	Proposed Model	18	8	98.00	98.15

7. Conclusions

This paper proposes an image encoding method using 3-axial sensor data of acceleration and gyro and a human activity recognition (HAR) model based on it. By visualizing the raw sensor data from the WISDM dataset, strong visual features of the data waveform could be extracted, which improved recognition accuracy and categories. To augment the 1D raw sensor data, we divided it into time intervals calculated based on the "walking" activity, which is one of the fundamental human activities, and normalized the representation range of the segmented 1D sensor data to values between 0 and 255. This enabled clustering of the finely represented sensor data into a larger range, making it possible to remove noise caused by fine changes, such as shaking. The data with the modified representation range creates a 2D image through the matrix dot product of the acceleration and gyro data, and this image includes areas of strong brightness and weak brightness depending on the position of the data waveform. However, this can show overly geometric patterns, which can actually degrade the performance of the model. Therefore, a second processing step is used to generate a standardized visual image.

The standardized visual image shows a propeller shape with different curves and brightness areas of the wings depending on the sensor data waveform, creating visual feature differences in similar types of human activities. Moreover, due to the clear input data, the hierarchical structure of the HAR model could be simplified to a relatively shallow eight layers compared to previous studies. In addition, it was possible to recognize 18 categories of human activity, which is three times higher than in previous HAR studies, and achieve a high accuracy of 98.15%.

Our proposed algorithm is a method for detecting various types of human body activities on a single device. Through this, we were able to recognize 18 categories of body activities. In future research, additional experiments are needed to recognize more types of body activities, and comparison and analysis with previous studies that use dimension expansion concepts such as image encoding will be necessary. Additionally, analysis of the correlation between increased computational load due to data expansion and changes in encoding images based on data waveforms will be needed.

If we design a self-big-data-measurement device for detecting human body activities and collecting the measured data, we can expect its usefulness in the development of customized healthcare services based on lifelogging.

Author Contributions: Writing—original draft, C.K.; Writing—review & editing, W.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Sangji University Research Fund, 2020. This research was supported by "Regional Innovation Strategy (RIS)" through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (MOE) (2022RIS-005).

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

References

1. Keusch, F.; Wenz, A.; Conrad, F. Do You Have Your Smartphone with You? Behavioral Barriers for Measuring Everyday Activities with Smartphone Sensors. *Comput. Hum. Behav.* **2022**, *127*, 107054. [CrossRef]
2. Yang, P.; Yang, C.; Lanfranchi, V.; Ciravegna, F. Activity Graph based Convolutional Neural Network for Physical Activity Recognition using Acceleration and Gyroscope Data. *IEEE Trans. Ind. Inform.* **2022**, *18*, 6619–6630. [CrossRef]
3. Alrazzak, U.; Alhalabi, B. A survey on human activity recognition using accelerometer sensor. In Proceedings of the Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR), Spokane, WA, USA, 30 May–2 June 2019; pp. 152–159.
4. Huang, J.; Kaewunruen, S.; Ning, J. AI-based quantification of fitness activities using smartphones. *Sustainability* **2022**, *14*, 1–19. [CrossRef]
5. Ehatisham-ul-Haq, M.; Murtaza, F.; Azam, M.A.; Amin, Y. Daily Living Activity Recognition In-The-Wild: Modeling and Inferring Activity-Aware Human Contexts. *Electronics* **2022**, *11*, 1–24. [CrossRef]
6. Wang, J.; Chen, Y.; Hao, S.; Peng, X.; Lisha, H. Deep Learning for Sensor-based Activity Recognition: A Survey. *Pattern Recognit. Lett.* **2017**, *119*, 3–11. [CrossRef]
7. Tian, Y.; Zhang, J.; Wang, J.; Geng, Y.; Wang, X. Robust human activity recognition using single accelerometer via wavelet energy spectrum features and ensemble feature selection. *Syst. Sci. Control. Eng.* **2020**, *8*, 83–96. [CrossRef]
8. Kang, J.; Shin, J.; Shin, J.; Lee, D.; Choi, A. Robust Human Activity Recognition by Integrating Image and Accelerometer Sensor Data Using Deep Fusion Network. *Sensors* **2022**, *22*, 174. [CrossRef]
9. Anguita, D.; Ghio, A.; Oneto, L.; Parra, X.; Reyes-Ortiz, J.L. Human Activity Recognition on Smartphones Using a Multiclass Hardware-Friendly Support Vector Machine. *Adv. Nonlinear Speech Process.* **2012**, *7657*, 216–223.
10. Sengül, G.; Karakaya, M.; Misra, S.; Abayomi-Alli, O.O.; Damaševi˘cius, R. Deep learning based fall detection using smartwatches for healthcare applications. *Biomed. Signal Process. Control* **2022**, *71*, 103242. [CrossRef]
11. Ignatov, A.D.; Strijov, V.V. Human activity recognition using quasiperiodic time series collected from a single tri-axial accelerometer. *Multimed. Tools Appl.* **2016**, *75*, 7257–7270. [CrossRef]
12. Gupta, A.; Semwal, V.B. Multiple task human gait analysis and identification: Ensemble learning approach. In *Emotion and Information Processing*; A Practical Approach; Springer: Cham, Switzerland, 2020; pp. 185–197. [CrossRef]
13. Barra, S.; Carta, S.M.; Corriga, A.; Podda, A.S.; Reforgiato Recupero, D. Deep Learning and Time Series-to-Image Encoding for Financial Forecasting. *IEEE/CAA J. Autom. Sin.* **2020**, *7*, 683. [CrossRef]
14. Ahmad, Z.; Khan, N. Inertial Sensor Data to Image Encoding for Human Action Recognition. *IEEE Sens. J.* **2021**, *9*, 10978–10988. [CrossRef]
15. Wang, D.; Wang, T.; Florescu, I. Is Image Encoding Beneficial for Deep Learning in Finance? *IEEE Internet Things J.* **2020**, *9*, 5617–5628. [CrossRef]
16. Estebsari, A.; Rajabi, R. Single residential load forecasting using deep learning and image encoding techniques. *Electronics* **2020**, *9*, 68. [CrossRef]
17. Bulling, A.; Blanke, U.; Schiele, B. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Comput. Surv.* **2014**, *46*, 1–33. [CrossRef]
18. Sadouk, L. CNN approaches for time series classification. In *Time Series Analysis-Data, Methods, and Applications*; IntechOpen: London, UK, 2019; pp. 1–23.
19. Vishwakarma, D.K.; Dhiman, C. A unified model for human activity recognition using spatial distribution of gradients and difference of Gaussian kernel. *Vis. Comput.* **2019**, *35*, 1595–1613. [CrossRef]
20. Semwal, V.B.; Nandi, G.C. Generation of joint trajectories using hybrid automate-based model: A rocking block-based approach. *IEEE Sens. J.* **2016**, *16*, 5805–5816. [CrossRef]
21. Teng, Q.; Wang, K.; Zhang, L.; He, J. The layer-wise training convolutional neural networks using local loss for sensor based human activity recognition. *IEEE Sens. J.* **2020**, *20*, 7265–7274. [CrossRef]
22. Agarwal, P.; Alam, M. A Lightweight Deep Learning Model for Human Activity Recognition on Edge Devices. *arXiv* **2019**, arXiv:1909.12917. Available online: https://arxiv.org/abs/1909.12917 (accessed on 8 July 2020).
23. Ignatov, A. Real-time human activity recognition from accelerometer data using Convolutional Neural Networks. *Appl. Soft Comput.* **2018**, *62*, 915–922. [CrossRef]
24. Xiao, Z.; Xu, X.; Xing, H.; Song, F.; Wang, X.; Zhao, B. A federated learning system with enhanced feature extraction for human activity recognition. *Knowl. Based Syst.* **2021**, *229*, 107338. [CrossRef]
25. Weiss, G.M. Wisdm smartphone and smartwatch activity and biometrics dataset. In *UCI Machine Learning Repository: WISDM Smartphone and Smartwatch Activity and Biometrics Dataset Data Set*; 2019; Available online: https://archive.ics.uci.edu/ml/machine-learning-databases/00507/WISDM-dataset-description.pdf (accessed on 8 July 2021).

26. Chen, L.J.; Stubbs, B.; Chien, I.C.; Lan, T.H.; Chung, M.S.; Lee, H.L.; Ku, P.W. Associations between daily steps and cognitive function among inpatients with schizophrenia. *BMC Psychiatry* **2022**, *22*, 87. [CrossRef] [PubMed]
27. Yuenyongchaiwat, K.; Pipatsitipong, D.; Sangprasert, P. Increasing walking steps daily can reduce blood pressure and diabetes in overweight participants. *Diabetol. Int.* **2018**, *9*, 75–79. [CrossRef] [PubMed]
28. Nagovitsyn, R.S.; Osipov, A.Y.; Ratmanskaya, T.I.; Loginov, D.V.; Prikhodov, D.S. The Program for Monitoring Students' Walking and Running according to the System "10,000 Steps a Day" During the Spread of COVID-19. In Proceedings of the Winter Conferences of Sports Science, Costa Blanca Sports Science Events Alicante, Alicante, Spain, 22–23 March 2021.
29. Willis, W.T.; Ganley, K.J.; Herman, R.M. Fuel oxidation during human walking. *Metabolism* **2005**, *54*, 793–799. [CrossRef]
30. Hallam, K.T.; Bilsborough, S.; De Courten, M. "Happy feet": Evaluating the benefits of a 100-day 10,000 step challenge on mental health and wellbeing. *BMC Psychiatry* **2018**, *18*, 19. [CrossRef]
31. Li, X.; Wang, Y.; Zhang, B.; Ma, J. PSDRNN: An efficient and effective HAR scheme based on feature extraction and deep learning. *IEEE Trans. Ind. Inform.* **2020**, *16*, 6703–6713. [CrossRef]
32. Xia, K.; Huang, J.; Wang, H. LSTM-CNN architecture for human activity recognition. *IEEE Access* **2020**, *8*, 56855–56866. [CrossRef]
33. 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.
34. 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]
35. Verma, U.; Tyagi, P.; Kaur, M. Single Input Single Head CNN-GRU-LSTM Architecture for Recognition of Human Activities. *Indones. J. Electr. Eng. Inform* **2022**, *10*, 410–420. [CrossRef]
36. Li, F.; Shirahama, K.; Nisar, M.; Köping, L.; Grzegorzek, M. Comparison of Feature Learning Methods for Human Activity Recognition Using Wearable Sensors. *Sensors* **2018**, *18*, 679. [CrossRef]
37. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* arXiv:1409.1556, 2014.
38. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27 June–28 July 2016; Volume 1, pp. 2818–2826.
39. 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.
40. Tan, M.; Le, Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *Proc. Mach. Learn. Res.* **2019**, *97*, 6105–6114.
41. Lu, L.; Zhang, C.; Cao, K.; Deng, T.; Yang, Q. A multichannel CNN-GRU model for human activity recognition. *IEEE Access* **2022**, *10*, 66797–66810. [CrossRef]
42. Zhang, Y.; Zhang, Z.; Zhang, Y.; Bao, J.; Zhang, Y.; Deng, H. Human Activity Recognition Based on Motion Sensor Using U-Net. *IEEE Access* **2019**, *7*, 75213–75226. [CrossRef]
43. Zhang, H.; Xiao, Z.; Wang, J.; Li, F.; Szczerbicki, E. A novel IoT-perceptive human activity recognition (HAR) approach using multihead convolutional attention. *IEEE Internet Things J.* **2019**, *7*, 1072–1080. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.