IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received January 8, 2020, accepted January 14, 2020, date of publication January 20, 2020, date of current version January 30, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2967845

Feature Extraction Using an RNN Autoencoder for Skeleton-Based Abnormal Gait Recognition

KOOKSUNG JUN, DEOK-WON LEE, KYOOBIN LEE, SANGHYUB LEE, AND MUN SANG KIM

School of Integrated Technology, Gwangju Institute of Science and Technology, Gwangju 61005, South Korea

Corresponding author: Mun Sang Kim (munsang@gist.ac.kr)

This work was supported by the Ministry of Trade, Industry and Energy of Korea under Grant 10063300.

ABSTRACT In skeleton-based abnormal gait recognition, using original skeleton data decreases the recognition performance because they contain noise and irrelevant information. Instead of feeding original skeletal gait data to a recognition model, features extracted from the skeleton data are normally used. However, existing feature extraction methods might include laborious processes and it is hard for them to minimize the irrelevant information while preserving the important information. To solve this problem, an automatic feature extraction method using a recurrent neural network (RNN)-based Autoencoder (AE) is proposed in this paper. We extracted features from skeletal gait data by using two RNN AEs: a long short-term memory (LSTM)-based AE (LSTM AE) and a gated recurrent unit (GRU)-based AE (GRU AE). The features of the RNN AEs are compared to the original skeleton data and other existing features. We evaluated the features by feeding them to various discriminative models (DMs) and comparing the recognition performances. The features extracted by using the RNN AEs are more easily recognized and robust than the original skeleton data and other existing features. In particular, the LSTM AE shows a better performance than the GRU AE. Compared to single DMs fed with the original skeleton directly, hybrid models where the features of the RNN AEs are fed to DMs show a higher recognition accuracy with fewer training epochs and learning parameters. Therefore, the proposed automatic feature extraction method improves the performance of skeleton-based abnormal gait recognition by reducing laborious processes and increasing the recognition accuracy effectively.

INDEX TERMS Abnormal gait recognition, skeleton-based recognition, RNN Autoencoder, feature extraction, hybrid model, deep learning.

I. INTRODUCTION

Gait recognition is a very important research problem because the weakness in a specific function of the human body can be detected by recognizing an abnormal and unbalanced gait. Since body functions are weakened as people age, abnormal gaits are frequently observed in elderly people. Traditionally, inertial sensors, such as accelerometers and gyro sensors, are used to measure gait patterns. These sensors are attached to the body to measure the data, so it is hard to collect and analyze the data in our daily lives. With the development of depth sensors, such as Kinect, gait patterns can be easily measured without attaching sensors to the body when using them. Many methods using a depth sensor for skeleton-based gait analyses, such as gait parameter measurement [1]–[6], human identification [7]–[11], and abnormal gait recognition [12]–[21] have been proposed in the past few years.

Spatial-temporal skeleton data of the human gait are used to analyze gait patterns with different approaches. Skeleton-based algorithms that measure gait parameters, such as the stride length, the step length, the walking speed, the cadence, and the angle of the foot and hips, have been proposed [1]–[6]. These parameters are closely related to human health. Therefore, measuring them accurately is the main consideration of these algorithms. There have been other approaches based on gait pattern recognition. Human identification using skeletal gait data has been actively researched [7]–[11]. Since physical characteristics and natures of individuals permeate to their gait patterns, it is possible to identify individuals by using gait data. Human identification can be simplified by noncontact methods using the skeleton-based gait data obtained through depth sensors. Additionally, skeleton-based abnormal gait

The associate editor coordinating the review of this manuscript and approving it for publication was Lefei Zhang.

recognition has also received much attention. Many methods using machine learning algorithms to recognize normal and abnormal gait patterns have been developed [12]–[21].

Skeleton-based gait data are sequential times series data. Recurrent neural networks (RNNs) are a powerful deep learning algorithm to analyze sequential data. Long short-term memory (LSTM) [22] and gated recurrent units (GRUs) [23] are popular RNN architectures. They can overcome the vanishing gradient problem of the basic RNN [24], [25]. RNN architectures can be used to build RNN-based discriminative models (RNN DMs) for data classification, and RNN-based Autoencoders (RNN AEs) for feature extraction and data reconstruction. They have shown great performance in handling sequential data, such as speech recognition [26], [27], machine translation [28], [29], video analysis [30]–[32], and skeleton-based action recognition [33]–[35]. Therefore, it is appropriate to use RNN architectures when treating skeleton-based gait data.

To increase the performance of gait recognition, feature extraction from skeleton data is needed. The size and orientation of skeleton data are normalized to increase the robustness of the data [12], [18]. The normalized skeleton can be used as features. Additionally, the joint angles calculated by using skeleton data can be used as features [13], [18]. These developed features can improve the recognition performance due to their robustness. However, it is laborious to find a proper feature extraction method for the purpose, and manual extraction does not have the ability to extract the discriminative information from the data. Additionally, it is necessary to find a well-matched model because the manual features sometimes show lower performance than the original data on some discriminative models. To solve these problems, we propose an automatic feature extraction method by using an RNN AE.

AEs are a typical unsupervised machine learning algorithm to extract features from original data or to reconstruct the data [36]. In the training of the AE, the input is the original data, and the output is the reconstructed data. It is trained by reducing the difference between the original data and the reconstructed data [37]. As it is trained, irrelevant and redundant data can be reduced in the extracted features and the reconstructed data. Therefore, using the extracted features or the reconstructed data as the input of the DM achieves a higher accuracy than using the original data as the input.

Many methods using an LSTM AE have been proposed in various fields [31], [38]–[42]. Srivastava *et al.* [31] used LSTM networks to build an AE for video representation. Patraucean *et al.* [38] introduced a spatial-temporal AE composed of a convolutional LSTM to extract features from videos. Tu *et al.* [39] proposed a spatial-temporal data augmentation method using an LSTM AE for skeleton-based recognition of human action. Compared to an LSTM AE, there are only a few studies on a GRU AE [43]–[45] because a GRU was recently proposed. Both of these methods are considered the best RNN architectures, and they have been compared repeatedly in several fields. Despite the less complicated structure, a GRU has shown similar performances compared to an LSTM, but it has not been confirmed yet whether it can replace an LSTM. Therefore, it is meaningful to compare their performances when used to build an AE and a DM for skeleton-based abnormal gait recognition.

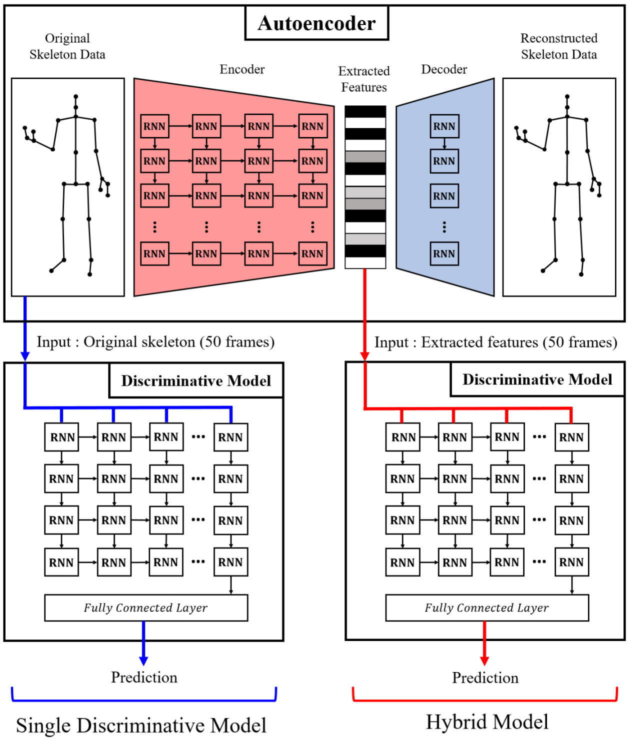


FIGURE 1. Illustration of the single discriminative model and the hybrid model.

In this paper, we propose a feature extraction method using an RNN AE to improve the performance of skeleton-based abnormal gait recognition. Instead of using a manual feature extraction method, we use an RNN AE to extract features from the original skeleton-based gait data. We evaluate the features by comparing them to the original skeleton data and other features. The original skeleton data or features are fed to various DMs for the evaluation, and in particular, an RNN DM is mainly treated. When the extracted features of an RNN AE are used as the input of an RNN DM, the model is defined as an RNN AE-DM hybrid model as shown in Fig. 1. A single RNN DM fed with the original skeleton data is used as a baseline to evaluate the hybrid model. We compare the recognition result of the hybrid model to the result of the single RNN DM to show that the features are more discriminative than the original skeleton data and that two-step training is more effective than a single-step training.

The following are the contributions of this paper:

- We propose an automatic feature extraction method from skeletal gait data by using two RNN AEs: an LSTM AE and a GRU AE. The extracted features are more discriminative and robust than the original skeleton data.
- The proposed RNN AEs minimize the amount of irrelevant information from the original skeleton data while

preserving the significant information to improve the recognition accuracy in abnormal gait recognition.
- The proposed feature extraction method shows better performance than other manual and automatic methods. It maximizes the recognition performance of the DMs without the laborious processes or loss of the important data, which decreases the performance in other existing methods.
- We compared the LSTM and GRU, which are considered the most powerful RNN architectures, when they are used to build an AE (LSTM AE vs GRU AE) and a DM (LSTM DM vs GRU DM).

This paper is organized as follows. In Section II, we review related works on skeleton-based abnormal gait recognition. In Section III, we briefly review RNN architectures used in the experiments and introduce the RNN DM, the RNN AE, and the RNN AE-DM hybrid model. In Section IV, we demonstrate the improvement by using the features of the RNN AEs and the effectiveness of the RNN AE-DMs. Finally, we provide a conclusion in Section V.

II. RELATED WORKS

In this section, we briefly review publicly accessible skeleton-based datasets of abnormal gaits. Then, we introduce machine learning methods for skeleton-based gait recognition. Finally, we describe feature extraction methods used in skeleton-based gait recognition.

A. 3D SKELETON DATASETS OF ABNORMAL GAITS

Many studies on skeleton-based abnormal gait recognition have been published [12]–[21]. However, there are a few publicly accessible 3D skeleton datasets of abnormal gaits [12], [17], [18], [46]. Publicly accessible datasets are collected by simulation of actors [12], [18] or using equipment [17], [46], such as padding a sole and attaching a weight, which cause abnormal gaits. The datasets are summarized as follows:

- Paiement *et al.* [18] collected normal, Parkinson's disease, and stoke gait data by using Kinect v1. Eleven subjects participated in the data collection, and only five of them simulated the abnormal gaits. Since the sensor looked down to the subjects when collecting the data, some joints were not well detected.
- Chaaraoui *et al.* [12] obtained abnormal gaits by using Kinect v2. Seven subjects participated in the data collection and simulated 4 abnormal gaits: right knee injury, left knee injury, right foot dragging, and left foot dragging. Since they collected only 4 datasets for the normal gait and 1 dataset for each abnormal gait from every subject, the number of datasets was small.
- Nguyen *et al.* [46] used equipment that can cause abnormal gaits. They used 5-cm, 10-cm, and 15-cm soles and a 4-kg weight. By padding the sole under the foot or attaching the weight to the ankle, they induced abnormal gaits. A treadmill was used to collect the gait data, and the data were obtained using Kinect v2. Nine subjects participated in the data collection, and each of them created 1 normal and 8 abnormal gait datasets. Each walking dataset contains 1,200 frames of the skeleton data.

TABLE 1. Description of gait datasets [46] used for evaluation.

Gait Type	Description
Gait 1	Normal gait
Gait 2	Padding a 5-cm-thick sole under the left foot
Gait 3	Padding a 10-cm-thick sole under the left foot
Gait 4	Padding a 15-cm-thick sole under the left foot
Gait 5	Weight of 4 kg on the left ankle
Gait 6	Padding a 5-cm-thick sole under the right foot
Gait 7	Padding a 10-cm-thick sole under the right foot
Gait 8	Padding a 15-cm-thick sole under the right foot
Gait 9	Weight of 4 kg on the right ankle

- Khokhlova *et al.* [17] also used a sole in a similar way to Nguyen *et al.* [46]. However, they used only a 7-cm sole in the data collection. They caused abnormal gaits by putting the sole into the right shoe or asking participants not to bend the right knee during walking. Twenty-seven subjects participated and walked between 5 and 7 times for each gait type.

Among these datasets, we chose the dataset collected by Nguyen *et al.* [46] to evaluate our methods. Other datasets have a small amount of data [12], [18], only a few abnormal gait types [17], [18], or much noise [18]. On the other hand, the selected dataset has the largest amount of data (9 people 9 gaits 1,200 frames), various gait types, and stable skeleton data. As shown in Table 1, the dataset contains similar gait patterns for which the only difference is the height of the sole, whereas other datasets are composed of easily distinguishable gait types. Classifying barely distinguishable gait types can effectively evaluate how powerful the model is.

B. SKELETON-BASED ABNORMAL GAIT RECOGNITION USING MACHINE LEARNING

Recently, skeleton-based abnormal gait recognition has received much attention because skeleton data can be easily collected in daily life without attaching sensors or markers by using a depth sensor. Various methods to recognize gait patterns have been proposed. In particular, machine learning algorithms have shown strength in gait recognition [12]–[21]. In this section, we briefly review studies using machine learning algorithms for abnormal gait recognition based on skeletal gait data.

Prochazka *et al.* [19] proposed a method for analysis of gait disorders and recognition of Parkinson's disease by applying a Bayesian classifier to skeleton data. Chaaraoui *et al.* [12] applied the bag of key poses algorithm to spatial-temporal gait features extracted from skeleton data for abnormal gait recognition. Tupa *et al.* [20] proposed a method to recognize

Parkinson's disease by applying deep learning to extracted gait parameters, such as the stride length and gait velocity. Nguyen *et al.* [13] used a hidden Markov model (HMM) to recognize abnormal gaits. Li *et al.* [21] applied the k-nearest neighbors (k-NN) classifier to a covariance matrix extracted from skeleton data for abnormal gait recognition. Khokhlova *et al.* [17] proposed an ensemble LSTM classifier to recognize abnormal gaits. They used dynamic features of low limb flexion extracted from skeleton data. Nguyen *et al.* [16] compared k-means clustering, Bayesian inference, a bag-of-words model, and an HMM for abnormal gait recognition.

C. FEATURE EXTRACTION FOR THE RECOGNITION OF ABNORMAL GAITS

Machine learning-based feature extraction has been widely applied in computer vision and data mining to improve classification accuracy [47], [48]. Transferring data to another domain helps a DM to more easily classify the data. However, machine learning-based feature extraction method for the skeleton-based recognition of abnormal gaits has not been proposed. Instead, various manual features have been introduced [12]–[14], [18]. The extracted features of the gait patterns are more easily recognized than the original skeleton data and are more robust to the environment. The manual feature extraction methods used for the recognition of abnormal gaits are described as follows:

- Chaaraoui *et al.* [12] used normalized skeleton data obtained by 3D transformation. They set the centroid of all joints as the origin coordinate and normalized the size of the skeleton based on the length between each joint and the centroid. Finally, they normalized the orientation of the skeleton by rotating it to align on the same axis.
- Paiement *et al.* [18] applied an averaging filter to skeleton data to reduce the noise. Then, they normalized the filtered skeleton data by using Procrustes analysis. Additionally, they used the angles between each joint and the hip center.
- Meng *et al.* [14] used the distances between two joints as features. Among obtainable 25 joints from Kinect v2, they selectively used 20 joints because the other 5 joints: the fingers, spine, and shoulders are sources of noise and redundancy in the recognition of abnormal gait.
- Nguyen *et al.* [13] used 7 joint angles from skeleton data as features for abnormal gait detection. The joint angles used are the left hip angle, right hip angle, left knee angle, right knee angle, left ankle angle, right ankle angle, and two feet angles.

The above features can improve the performance of abnormal gait recognition. However, some of the features perform differently depending on the DM. Therefore, it is required to find the DM that matches the manually extracted features well. In addition, some features require multiple processes, such as filtering, normalization, and joint selection.

III. METHODS

In this section, we briefly review three RNN architectures: basic RNN, LSTM, and GRU architectures, which are used in this paper. Then, we describe the RNN DM, the RNN AE, and the RNN AE-DM hybrid model. The purpose of the RNN DM is to recognize abnormal gaits. In this paper, the RNN DM is used to evaluate various features and to compose the RNN AE-DM. The purpose of the RNN AE is to extract features from the skeleton data, and it is used to align the data in a more discriminative way and to compose the RNN AE-DM. In the RNN AE-DM, the features of the RNN AE are fed to the RNN DM. This model is compared to a single RNN DM where skeleton data are directly fed to the DM.

A. RNN ARCHITECTURES

RNNs are powerful neural networks for handling sequential data, such as text, sounds, and human gestures [26]–[35]. Compared to the multilayer perceptron (MLP), considered the simplest deep neural network, RNNs also have learning parameters called hidden states. A key to handle sequential data is based on these hidden states. During training, they are updated based on both the previous and current information of the sequential data.

The hidden state h_t and the output y_t of the recurrent layer, where $t \in \{1, \ldots, T\}$ denotes the index of the frame, are calculated as follows:

$$h_t = \tanh\left(W_{hh}h_{t-1} + W_{xh}x_t + b_h\right) \tag{1}$$
$$y_t = W_{hy}h_t + b_y \tag{2}$$

where W and b denote the weights and biases, respectively, between elements.

The hidden state h_t is updated by using the current input x_t and the previous hidden state h_{t-1}. Thus, the previous information influences the neural network calculation of the current information in sequential data. The output values of the recurrent layer are calculated through (2). Basically, an LSTM and a GRU follow this equation to calculate the output value.

The basic RNN has the vanishing gradient problem for learning long-term dependencies [24], [25]. An LSTM [22] can solve the problem by updating the hidden states using additional learning variables composed of the forget gate value f_t, the input gate value i_t, the output gate value o_t, and the cell state value C_t as follows:

$$f_t = \sigma\left(W_{xf}x_t + W_{hf}h_{t-1} + b_f\right) \tag{3}$$
$$i_t = \sigma\left(W_{xi}x_t + W_{hi}h_{t-1} + b_i\right) \tag{4}$$
$$C_t = f_t \circ C_{t-1} + i_t \circ \tanh\left(W_{xC}x_t + W_{hC}h_{t-1} + b_C\right) \tag{5}$$
$$o_t = \sigma\left(W_{xo}x_t + W_{ho}h_{t-1} + b_o\right) \tag{6}$$
$$h_t = o_t \circ \tanh\left(C_t\right) \tag{7}$$

where W and b denote the weights and biases, respectively.

A GRU [23] can also solve the vanishing gradient problem of the basic RNN. It is known that a GRU achieves a similar

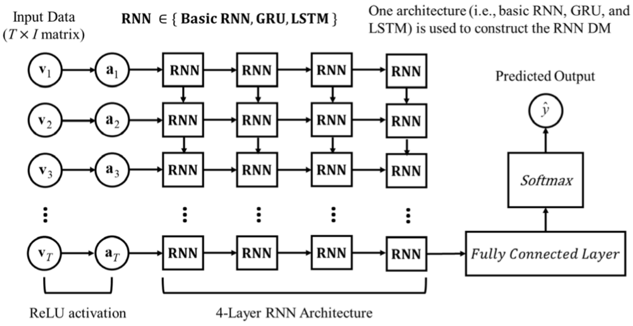


FIGURE 2. Structure of the RNN DM.

performance to an LSTM, but a GRU involves a smaller number of variables. Similar to an LSTM, a GRU is based on a gate structure. A GRU updates the hidden states by using the reset gate value r_t and the update gate value z_t as follows:

$$r_t = \sigma\left(W_{xr}x_t + W_{hr}h_{t-1} + b_r\right) \quad (8)$$

$$z_t = \sigma\left(W_{xz}x_t + W_{hz}h_{t-1} + b_z\right) \quad (9)$$

$$h_t = (1 - z_t) \circ h_{t-1} + z_t \circ \tanh\left(W_{xh}x_t + r_t \circ W_{hh}h_{t-1} + b_h\right) \quad (10)$$

where W and b denote the weights and biases, respectively.

B. RNN DM

In this paper, we apply an RNN DM to recognize the abnormal gait from the skeleton data. The RNN DM is mainly used to evaluate features extracted from the skeletal gait data. It is composed of a basic RNN, GRU, or LSTM, as shown in Fig. 2. The input data $\mathbf{v}_t$ is fed to the input layer of the RNN DM. The equation for the input layer is:

$$\mathbf{a}_t = \mathrm{ReLU}\left(\mathbf{W}_{\mathbf{v}a}\mathbf{v}_t + \mathbf{b}_\mathbf{a}\right) \quad (11)$$

where $\mathbf{a}_t$ denotes the activated value before putting it into the RNN structure. In the input layer, the input data $\mathbf{v}_t$ are used to calculate the activated value $\mathbf{a}_t$. The neural calculations based on the weight $\mathbf{W}_{\mathbf{v}a}$ and the bias $\mathbf{b}_\mathbf{a}$ are conducted, and then the results are activated by using a rectified linear unit (ReLU).

Each $\mathbf{v}_t$ contains x_i, where $i \in \{1, \ldots, I\}$ denotes the index of the input data in each frame. The value of I is the same as the number of input data in a single frame. If the original skeleton data are directly fed into the RNN DM, I is equal to 75 because the 3D coordinates of 25 joints compose a single frame of the data. On the other hand, I is the same as the number of features in a single frame when features are fed to the model.

We construct a 4-layer RNN architecture for the RNN DM. In the RNN layer, the hidden states $\mathbf{h}_t$ are updated by:

$$\mathbf{h}_t = \mathrm{RNN}\left(\mathbf{a}_t, \mathbf{h}_{t-1}, \mathbf{C}_{t-1}\right) \quad (12)$$

where RNN$(\cdot)$ denotes the selected RNN architecture among the basic RNN, LSTM, and GRU. For an LSTM, the previous cell state $\mathbf{C}_{t-1}$ is additionally used to update the hidden state, whereas the other architectures use only $\mathbf{a}_t$ and $\mathbf{h}_{t-1}$.

The last hidden state $\mathbf{h}_T$ is used to classify the data. The equation for the classification is as follows:

$$\hat{y} = \mathrm{softmax}\left(\mathbf{W}_{\mathbf{h}\hat{y}}\mathbf{h}_T + \mathbf{b}_{\hat{y}}\right) \quad (13)$$

where $\hat{y}$, $\mathbf{W}_{\mathbf{h}\hat{y}}$, and $\mathbf{b}_{\hat{y}}$ denote the predicted gait type, the output weight, and the output bias, respectively. A softmax classifier is used to recognize the gait pattern. During training, the cross-entropy cost function, L2 regularization, and adaptive moment estimation (Adam) [49] are applied to update the learning parameters.

C. RNN AE

We propose a sequence to sequence RNN AE to extract the sequential features from the original skeleton data. In this paper, we use two RNN AEs: LSTM AE and GRU AE. They have the same structure except for the architecture used to compose the RNN layers. As shown in Fig. 3, the RNN AE consists of an encoder and a decoder. In the encoding layers, the dimensions are reduced, and the features are extracted. The equations for the encoder are:

$$\mathbf{h}_t^{(E)} = \mathrm{RNN}\left(\mathbf{v}_t, \mathbf{h}_{t-1}^{(E)}, \mathbf{C}_{t-1}\right) \quad (14)$$

$$\mathbf{f}_t = \mathbf{W}_{\mathbf{h}f}^{(E)}\mathbf{h}_t^{(E)} + \mathbf{b}_\mathbf{f}^{(E)} \quad (15)$$

where $\mathbf{h}_t^{(E)}$ and $\mathbf{f}_t$ denote the hidden states of the encoder and the extracted features, respectively, at frame t. $\mathbf{f}_t$ is composed of f_{tn}, where $n \in \{1, \ldots, N\}$ and N denotes the number of features. Thus, the extracted features of all the frames are formatted as a $T \times N$ matrix. $\mathbf{f}_t$ is calculated by using the weight $\mathbf{W}_{\mathbf{h}f}^{(E)}$ and the bias $\mathbf{b}_\mathbf{f}^{(E)}$. Four RNN layers compose the encoder, and the number of hidden neurons in one RNN unit is equal to the number of features in a single frame.

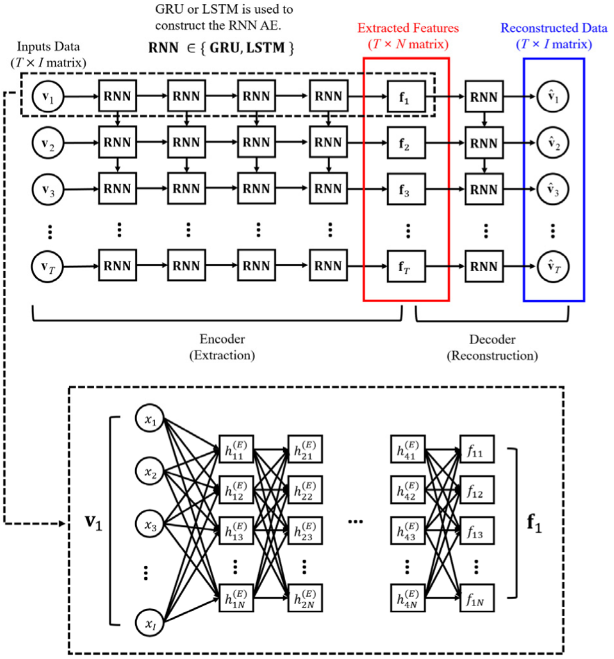


FIGURE 3. Structure of the RNN AE.

The extracted features of all the frames are fed to the decoder. In the decoding layer, the dimensions are expanded, and the features are used to reconstruct the data. The equations for the hidden state of the decoder $\mathbf{h}_t^{(D)}$ and the reconstructed input data $\hat{\mathbf{v}}_t$ are as follows:

$$\mathbf{h}_t^{(D)} = \mathrm{RNN}\left(\mathbf{f}_t, \mathbf{h}_{t-1}^{(D)}, \mathbf{C}_{t-1}\right) \tag{16}$$

$$\hat{\mathbf{v}}_t = \mathbf{W}_{\mathbf{h}\hat{\mathbf{v}}}^{(D)}\mathbf{h}_t^{(D)} + \mathbf{b}_{\hat{\mathbf{v}}}^{(D)} \tag{17}$$

$$L = \sum_t (\mathbf{v}_t - \hat{\mathbf{v}}_t)^2 \tag{18}$$

where $\mathbf{W}_{\mathbf{h}\hat{\mathbf{v}}}^{(D)}$ and $\mathbf{b}_{\hat{\mathbf{v}}}^{(D)}$ denote the weight and the bias, respectively, from the hidden states to the reconstructed data. The loss L is defined as sum of mean squared error between the input skeleton data and the reconstructed skeleton data. A single RNN layer composes the decoder, and the number of hidden neurons is the same as the amount of the skeleton data in a single frame. To maximize the representation performance of the features, we use a single RNN layer in the decoder, whereas multiple RNN layers compose the encoder. The training is conducted by reducing the difference between the original input $\mathbf{v}_t$ and the reconstructed data $\hat{\mathbf{v}}_t$.

D. RNN AE-DM HYBRID MODEL

Whereas the single RNN DM is fed with the original skeleton data directly, the features extracted by using the RNN AE are fed to the RNN DM in the proposed hybrid model, as shown in Fig. 4. The original skeleton data of size $T \times I$ are transformed to the features of size $T \times N$ through the RNN AE. Then, they are fed to the RNN DM, and the input gait pattern is recognized. The training is conducted separately on the RNN AE and the RNN DM. After the RNN AE is trained to successfully extract the features, the RNN DM is trained to recognize gaits by using the features. We supposed that the features are successfully extracted when the difference between the input data and the reconstructed data is minimized.

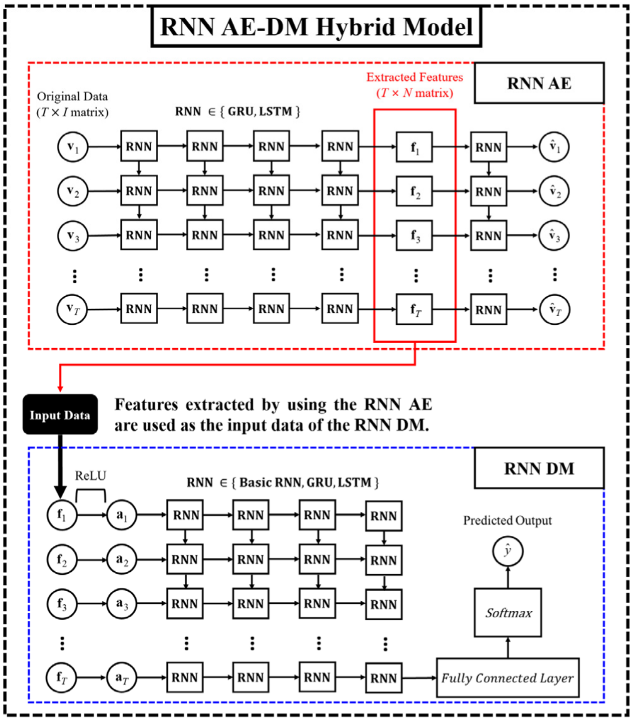


FIGURE 4. Structure of the RNN AE-DM hybrid model.

IV. EXPERIMENTS

In this section, we conducted various experiments to evaluate the features of the RNN AEs for skeleton-based abnormal gait recognition. First, we evaluated the features extracted through the RNN AEs by comparing them to the original skeleton data. Then, we conducted experiments to show effectiveness of the proposed RNN AE-DMs compared to the single RNN DMs. Finally, we compared the performance between the features of the RNN AEs and other developed features used in skeleton-based abnormal gait recognition.

We evaluated our model on the gait datasets by Nguyen *et al.*[46], which are composed of 1 normal and 8 abnormal gaits. There are 81 gait datasets for 9 people and 9 gaits. Each dataset contains 1,200 frames of skeleton data. However, this number of frames could be obtained because the data were collected by using a treadmill. The purpose of this paper is to develop a gait recognition model that is applicable in real life. Therefore, we needed to set the number of frames obtainable by just walking in front of the sensor. According to Microsoft, Kinect v2 can obtain skeleton data stably when the distance between the sensor and a person ranges from 1.2 m to 3.5 m. Since approximately 50 frames of skeleton data can be obtained when a person walks in the recommended range, we set the number of frames to feed into the model to 50.

We divided the 1,200 frames of each gait dataset into 24 smaller datasets. Thus, each sliced dataset contains 50 frames of the skeleton data, and there are total 1,944 sliced gait datasets. The datasets are fed to DM or AE. When the sliced gait datasets or the extracted features are fed to DM, 3-fold cross-validation is used to evaluate the model. The 3-fold cross-validation is conducted 5 times, so a total of 15 trainings are conducted and the average test accuracies are used for the evaluation.

The configuration of computer used for experiments is Intel(R) Core(TM) i7-7700K central processing unit (CPU), 8.00 GB random-access memory (RAM), and NVIDIA GeForce 1050-Ti. The computational cost for the RNN AEs and RNN DMs depends on the RNN architecture used. For the RNN DM, the running times of the basic RNN, LSTM, and GRU for a single epoch are 0.33s, 0.76s, and 0.82s, respectively. This method requires a minimum of 300 epochs to satisfactorily train the RNN DMs. For the RNN AE, the running times of the LSTM and GRU for a single epoch are 0.14s and 0.17s, respectively. This approach also requires a minimum of 50,000 epochs to effectively train the RNN AEs.

TABLE 2. Recognition accuracy as the number of features of the RNN AEs changes.

Input Data	Number of Features	Data Compression Ratio	Discriminative Model		
			Basic RNN	LSTM	GRU
Original Skeleton	/	/	88.3%	91.3%	91.4%
Features of the LSTM AE	60	0.80	92.2%	94.5%	95.4%
	50	0.67	92.8%	94.7%	95.9%
	40	0.53	91.2%	92.7%	93.6%
	30	0.40	89.1%	91.9%	92.1%
	20	0.27	86.8%	88.2%	87.6%
	10	0.13	82.0%	83.3%	84.5%
Features of the GRU AE	60	0.80	90.0%	92.1%	93.2%
	50	0.67	92.5%	94.0%	94.9%
	40	0.53	92.2%	92.3%	93.3%
	30	0.40	90.2%	93.0%	93.0%
	20	0.27	90.0%	91.4%	90.3%
	10	0.13	79.5%	78.8%	79.5%

TABLE 3. Improvement by using features of the RNN AEs compared to the original skeleton data.

Discriminative Model	Input Type			Improvement by Using Features	
	Original Skeleton	Features of the GRU AE	Features of the LSTM AE	GRU AE	LSTM AE
Basic RNN	88.3 %	92.5 %	92.8 %	**5.6 %**	**5.9 %**
LSTM	91.3 %	94.0 %	94.7 %	**2.4 %**	**3.1 %**
GRU	91.4 %	94.9 %	95.9 %	**4.8 %**	**5.8 %**
CNN	86.8 %	93.3 %	93.4 %	**6.5 %**	**6.6 %**
MLP	84.5 %	90.3 %	91.5 %	**5.8 %**	**7.0 %**
k-means	79.8 %	83.3 %	90.9 %	**3.5 %**	**11.1 %**
k-NN	80.4 %	84.5 %	91.2 %	**4.2 %**	**10.8 %**
Random Forest	79.4 %	86.0 %	91.0 %	**6.6 %**	**11.6 %**

A. IMPROVEMENT BY USING THE FEATURES OF THE RNN AE

We conducted the experiments to evaluate the features extracted by using the RNN AEs. In the experiments, we compared the recognition accuracies when using the features of the RNN AEs and the original skeleton data. The features were extracted by using two RNN AEs: the LSTM AE and the GRU AE. Through the experiments, we also compared the performances of the features between the two RNN AEs. We fed the extracted features or the original skeleton data into DMs for the comparison.

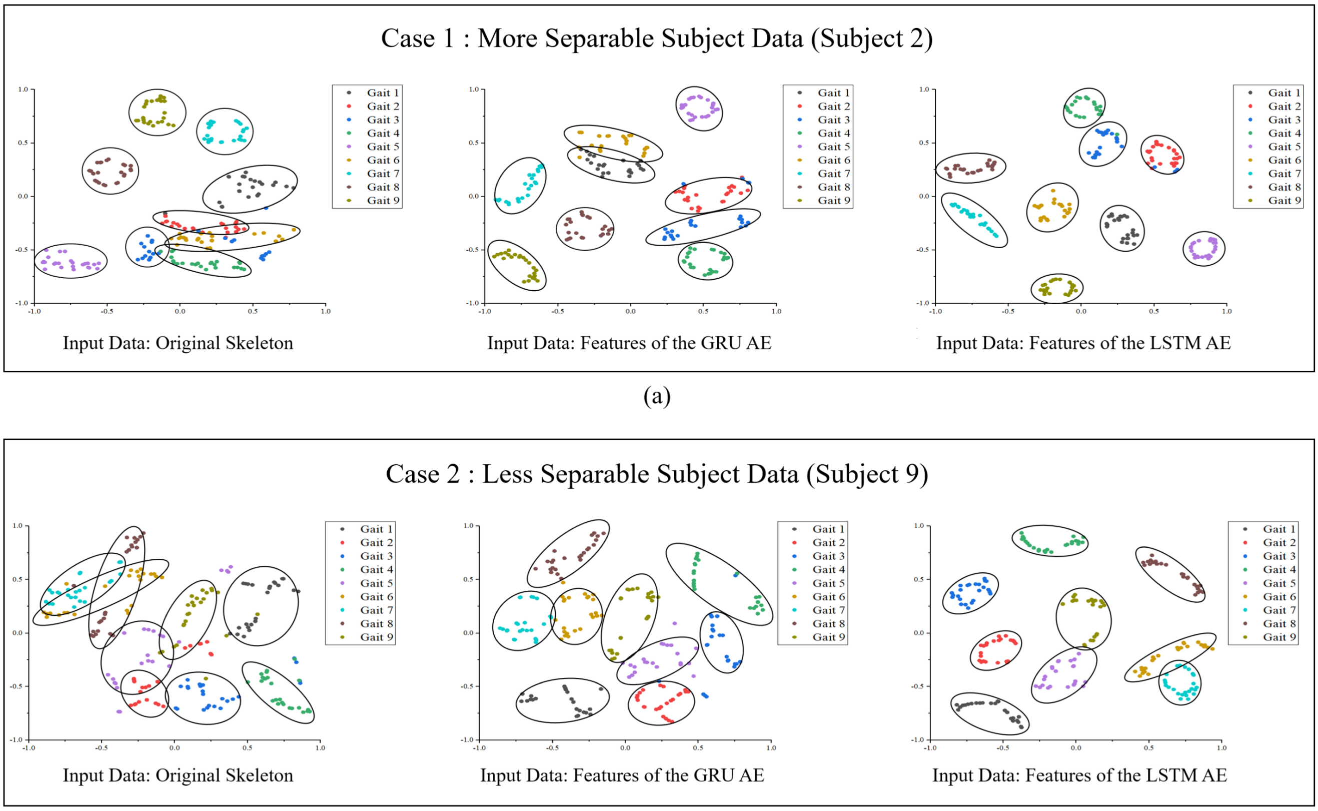
The number of extracted features is related to representativeness and loss of the data. If the number of extracted features is too few or too many, they do not effectively represent the data. There is a loss of important data when the number of extracted features is too few. On the other hand, there are irrelevant data in the extracted features if there are too many extracted features. Therefore, it is important to find the proper number of extracted features through the experiment. The number of data points in a single frame is 75 because the 3D coordinates of the 25 joints compose the skeleton. We extracted the features by changing the number in a single frame from 10 to 60 and compared the recognition accuracy between them. The data compression ratio is defined as the number of extracted features over the number of original data points. We fed the extracted features of different sizes and the original skeleton data into the three RNN DMs to find the proper number of features.

Table 2 shows the recognition accuracy as the number of the extracted features changes. When the number of features in a single frame is 30 or more, feeding the extracted features of the RNN AEs shows a higher accuracy than feeding the original skeleton data. Although the features have less data than the original data, the results using them show better performances. If the number of the features of the LSTM AE and GRU AE is less than 30, there is a loss of the important data, and the features do not represent the gait patterns effectively. Therefore, it is not proper to use these features. One of the most important considerations in feature extraction is to preserve the important data while reducing the dimensionality.

When the number of features is equal to 50, the highest recognition accuracy is achieved. If the number of features is more than 50, they contain unnecessary data for abnormal gait recognition, which reduces the performance. Therefore, we considered the features containing 50 data points in a single frame as the best effective features and conducted the following experiments with them.

Table 3 shows the improvement by using the features of the RNN AEs compared to using the original skeleton data directly. The features were fed to various DMs, including the RNN DMs and other machine learning algorithms. Both the features of the LSTM AE and the GRU AE improve the performance of the DMs without exception. In particular, the features of the LSTM AE achieve a higher accuracy than the features of the GRU AE in the overall results. The results show that the extracted features of the RNN AEs are more easily recognizable than the original skeleton data. Through the neural calculations during the training of the RNN AEs, the noise in the original data is reduced, and the features are aligned in a more discriminative way. Therefore, although the amount of the input data is reduced, the recognition accuracy is higher when using the extracted features.

Non-RNN models, such as convolutional neural network (CNN), MLP, k-means clustering, k-NN, and random forest algorithms, have a lower ability to recognize the sequential data than the RNN DMs. However, they achieve a higher improvement than the RNN DMs when feeding the



(a)

(b)

FIGURE 5. Improvement by using the features of the RNN AEs. The t-SNE Visualization of the input data is used to show the effectiveness of the features. (a) Visualization of more separable subject data (Subject 2). (b) Visualization of less separable subject data (Subject 9).

TABLE 4. Recognition accuracy of the RNN AE-DMs and the end-to-end models.

Encoding Layers	Discriminative Layers	End-to-End Model	RNN AE-DM
GRU	Basic RNN	91.5 %	92.5 %
GRU	LSTM	88.1 %	94.0 %
GRU	GRU	91.5 %	94.9 %
LSTM	Basic RNN	90.3 %	92.8 %
LSTM	LSTM	89.5 %	94.7 %
LSTM	GRU	91.3 %	95.9 %

features of the RNN AEs. The results show that the features of the RNN AEs can reduce the nonlinearity caused by the sequential characteristics of the gait data and can help the non-RNN models recognize the data more easily. Although greater improvements are achieved by the non-RNN DMs, the recognition accuracy itself is higher in the RNN DMs.

The LSTM AE + GRU DM hybrid model achieves the best performance. The performances of LSTM and GRU are different depending on the model with them. When they are used to compose the RNN AE, a better performance is achieved by the LSTM. On the other hand, the GRU shows higher recognition accuracy compared to the LSTM when they compose the RNN DM.

B. VISUALIZATION OF THE INPUT DATA

We applied t-distributed stochastic neighbor embedding (t-SNE) to visually show that the extracted features of the RNN AEs are more recognizable than the original skeleton data. The t-SNE method reduces the dimensionality of the input data. In general, it is used to visualize the input data in 2D or 3D space. The better grouped the data are, the more discriminative they are. Therefore, the improvement by using the features of the RNN AEs compared to the original data can be visually shown by using the t-SNE method. We fed the extracted features or the original skeleton data of the 9 gait patterns of each subject into the t-SNE model. Among the subjects, we selected two cases (i.e., a more separable subject data and a less separable subject data) to evaluate the features. The number of features in a single frame used in this experiment is also 50. Fig. 5 shows the results of the t-SNE visualization in 2D space.

In the results of the more separable subject data, there are some overlapping areas for the groups when feeding the original skeleton data of the 9 gait patterns. When feeding the features of the RNN AEs, the groups are more recognizable, although there are still overlapping areas in the GRU AE. The results show that the LSTM AE achieves the best performance. The gait pattern groups are clearly recognizable when feeding them. In the results of the less separable subject data, it is hard to recognize the gait patterns when using the original skeleton data. We applied t-SNE to these data by changing the learning configurations many times, but the results did not change. However, when feeding the features of the GRU AE and the LSTM AE, the gait patterns are much more easily grouped together. In particular, when using the features of the LSTM AE, the gait patterns can be clearly recognized.

The results show that the gait patterns become more recognizable in both of the more and less separable cases when using the features extracted by using the RNN AEs. In particular, the features of the LSTM AE achieve the best performance. The gait patterns are clearly recognizable even in the less separable subject data. The LSTM AE extracts more recognizable and robust features than the GRU AE. The results of the t-SNE visualization support the improvement by using the features of the RNN AEs compared to the results using the original skeleton data.

C. EFFECTIVENESS OF THE RNN AE-DM

We compared the RNN AE-DM hybrid models and the single RNN DMs with respect to effectiveness by showing the training curve of the models and the training results of the models as the number of the learning parameters changes. The number of extracted features was fixed at 50 in the experiments.

Fig. 6 shows the test accuracy curves when feeding the features of the RNN AEs and the original skeleton data into the three RNN DMs. The results show that the hybrid models are more effective than the single DMs in terms of the training speed. The training speed is related to how easily the input data can be recognized. In all the graphs, the recognition accuracy when feeding the features of the RNN AEs increases faster than the result using the original skeleton data. The features of the RNN AEs accelerate the training and increase the recognition accuracy. In particular, the fastest increase in the recognition accuracy is achieved when the RNN DMs are fed with the features of the LSTM AE.

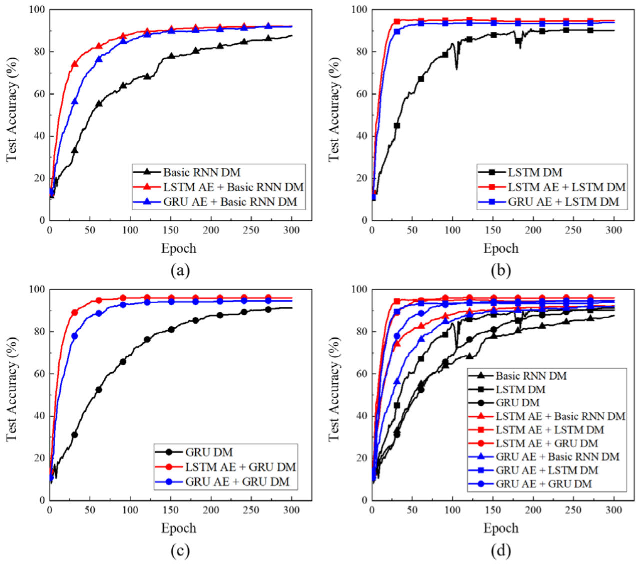


FIGURE 6. Test accuracy curves when feeding the three types of input data to the three RNN DMs: (a) the basic RNN DM, (b) the LSTM DM, and (c) the GRU DM. (d) Comparison between the results.

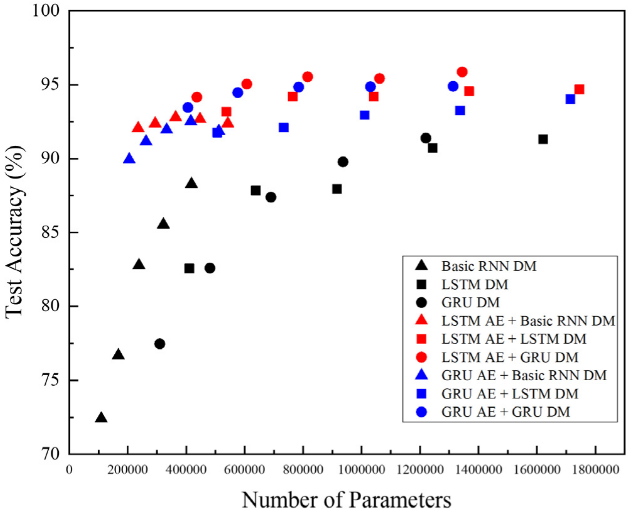


FIGURE 7. Training results of the models as the number of parameters changes.

A likely question is whether the RNN AE-DM hybrid models achieve a better performance at the expense of more learning parameters than do single RNN DMs. Given that the use of the RNN AEs means including more learning parameters, we therefore conducted an experiment to show that the hybrid models show higher accuracy with a similar number of parameters than do single DMs. Fig. 7 shows the recognition accuracy of each model as the number of learning parameters increases. We changed the number of parameters by increasing or decreasing the hidden neurons of the DMs while keeping the number of hidden layers constant. The results show that the hybrid models using the features of the RNN AEs outperform the single DMs with the same or smaller number of parameters.

D. EVALUATION OF TWO-STEP TRAINING OF THE RNN AE-DM

The RNN AE-DM hybrid model is a two-step training model. The purposes of each step are different. First, the RNN AE is trained to extract the features from skeleton data. In this training, it is focused on extracting the meaningful features and reconstructing the data as much as possible. Then, the RNN DM is trained to recognize the abnormal gait patterns by using the extracted features. This training focuses on classifying the gait patterns accurately. We evaluated the two-step training of the RNN AE-DMs by comparing the recognition accuracies of the RNN AE-DMs to those of End-to-End models that have a single-step training.

Fig. 8 shows the structure of the End-to-End model. The data flow is exactly same as the RNN AE-DM. The features extracted from the encoding layers are fed to the discriminative layers. However, interaction between the encoding layers and the discriminative layers is different. In the

TABLE 5. Recognition accuracy when using the features of the RNN AEs and other features.

Input Data Type	Discriminative Model							
	GRU	LSTM	Basic RNN	CNN	MLP	k-means	k-NN	Random Forest
Original Skeleton	91.4 %	91.3 %	88.3 %	86.8 %	84.5 %	79.8 %	80.4 %	79.4 %
Features of the LSTM AE	**95.9 %**	**94.7 %**	**92.8 %**	**93.4 %**	**91.5 %**	**90.9 %**	91.2 %	**91.0 %**
Features of the GRU AE	94.9 %	94.0 %	92.5 %	93.3 %	90.3 %	83.3 %	84.5 %	90.4 %
Features of PCA	84.8 %	86.6 %	82.3 %	89.0 %	87.7 %	77.3 %	80.7 %	84.0 %
Features of SVD	83.8 %	83.8 %	82.6 %	89.4 %	87.6 %	77.1 %	80.2 %	83.3 %
Normalized Skeleton [12]	93.8 %	91.7 %	89.2 %	91.9 %	89.5 %	87.1 %	91.4 %	84.9 %
Filtering + Normalized Skeleton [18]	93.1 %	91.9 %	89.6 %	93.1 %	91.4 %	87.6 %	**92.2 %**	87.7 %
Filtering + Angles Between Each Joint and the Hip [18]	83.7 %	89.1 %	74.9 %	83.9 %	81.1 %	83.2 %	86.8 %	84.5 %
Lengths Between 20 Joints [14]	92.4 %	91.4 %	86.3 %	91.2 %	86.3 %	88.0 %	92.1 %	89.5 %
7 Joint Angles [13]	86.9 %	89.5 %	66.9 %	84.4 %	66.9 %	77.2 %	84.4 %	75.5 %

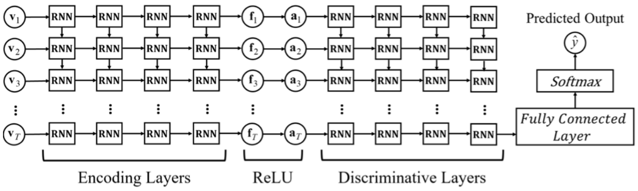


FIGURE 8. Structure of the End-to-End model.

RNN AE-DM, there is no interaction because the encoding and the discriminative layers are trained separately. The training of the RNN DM does not affect the parameters of the RNN AE. On the other hand, all the parameters in the encoding and the discriminative layers are updated at the same time in the training of the End-to-End model. Additionally, the End-to-End model does not have the decoding layers that are needed to reconstruct the data in the RNN AE.

We conducted two experiments with the End-to-End model: 1) training with randomly initialized parameters, and 2) fine-tuning with parameters of the most accurate results of two-step training. Table 4 shows the results of the first experiment. The RNN AE-DMs achieve a higher accuracy than that of the End-to-End models with randomly initialized parameters in all of the combinations of the encoding and the discriminative layers. In the fine-tuning experiment, the recognition accuracy does not increase. Therefore, we verified the effectiveness of the two-step training of the RNN AE-DM. The separate training steps for the different purposes are more effective than the single training step of the End-to-End model. Consequently, the complicated recognition problem can be changed into two simpler problems by using the RNN AE-DM hybrid model, so it improves the recognition accuracy.

E. COMPARISON BETWEEN THE RNN AE FEATURES AND OTHER FEATURES

We evaluated the features of the RNN AEs by comparing them with other existing features. Since there are noise and unnecessary data in skeleton data for abnormal gait recognition, the manual features are normally used instead of directly feeding the original skeleton data into a learning model [12]–[14], [18]. The detailed descriptions of the other features are in Section II. We compared our features to the manual features with respect to the recognition accuracy. Furthermore, we compared our features with other automatically extracted features obtained by using principal component analysis (PCA) and singular value decomposition (SVD) which are typical dimension reduction methods. In this experiment, the number of features extracted by using the RNN AEs was fixed at 50, which achieves the best performance, as previously mentioned. In the experiment, we separately set the learning configurations of the DMs, such as the learning rate and the number of training iterations, for each model to maximize their performance. Table 5 shows the results of the experiment.

According to the results, the features of the LSTM AE achieve the best performance in skeleton-based abnormal gait recognition. The highest accuracies are achieved when the features of the LSTM AE are fed to each DM, except for k-NN. The features of the GRU AE are the next best, but they show a relatively low performance when fed to k-means and k-NN. Compared to other features, in most cases, the RNN AEs achieve a better performance when they are fed to the DMs. There is a loss of significant information in the other existing methods, but this loss can be minimized by using the feature extraction of the RNN AEs. Among the other features, the filtered and normalized skeleton [18] achieves the best performance.

PCA and SVD can also extract features automatically. They reduce the dimensionality of the data in an unsupervised way. However, their features do not improve the recognition accuracy of all the DMs. As shown in Table 5, they improve the recognition performance only in the CNN, MLP, and random forest. In particular, the performance of the RNN DMs are remarkably decreased. It is hard for them to preserve the sequential characteristics of the data when reducing the dimensionality. On the other hand, the feature extraction of the RNN AEs not only reduces the dimensionality but also aligns the data in a more discriminative way that also preserves the sequential characteristics. The RNN AEs require the longer training time to extract the meaningful features than PCA and SVD. However, the features of the RNN AEs show much better performance than the features of PCA and SVD.

The performance of the other features is heavily influenced by the DM. In particular, the features based on angles show remarkably different results depending on DMs. The results are worse than the original skeleton data in some DMs while achieving fairly good performances with other models. Therefore, when using the other features, the well-matched DM must be found. On the other hand, the features of the RNN AEs achieve a higher recognition accuracy than the original skeleton data regardless of the DM.

V. CONCLUSION

In skeleton-based abnormal gait recognition, feature extraction from skeleton data is necessary because the original data contain noise and irrelevant information, which decrease the recognition accuracy. In this paper, we propose the feature extraction method using the RNN AEs. The RNN AEs align the original skeleton data in a more discriminative way and minimize the irrelevant information, especially notable with the LSTM AE which shows better performance than the GRU AE. Therefore, the features improve the recognition performance of the RNN DMs and the other DMs. Among the DMs, the GRU DM shows the best performance in the most features. Whereas the LSTM AE shows a better performance than the GRU AE, the GRU DM achieves a higher accuracy than the LSTM DM. The RNN AE-DM hybrid models fed with the features show better performance than the single RNN DMs fed with the original skeleton data, and do so with less training time and fewer learning parameters. Furthermore, the two-step training of the RNN AE-DM is more effective than the single-step training of the End-to-End model that has the same data flow.

ACKNOWLEDGMENT

(Kooksung Jun and Deok-Won Lee contributed equally to this work.)

REFERENCES

[1] M. Gabel, R. Gilad-Bachrach, E. Renshaw, and A. Schuster, "Full body gait analysis with Kinect," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2012, pp. 1964–1967.

[2] E. E. Stone and M. Skubic, "Unobtrusive, continuous, in-home gait measurement using the microsoft Kinect," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 10, pp. 2925–2932, Oct. 2013.

[3] E. Auvinet, F. Multon, C.-E. Aubin, J. Meunier, and M. Raison, "Detection of gait cycles in treadmill walking using a Kinect," *Gait Posture*, vol. 41, no. 2, pp. 722–725, Feb. 2015.

[4] E. Cippitelli, S. Gasparrini, S. Spinsante, and E. Gambi, "Kinect as a tool for gait analysis: Validation of a real-time joint extraction algorithm working in side view," *Sensors*, vol. 15, no. 1, pp. 1417–1434, Jan. 2015.

[5] X. Xu, R. W. Mcgorry, L.-S. Chou, J.-H. Lin, and C.-C. Chang, "Accuracy of the Microsoft Kinect for measuring gait parameters during treadmill walking," *Gait Posture*, vol. 42, no. 2, pp. 145–151, Jul. 2015.

[6] D. J. Geerse, B. H. Coolen, and M. Roerdink, "Kinematic validation of a multi-Kinect v2 instrumented 10-meter walkway for quantitative gait assessments," *PLoS ONE*, vol. 10, no. 10, Oct. 2015, Art. no. e0139913.

[7] A. Sinha, "Person identification using skeleton information from kinect," in *Proc. 6th Int. Conf. Adv. Comput. Hum. Interact. (ACHI)*, 2013, pp. 101–108.

[8] S. Jiang, Y. Wang, Y. Zhang, and J. Sun, "Real time gait recognition system based on Kinect skeleton feature," in *Proc. ACCV*, 2014, pp. 46–57.

[9] E. Gianaria, "Human classification using gait features," in *Biometric Authentication*. Sofia, Bulgaria: International Workshop on Biometric Authentication, 2014, pp. 16–27.

[10] J. Preis, "Gait recognition with Kinect," in *Proc. 1st Workshop Kinect Pervas. Comput.*, 2012, pp. 1–4.

[11] D. Kastaniotis, I. Theodorakopoulos, C. Theoharatos, G. Economou, and S. Fotopoulos, "A framework for gait-based recognition using Kinect," *Pattern Recognit. Lett.*, vol. 68, pp. 327–335, Dec. 2015.

[12] A. A. Chaaraoui, J. R. Padilla-Lopez, and F. Florez-Revuelta, "Abnormal gait detection with RGB-D devices using joint motion history features," in *Proc. 11th IEEE Int. Conf. Workshops Autom. Face Gesture Recognit. (FG)*, vol. 7, May 2015, pp. 1–6.

[13] T.-N. Nguyen, H.-H. Huynh, and J. Meunier, "Skeleton-based abnormal gait detection," *Sensors*, vol. 16, no. 11, p. 1792, Oct. 2016.

[14] M. Meng, H. Drira, M. Daoudi, and J. Boonaert, "Detection of abnormal gait from skeleton data," in *Proc. 11th Int. Joint Conf. VISIGRAPP*, 2016, pp. 133–139. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01703237

[15] T.-N. Nguyen, H.-H. Huynh, and J. Meunier, "Estimating skeleton-based gait abnormality index by sparse deep auto-encoder," in *Proc. IEEE 7th Int. Conf. Commun. Electron. (ICCE)*, Jul. 2018, pp. 311–315.

[16] T.-N. Nguyen and J. Meunier, "Applying adversarial auto-encoder for estimating human walking gait abnormality index," *Pattern Anal. Appl.*, vol. 22, no. 4, pp. 1597–1608, Nov. 2019.

[17] M. Khokhlova, C. Migniot, A. Morozov, O. Sushkova, and A. Dipanda, "Normal and pathological gait classification LSTM model," *Artif. Intell. Med.*, vol. 94, pp. 54–66, Mar. 2019.

[18] A. Paiement, L. Tao, M. Camplani, S. Hannuna, D. Damen, and M. Mirmehdi, "Online quality assessment of human motion from skeleton data," in *Proc. Brit. Mach. Vis. Conf.*, 2014, pp. 153–166.

[19] A. Procházka, O. Vyšata, M. Vališ, O. Ťupa, M. Schätz, and V. Mařík, "Bayesian classification and analysis of gait disorders using image and depth sensors of Microsoft Kinect," *Digit. Signal Process.*, vol. 47, pp. 169–177, Dec. 2015.

[20] O. Ťupa, A. Procházka, O. Vyšata, M. Schätz, J. Mareš, M. Vališ, and V. Maŕík, "Motion tracking and gait feature estimation for recognising Parkinson's disease using MS Kinect," *Biomed. Eng. Online*, vol. 14, no. 1, p. 97, Dec. 2015.

[21] Q. Li, Y. Wang, A. Sharf, Y. Cao, C. Tu, B. Chen, and S. Yu, "Classification of gait anomalies from Kinect," *Vis. Comput.*, vol. 34, no. 2, pp. 229–241, Feb. 2018.

[22] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[23] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," Dec. 2014, *arXiv:1412.3555*. [Online]. Available: https://arxiv.org/abs/1412.3555

[24] S. Hochreiter, "The vanishing gradient problem during learning recurrent neural nets and problem solutions," *Int. J. Uncertainty Fuzziness Knowl.-Based Syst.*, vol. 6, no. 2, pp. 107–116, Apr. 1998.

[25] R. Dey and F. M. Salemt, "Gate-variants of gated recurrent unit (GRU) neural networks," in *Proc. IEEE 60th Int. Midwest Symp. Circuits Syst. (MWSCAS)*, Aug. 2017, pp. 1597–1600.

[26] A. Graves, N. Jaitly, and A.-R. Mohamed, "Hybrid speech recognition with deep bidirectional LSTM," in *Proc. IEEE Workshop Autom. Speech Recognit. Understand.*, Dec. 2013, pp. 273–278.

[27] A. Graves, A.-R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 6645–6649.

[28] K. Cho, B. van Merrienboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," Sep. 2014, *arXiv:1409.1259*. [Online]. Available: https://arxiv.org/abs/1409.1259

[29] M.-T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," Aug. 2015, *arXiv:1508.04025*. [Online]. Available: https://arxiv.org/abs/1508.04025

[30] J. Yue-Hei Ng, M. Hausknecht, S. Vijayanarasimhan, O. Vinyals, R. Monga, and G. Toderici, "Beyond short snippets: Deep networks for video classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 4694–4702.

[31] N. Srivastava, "Unsupervised learning of video representations using lstms," in *Proc. ICML*, 2015, pp. 843–852.

[32] N. Ballas, L. Yao, C. Pal, and A. Courville, "Delving deeper into convolutional networks for learning video representations," Nov. 2015, *arXiv:1511.06432*. [Online]. Available: https://arxiv.org/abs/1511.06432

[33] Y. Du, W. Wang, and L. Wang, "Hierarchical recurrent neural network for skeleton based action recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1110–1118.

[34] R. Cui, A. Zhu, S. Zhang, and G. Hua, "Multi-source learning for skeleton -based action recognition using deep LSTM networks," in *Proc. 24th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2018, pp. 547–552.

[35] J. Liu, G. Wang, L.-Y. Duan, K. Abdiyeva, and A. C. Kot, "Skeleton-based human action recognition with global context-aware attention LSTM networks," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 1586–1599, Apr. 2018.

[36] B. Boehmke and B. Greenwell, "Autoencoders," *Deep learning*. Cambridge, MA, USA: MIT Press, 2016, pp. 499–523.

[37] G. E. Hinton, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006.

[38] V. Patraucean, A. Handa, and R. Cipolla, "Spatio-temporal video autoencoder with differentiable memory," Nov. 2015, *arXiv:1511.06309*. [Online]. Available: https://arxiv.org/abs/1511.06309

[39] J. Tu, H. Liu, F. Meng, M. Liu, and R. Ding, "Spatial-temporal data augmentation based on LSTM autoencoder network for skeleton-based human action recognition," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 3478–3482.

[40] E. Marchi, F. Vesperini, F. Eyben, S. Squartini, and B. Schuller, "A novel approach for automatic acoustic novelty detection using a denoising autoencoder with bidirectional LSTM neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2015, pp. 1996–2000.

[41] F. Zhao, J. Feng, J. Zhao, W. Yang, and S. Yan, "Robust LSTM-autoencoders for face de-occlusion in the wild," *IEEE Trans. Image Process.*, vol. 27, no. 2, pp. 778–790, Feb. 2018.

[42] J. Li, M.-T. Luong, and D. Jurafsky, "A hierarchical neural autoencoder for paragraphs and documents," Jun. 2015, *arXiv:1506.01057*. [Online]. Available: https://arxiv.org/abs/1506.01057

[43] H. Liu, J. Zhou, Y. Zheng, W. Jiang, and Y. Zhang, "Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders," *ISA Trans.*, vol. 77, pp. 167–178, Jun. 2018.

[44] T. Zhao, R. Zhao, and M. Eskenazi, "Learning discourse-level diversity for neural dialog models using conditional variational autoencoders," Mar. 2017, *arXiv:1703.10960*. [Online]. Available: https://arxiv.org/abs/1703.10960

[45] J. Cowton, "A combined deep learning gru-autoencoder for the early detection of respiratory disease in pigs using multiple environmental sensors," *Sensors*, vol. 18, no. 8, p. 2521, 2018.

[46] T.-N. Nguyen and J. Meunier. *Walking Gait Dataset: Point Clouds, Skeletons and Silhouettes*. Accessed: Apr. 2018. [Online]. Available: http://www-labs.iro.umontreal.ca/~labimage/GaitDataset/

[47] Z. Wang, B. Du, and Y. Guo, "Domain adaptation with neural embedding matching," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published.

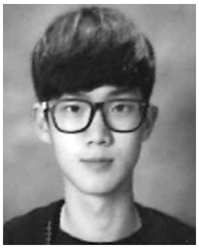
[48] L. Zhang, Q. Zhang, B. Du, X. Huang, Y. Y. Tang, and D. Tao, "Simultaneous spectral-spatial feature selection and extraction for hyperspectral images," *IEEE Trans. Cybern.*, vol. 48, no. 1, pp. 16–28, Jan. 2018.

[49] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," Dec. 2014, *arXiv:1412.6980*. [Online]. Available: https://arxiv.org/abs/1412.6980

KOOKSUNG JUN received the B.S. degree in mechanical engineering and the M.S. degree in intelligent robotics from the Gwangju Institute of Science and Technology, Gwangju, South Korea, in 2018 and 2019, respectively, where he is currently pursuing the Ph.D. degree. His current research interests include healthcare robotics, artificial intelligence, and pattern recognition.

DEOK-WON LEE received the B.S. degree in electronics and electric wave and information engineering from Chungnam National University and the M.S. degree in electrical and electronics engineering from Yonsei University, in 2009 and 2013, respectively. He is currently pursuing the Ph.D. degree with the Gwangju Institute of Science and Technology. His current research interests include healthcare robotics, artificial intelligence, and pattern recognition.

KYOOBIN LEE received the B.S., M.S., and Ph.D. degrees in mechanical engineering from the Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 1998, 2000, and 2008, respectively. From 2012 to 2017, he was the Principal Researcher with the Samsung Advanced Institute of Technology. He is currently an Assistant Professor with the School of Integrated Technology, Gwangju Institute of Science and Technology. His current research interests include bio-signal deep learning, pattern recognition, and artificial intelligence.



SANGHYUB LEE received the B.S. degree in biomedical engineering from the University of Ulsan and the M.S. degree in intelligent robotics from the Gwangju Institute of Science and Technology, in 2017 and 2019, respectively, where he is currently pursuing the Ph.D. degree. His current research interests include image processing, healthcare robotics, and pattern recognition.

MUN SANG KIM received the B.S. and M.S. degrees in mechanical engineering from Seoul National University, Seoul, South Korea, in 1980 and 1982, respectively, and the Dr. Ing. degree in robotics from the Technical University of Berlin, Berlin, Germany, in 1987. From 1987 to 2016, he was a Research Scientist with the Korea Institute of Science and Technology, Seoul, South Korea. He led the Advanced Robotics Research Center, in 2000, and became the Director of the "Intelligent Robot—The Frontier 21 Program," in October 2003, which is one of the most challenging research programs in South Korea. He is currently a Professor with the School of Integrated Technology, Gwangju Institute of Science and Technology. His current research interests include healthcare robotics, UWB-based indoor localization systems, and culture technology.