International Journal of Engineering

Journal Homepage: www.ije.ir

Fuzzy Centralized Coordinate Learning and Hybrid Loss for Human Activity Recognition

M. Bourjandi, M. Yadollahzadeh Tabari*, M. Golsorkhtabaramiri

Department of Computer Engineering, Babol Branch, Islamic Azad University, Babol, Iran

P A P E R I N F O

Paper history:
Received 05 August 2021
Received in rivised form 30 August 2021
Accepted 29 September 2021

Keywords:
Human Activity Recognition
Deep Learning
Fuzzy Centralized Coordinate Learning
Hybrid Loss Function

A B S T R A C T

Human activity recognition has been a popular research topic in recent years. The rapid development of deep learning techniques has greatly helped researchers to achieve success in this field. But the researchers usually over look the distribution of features in the coordinate space despite its significant effect on the convergence status of network and classification of activities. This paper proposes a combined method based on fuzzy centralized coordinate learning (FCCL) and a hybrid loss function to overcome the explained constraint. The FCCL induces features to be dispersedly spanned across all quadrants of the coordinate space. For this reason, the angle between the feature vectors of the activity classes increases significantly. Furthermore, a hybrid loss function is presented to increase the discriminative power of the proposed method. Our experiments were carried out on the opportunity and the PAMAP2 datasets. The proposed method has been compared with six machine learning and three deep learning methods for activity recognition. Experimental results showed that the proposed method outperformed all of the comparative methods due to identifying discriminative features. The proposed method successfully enhanced the average accuracy by 17.01% and 3.96% on the PAMAP2 and opportunity datasets, respectively, compared to the deep learning methods.

***doi**: 10.5829/ije.2022.35.01a.12*

NOMENCLATURE

P_k	predicted posterior probability(*k*-th class)	$S_o(x_1.x_2)$	cosine similarity of the two feature vectors
μ_u	mean of u	$\theta_{y_i,i}$	intersection angle between $\frac{w}{\|\|w\|\|}$ and $\vartheta(x)$
σ_u^2	variance of u	β	Balance parameter
$\Gamma(\,)$	gamma function	τ	adaptive parameter
D	degree of freedom	$L_{softmax}$	Soft-Max loss function
γ_1 , γ_2	fuzzy decay factors	L_{SAAM}	Simple adaptive angular margine loss function
o_b	mean vector produced by the current mini batch	ΔO	difference between the old mean vector and o_b
σ_b	standard deviation vector produced by the current mini batch	$\Delta\sigma$	difference between the old standard deviation vector and σ_b

1. INTRODUCTION

Human activity recognition (HAR) is a field of study to identify and analyze the activities performed by a person (or persons). Today, the HAR methods are widely used in various areas, including the healthcare [1-2], smart cities [3], and affordable mobile devices [4]. Machine learning algorithms and deep learning methods are the most common and popular methods widely used in HAR. HAR still faces two issues: 1) How to extract discriminative features from raw data, and 2) How to apply discriminative loss functions in feature classification.

The top well-known methods in the machine learning algorithms include the k-Nearest Neighbor (k-NN) algorithm [5], Artificial Neural Networks (ANN), Support Vector Machine (SVM) [6], the Random Forest (RF) [7], Decision Tree (DT), and Naive Bayes (NB) [8].

* Corresponding Author Institutional Email: *m_tabari@baboliau.ac.ir* (M. Yadollahzadeh Tabari)

Please cite this article as: M. Bourjandi, M. Yadollahzadeh Tabari, M. Golsorkhtabaramiri, Fuzzy Centralized Coordinate Learning and Hybrid Loss for Human Activity Recognition, ***International Journal of Engineering, Transactions A: Basics***, Vol. 35, No. 01, (2022) 130-141

Bustoni *et al.* [9] compared the recognition performance of SVM, k-NN, and Random Forest machine learning algorithms on HAR. Their results showed that the highest accuracy and recall were achieved by the SVM method with Support Vector Classifier (SVC) and Radial Basis Function (RBF) kernels, which are 87 and 85%, respectively. The drawback of these methods is that they use hand-crafted features that rely heavily on human experience or knowledge.

In contrast to machine learning algorithms with shallow statistical features, deep learning methods include high-level and meaningful features and have achieved good performance in HAR [10-14]. Various deep learning-based techniques, such as AutoEncoders (AEs), convolutional neural networks (CNNs), recurrent neural networks (RNNs), have successfully been used in HAR [1, 11-12]. Panwar et al. [15] designed a CNN model to recognize three fundamental movements of the human forearm using a single wrist-worn accelerometer sensor. Their results showed that the CNN model outperformed SVM and K-means because it automatically extracts the high-level features. The limitation of the CNN model is that it ignores the temporal dependencies within the data [11, 16].

In recent articles, researchers have extensively employed hybrid networks that benefit from the advantages of different networks [1]. Javier et al. [17] proposed a hybrid method for HAR based on convolutional and Long Short-Term Memory recurrent (LSTM) networks called Deep ConvLSTM. The Deep ConvLSTM method consists of convolutional, recurrent, and Soft-Max layers. This method achieved an F1-score 69% performance using only signals obtained from accelerometers on the OPPORTUNITY dataset. This value was improved by 20% using signals acquired from accelerometers, gyroscopes, and magnetic sensors. The low efficiency of this method is due to the Soft-Max Loss which cannot obtain discriminative features for activity recognition [18]. In litertaure [19], a hybrid method, called CNN-LSTM-ELM, was proposed, which used ELM in the last layer for classification purposes. This method achieved an F1-score of 90.8% on the OPPORTUNITY dataset. The limitation of this method is the high sensitivity to hidden nodes [20].

Recently, the combination of the margin-based loss with deep learning methods has led to acquiring discriminative features and successful results in face recognition [21-22]. Zhang et al. [23] proposed a new learning method called centralized coordinate learning (CCL) to identify differential features in facial recognition. They conducted experiments on six datasets: LFW [24], CACD [25], SLLFW [26], CALFW [27], YouTube Face [28], and MegaFace [29]. The accuracy level reached 99.4% accuracy on the LFW dataset.

A combined method is presented in this paper to solve the limits mentioned in the recent HAR methods. This combined method is based on fuzzy centralized coordinate learning (FCCL) and a hybrid loss function. The proposed FCCL method dispersedly spans the features in the coordinate space to increase the angle between different activity classes. The hybrid loss function classifies the obtained discriminative features with high accuracy, and its contributions are as follows:

1. we present the FCCL to learn the discriminative features, which increase intra-class compactness and inter-class diversity in activity classes.
2. Using a new loss function called the hybrid loss for enhancing the separability of different activity classes.

This article is organized as follows: in Section 2, the architecture of the proposed method has been expressed for the accurate identification of human activity. Two used datasets, the performance metrics, and experimental settings are introduced in Section 3. Finally, in Section 4, the experiments and results are discussed.

2. PROPOSED METHOD

This research has attempted to overcome the limitations mentioned in the previous section by presenting a combined method based on the FCCL and a hybrid loss function. Figure 1 presents a workflow of the proposed method. According to this figure, the first step in the proposed method is to receive raw data from wearable sensors. Many HAR methods [10-11, 30] have used statistical features such as symbolic representation [31], statistics of raw data, and transform coding [32]; however, they have overlooked the short-term and long-term temporal dependencies between the features. To solve this problem, we use a combination of the CNN and LSTM networks to extract the features, shown in Figure 2. In the second step, we presented the FCCL to learn discriminative features which improve intra-class compactness and inter-class diversity. Finally, according to Figure 2, a hybrid loss function is proposed to improve the separability of different activity classes. In the following, the proposed method is described in detail.

2. 1. Feature Extraction Using the Hybrid Deep Learning Method An essential step in HAR is to extract high-level features from time-series data. Deep neural networks (DNNs) can be proposed to extract meaningful features. Most researchers use CNN to recognize human activities with wearable sensors [33]. The main layers used in deep convolution networks are the convolution, the max-pooling, and the fully connected layers. In the convolution layer, filters are used to represent an abstract of the input data [34]. CNNs automatically learn the local and short-term features of time series data but ignore temporal dependencies within data [11]. The LSTM is a type of DNNs that extracts temporal dependencies within data and widely combines with CCNs. The advantage of LSTM is the ability to learn

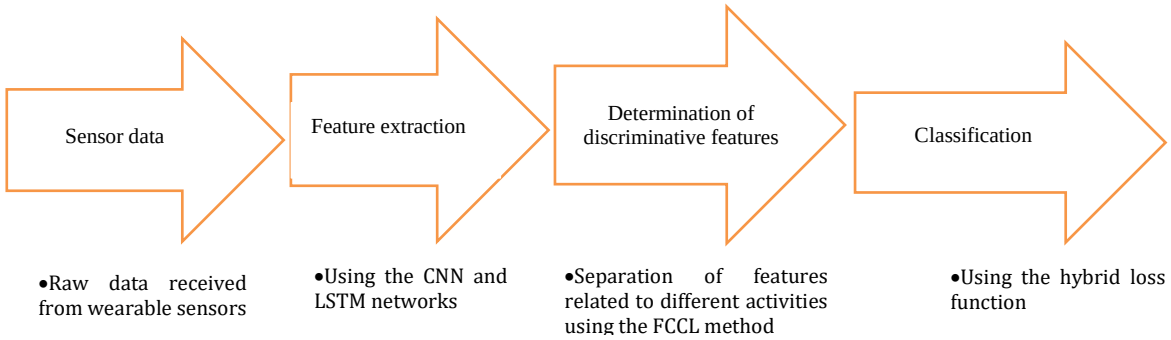


Figure 1. Basic block diagram of the work in this research

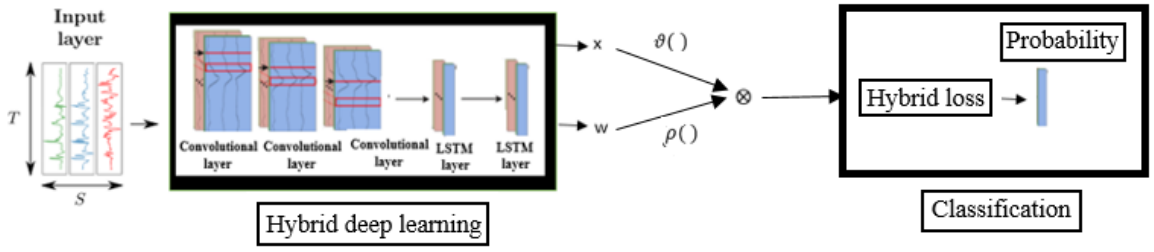


Figure 2. The architecture of the proposed method; x and w are the feature and classification vectors, respectively. $\rho(\,)$ and $\vartheta(\,)$ are transformation functions on parameters w and x, respectively

long-term dependency, which is not possible by recurrent neural networks [35]. The present study uses a hybrid deep learning method based on CNN and LSTM networks to take advantage of both, in which the output of the last convolutional layer feeds into the LSTM layer. According to Figure 2, three convolutional layers and two LSTM layers are used for feature extraction.

2. 2. Fuzzy Centrolized Coordinate Learning

Soft-Max loss [18] is a standard multiclass classification loss function in DNN. Soft-Max loss projects an input feature into a probability distribution [39]. The Soft-Max loss is defined as:

$$P_k = \frac{\exp(Z_k)}{\sum_{j=1}^{K} \exp(Z_j)} \quad (1)$$

$$Z_j = w_j^T x + b_j \quad (2)$$

where P_k is the predicted posterior probability for the k-th class, x is the extracted feature at the last layer of the hybrid deep learning method, w_j and b_j are classification vector, and the bias belonging to j-th class, respectively. According to Equation (2), the features x and the classification vectors w_j are parameters learned in the training step. Proper formulation of features x and the classification vectors w greatly affects the network convergence [23]. An appropriate formulation of w reduces the classification gap between the training and test steps. The efficient formulation of x can increase the angle between feature vectors related to different classes. For that reason, two functions $\rho(\,)$ and $\vartheta(\,)$ on parameters w and x, respectively, are used during the CCL for more effective network training [23]. In addition, we set b_j to 0 for simplicity. With considering the two functions $\rho(\,)$ and $\vartheta(\,)$, Equation (2) is rewritten as follows:

$$Z = \rho(w)^T \vartheta(x) \quad (3)$$

Based on the previous research [23], $\rho(w)$ is set to $\frac{w}{||w||}$. According to Equations (1) and (2), $\vartheta(x)$ will significantly affect the predicted posterior probability of a class (P_k) and the final output of the deep network. If $\| \vartheta(x)\|$ is small, the predicted posterior probability of all samples will be similar so that the loss function will be less discriminative between activities of different classes. Once $\| \vartheta(x)\|$ is large, the probabilities may vary much and cause instability in the learning of the deep network. Hence, this function should be formulated to separate the feature vectors of different classes by large angles. Therefore, in this article, the extracted features are spanned across all quadrants of the coordinate space using the FCCL method. For each dimension i of the feature vector x, $\vartheta(x_i)$ is defined as follows [23]:

$$\vartheta(x_i) = \frac{x(i)-o(i)}{\sigma(i)} \quad (4)$$

where $o = E[x]$ is the mean vector of x, and $\sigma(i)$ is the standard deviation of $x(i)$. To simplify the analysis, we use the L_2 norm of $\vartheta(x_i)$, i.e., $||\vartheta(x)||$.

$$u = ||\vartheta(x)|| = \sqrt{\sum_{j=1}^{D}\left(\vartheta(x(j))\right)^2} \quad (5)$$

The mean and variance of u are formulated as follows:

$$\mu_u = E[u] = \sqrt{2}\,\frac{\Gamma(\frac{D+1}{2})}{\Gamma(\frac{D}{2})} \quad (6)$$

$$\sigma_u^2 = D - \mu_u^2 \quad (7)$$

The $\Gamma(\,)$ is a Gamma function and the values of μ_u and σ_u^2 are determined according to the degree of freedom D. According to the experiments performed by Zhang et al., the best rational choice for D is 374 [23]. The origin vector (o) and the standard deviation (σ), which are two key factors in the function ϑ(x), are learned during deep neural network training. In the following, an example is given to illustrate how o affects the function $\vartheta(x)$. The cosine similarity of the two feature vectors x_1 and x_2 is formulated in Equation (8), by taking o as the coordinate origin.

$$S_o(x_1.x_2) = \frac{(x_1-o)^T(x_2-o)}{||x_1-o||\,||x_2-o||} \quad (8)$$

According to Equation (8), o affects the similarity between x_1 and x_2. For example, if x_1= [0.1, 0.1, 0.1], x_2= [0.1, 0.2, 0.1], and o = [0, 0, 0], the intersection angle between x_1 and x_2 is 19.6° and $S_o(x_1, x_2)$ = 0.942. If the origin vector shifts to [-0.03, -0.03, -0.03], then the intersection angle is 16.26 ° and $S_o(x_1.x_2)$ = 0.96. According to the example above, if the origin vector o moves a little, the angle and similarity between the feature vectors are greatly affected. Therefore, the values of o and σ in each update should not change much during DNN training. For that reason, in this article, the fuzzy decay factors of γ_1 and γ_2 are proposed to determine the appropriate balance between the new and old values of o and σ, respectively. The parameters σ and o are updated with decay factors as follows:

$$o_{new} = \gamma_1\, o_{old} + (1-\gamma_1)o_b \quad (9)$$

$$\sigma_{new} = \gamma_2\, \sigma_{old} + (1-\gamma_2)\sigma_b \quad (10)$$

where o_b and σ_b are the mean and standard deviation vectors produced by the current mini batch. In the previously conducted research [23], γ_1 and γ_2 were set to 0.99, but in the present article, the exact value of these variables is determined in a fuzzy way. For this purpose, we used the variables ΔO and $\Delta\sigma$. Equations (11) and (12) show the values of ΔO and $\Delta\sigma$.

$$\Delta O = ||o_{old} - o_b|| \quad (11)$$

$$\Delta\sigma = ||\sigma_{old} - \sigma_b|| \quad (12)$$

In the following, a fuzzy system is presented to determine the fuzzy value of γ_1 based on the variable ΔO. In a similar way, the value of γ_2 in Equation (10) is determined by a fuzzy system based on variable $\triangle\sigma$. Figure 3 shows the structure of the fuzzy system to find the appropriate value of γ_1 . As can be seen in this figure, the input of this system is the fuzzy values corresponding to the variable ΔO. The fuzzy sets of input variable ΔO are illustrated in Figure 4. These fuzzy sets are based on the Gaussian membership function.

The fuzzy sets of input and the competitive fuzzy rules in Table 1 are obtained by performing different experiments on the PAMAP2 and OPPORTUNITY datasets. As shown in Figure 4, we use five fuzzy sets for the variable ΔO. The value of γ_1 is set based on the membership value of ΔO to each of the fuzzy sets. When maximum membership value of ΔO is related to very low or low fuzzy sets, we can increase the effectiveness of o_b more reliably for updating parameter o_{new}. For this reason, γ_1 is set by a small value. The range of γ_1 is [0,0.55] or(0.55,0.67) according to $\mu_{very\ low}(\Delta O)$ in rule1 or $\mu_{low}(\Delta O)$ in rules 2 and 3, respectively. Similarly, when maximum membership value of ΔO is related to very high or high fuzzy sets, the fuzzy value of γ_1 is increased to prevent creating an unstable state in the network and increase the effect of the parametero_{old} . Therefore, the range of γ_1 is [0.82,0.96) or 0.99 according to $\mu_{high}(\Delta O)$ in rules 6 and 7 or $\mu_{very\ high}(\Delta O)$ in rule 8, respectively. The value of γ_2 in Equation (10) is determined similarly based on the fuzzy

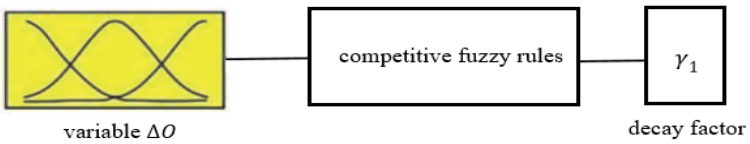


Figure 3. The structure of the fuzzy system to find the appropriate value of γ_1

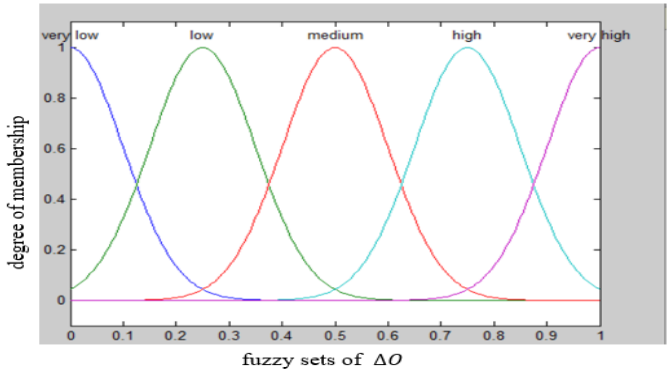


Figure 4. The fuzzy sets of input variable ΔO

TABLE 1. Competitive fuzzy rules

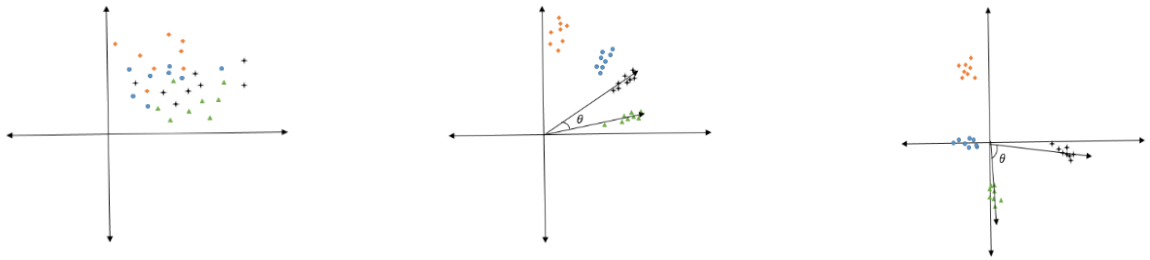
1-	If $(0.45 \leq \mu_{very\ low}(\Delta O) \leq 1)$ then $(\gamma_1 = 1 - \mu_{very\ low}(\Delta O))$
2-	If $(0.45 < \mu_{low}(\Delta O) \leq 1)$ and $(\mu_{very\ low}(\Delta O) < 0.45)$ then $(\gamma_1 = 0.51 + \frac{\mu_{low}(\Delta O)}{10})$
3-	If $(0.45 < \mu_{low}(\Delta O) < 1)$ and $(\mu_{medium}(\Delta O) < 0.45)$ then $(\gamma_1 = 0.61 + \frac{1-\mu_{low}(\Delta O)}{10})$
4-	If $(0.45 \leq \mu_{medium}(\Delta O) \leq 1)$ and $(\mu_{low}(\Delta O) < 0.45)$then $(\gamma_1 = 0.6 + \frac{\mu_{medium}(\Delta O)}{7})$
5-	If $(0.45 < \mu_{medium}(\Delta O) < 1)$ and $(\mu_{high}(\Delta O) < 0.45)$then $(\gamma_1 = 0.74 + \frac{1-\mu_{medium}(\Delta O)}{7})$
6-	If $(0.45 \leq \mu_{high}(\Delta O) < 1)$ and $(\mu_{medium}(\Delta O) < 0.45)$then $(\gamma_1 = 0.73 + \frac{\mu_{high}(\Delta O)}{5})$
7-	If $(0.8 < \mu_{high}(\Delta O) \leq 1)$ and $(\mu_{very\ high}(\Delta O) < 0.2)$then $(\gamma_1 = 0.91 + \frac{1-\mu_{high}(\Delta O)}{5})$
8-	If $(0.2 \leq \mu_{veryhigh}(\Delta O))$ then $(\gamma_1 = 0.99)$

sets of input variable $\triangle\sigma$. Therefore, these fuzzy values (γ_1 and γ_2) effectively determine variables o_{new} and σ_{new} at each step.

2. 3. Hybrid Loss Various classification functions are used to determine the final output in the last layer of deep neural networks. In literature, different classification functions have been proposed, but all of them have some drawbacks. To provide a more compact classification boundary for accurate user activities identification for HAR, we presented a combination of Soft-Max loss and Simple Adaptive Angular Margin (SAAM) loss, called hybrid loss. According to the definition of cosine similarity, the Soft-Max and SAAM loss functions are formulated as Equations (13) and (14), respectively, as follows:

$$L_{Soft-max} = \sum_i^N -\log\left(\frac{\exp\left(||\vartheta(x_i)||\cos\left(\theta_{y_i,i}\right)\right)}{\sum_{k=1}^{K}\exp\left(||\vartheta(x_i)||\cos(\theta_{k,i})\right)}\right) \tag{13}$$

$$L_{SAAM} = \sum_i^N -\log\left(\frac{\exp\left(||\vartheta(x_i)||\cos\left(\tau\theta_{y_i,i}\right)\right)}{\exp\left(||\vartheta(x_i)||\cos\left(\tau\theta_{y_i,i}\right)\right)+\sum_{k\neq y_i}\exp\left(||\vartheta(x_i)||\cos(\theta_{k,i})\right)}\right) \tag{14}$$



(a) original data distribution (b) Soft-Max loss (c) Hybrid loss

Figure 5. Illustration of the effects of various loss functions. The "orange rhombuses", "blue dots", "black stars" and "green triangles" represent the samples of four activities classes. (a) Original data distribution. (b) Converged features by using the Soft-Max loss. (c) Converged features by using the hybrid loss

where $\theta_{y_i.i}$ is the intersection angle between $\frac{w}{||w||}$ and $\vartheta(x)$ and range of $\theta_{y_i.i}$ is $[0.\pi]$. The adaptive parameter τ is set based on the range of $\theta_{y_i.i}$ [23].

According to Equations (13) and (14), $\left|\left|\vartheta(x_i)\right|\right|$ and $\cos(\theta_{y_i.i})$ will affect the loss functions and, both depend on the form of $\vartheta(x_i)$. In Sub-section 2.2, we formulated $\vartheta(x)$ well. The hybrid loss function is shown in Equation (15) based on $L_{Soft-max}$ and L_{SAAM}.

$$L_{hybrid\ loss} = \frac{\beta\ L_{softmax} + L_{SAAM}}{\beta + 1} \quad (15)$$

Parameter β is a value to balance between the Soft-Max loss and the SAAM loss. The value of β is determined based on the value of $\theta_{y_i.i}$ as follows:

$$\beta = \begin{cases} 3. & \frac{\pi}{3} < \theta_{y_i.i} \leq \pi; \\ 2. & \frac{\pi}{30} < \theta_{y_i.i} \leq \frac{\pi}{3}; \\ 2 \times \cos(\theta_{y_i.i}). & \theta_{y_i.i} \leq \pi/30 \end{cases} \quad (16)$$

According to Equation (16), the range of $\theta_{y_i.i}$ is partition into three distances, based on which β is set. When $\frac{\pi}{3} < \theta_{y_i.i} \leq \pi$, the angle is big enough to supply enough gradient information for training. thus, the Soft-Max loss function can classify the activities accurately, and we fix the parameter β =3. When $\frac{\pi}{30} < \theta_{y_i.i} \leq \frac{\pi}{3}$, the ability of Soft-Max loss to extract discriminative features decreases, so we try to reduce its effect on the proposed hybrid loss, and β is set to 2. As the angle decreases, the SAAM loss function can enhance the separability of neighboring classes by using parameter τ. Thus, when $\theta_{y_i.i} \leq \pi/30$, the effect of Soft-Max loss decreases slowly based on $\theta_{y_i.i}$, and we set β =2 × cos $(\theta_{y_i.i})$. In Figure 5, we show the differences of the two loss functions on the classification of features, including the Soft-Max loss in Equation (13) and the hybrid loss in Equation (15). According to Figure 5(b), the centers of activity classes are placed in a unit hypersphere and are very close to each other. Therefore, the angle θ between the features of two different classes is small. The hybrid loss causes the extracted features to span across all quadrants of the coordinate space. As shown in Figure 5(c), the dispersed distribution of features increases the angle between neighboring classes and improves the separability.

3. EXPERIMENTAL EVALUATION

Several experiments were performed to determine the effectiveness of the proposed method. To this end, two benchmark datasets were used, which contain the human activity of daily living. These data were recorded using a combination of environmental sensors and body-connected sensors in the environment. The experiments were performed using the Keras framework in Python 3 on a system with Windows 10. The system uses a quad-core processor with 2.30 GHz speed and a GeForce GTX 950M graphics card. Section 3.1 provides a brief overview of the datasets used for the evaluation, and Section 3.2 describes the evaluation method. Finally, the settings of the deep learning methods are presented in Section 3.3.

3. 1. Datasets The proposed method was evaluated on two benchmark datasets widely used in literature. These datasets include data streams received from the sensors embedded in different positions of the participants' bodies. The activity recognition datasets usually include various activities such as walking, cycling, and goal-oriented activities. In the experiments performed in this article, the OPPORTUNITY [36] and PAMAP2 [37] datasets were used. The OPPORTUNITY dataset has an unbalanced class distribution such that most of the samples belong to the NULL class. In contrast, the PAMAP2 dataset contains a balanced distribution of human activities. These datasets are available at the UCI Machine learning repository [2].

• OPPORTUNITY

The OPPORTUNITY dataset [36] includes annotated recordings from four subjects performing 17 different daily activities in a kitchen scenario. A NULL class also exists that is not associated with any of the daily activities. The data were collected using body-worn sensors. These sensors contain five commercial RS485-networked XSense inertial measurement units, 12 Bluetooth acceleration sensors, and commercial InertiaCube3 inertial sensors. They were placed on a custom-made motion jacket, the limbs, and each foot, respectively. These sensors perform sampling with a frequency of 30 Hz. During the recordings, for each subject, six different runs were recorded. Five runs were termed Activity of Daily Life (ADL), and one run was termed drill run (Drill). In ADL1 to ADL5, the subjects performed all of the different scripted activities, and in the Drill, they replicated each activity 20 times. To deal with the problem of missing data, 38 sensor channels (including accelerometer recordings) were removed and the remaining 107-dimensional data were used for our experiments. We selected the ADL1-3 and Drill runs as the training set and the ADL4-5 runs as the testing set. For the segmentation phase, a fixed-length sliding window was applied to slicing the data. Based on the previous research, a sliding window of 2 s and a sliding stride of 3 were used in this article [1].

• PAMAP2

The PAMAP2 Physical Activity Monitoring dataset consists of 12 daily living activities, i.e., lie, sit, stand,

[2] http://archive.ics.uci.edu/ml

walk, run, cycle, Nordic walk, iron, vacuum, jump rope, ascend, and descend stairs, which are performed by nine subjects. The data were recorded by three inertial measurement units (accelerometer, gyroscope, and magnetometer) and a heart rate monitor attached to the participant's hands, chest, and ankles [37]. These sensors performed sampling with a frequency of 33 Hz. Data were collected in more than 10 hours, and the obtained dataset had 52 dimensions. For the segmentation phase, sliding windows of 5.12 s were used with 1 s stepping between adjacent windows (78% overlap). Runs 1 and 2 for subject5 were used in the validation set and runs 1 and 2 for subject6 in the test set. The remaining data were used in the train set.

3. 2. Model Evaluation It is now necessary to examine the effectiveness of the proposed method in human activity recognition. Since the recognition rate of the majority class greatly affects the performance statistics in the minority classes and the OPPORTUNITY dataset is highly unbalanced (the NULL class includes more than 75% of the total data). Thus, in addition to the accuracy criterion, the weighted F1-score and the average F1-score were used, which were independent of the class distribution. The weighted F1-score is formulated by Equation (17) as follows:

$$\text{weighted F1} - \text{score } (\text{F1}_{\text{W}}) = \sum_i 2 \times w_i \frac{precision_i \times recall_i}{precision_i + recall_i} \quad (17)$$

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

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

$$w_i = {m_i}/{M} \quad (20)$$

where TP refers to the number of samples correctly recognized in the positive class, FP refers to the number of samples incorrectly identified in the negative class. TN refers to the number of samples correctly recognized in the negative class, and FN refers to the number of samples incorrectly identified in the positive class. In Equation (20), parameter i is the class index, w_i is the ratio of data in class i, m_i is the number of samples in class i, and M is the total number of samples considered. The average F1-score is not dependent on the class distribution and is formulated by Equation (21) as follows:

$$\text{average F1} - \text{score } (\text{F1}_{\text{A}}) = \frac{2}{m_i} \sum_i \frac{precision_i \times recall_i}{precision_i + recall_i} \quad (21)$$

TABLE 2. Settings of hyperparameters related to the deep learning methods

Model	Layer	Parameter	Value	Classifier
CNN-LSTM-ELM [19]	2-5	Convolutional(Kernel Size)	(5,1)	ELM
		Convolutional (sliding stride)	(1,1)	
		Convolutional(Kernels)	50^2,40^3,30^4,40^5	
	6-7	LSTM (number of neurons)	128	
Hybrid [1]	2	Convolutional(Kernel Size)	(11,1)	Soft-Max
		Convolutional (sliding stride)	(1,1)	
		Convolutional(Kernels)	50	
		Convolutional (Pooling Size)	(2,1)	
	3-4	LSTM(cells)	27	
		Output of LSTM cells	600	
	5	Fully connected	512	
Deep ConvLSTM [17]	2-5	Convolutional(Kernel Size)	(5,1)	Soft-Max
		Convolutional (sliding stride)	(1,1)	
		Convolutional(Kernels)	64	
	6	LSTM (number of neurons)	128	
	7	Fully connected	n_c	
Proposed method	2-4	Convolutional(Kernel Size)	(11,1)	Hybrid Loss
		Convolutional (sliding stride)	(1,1)	
		Convolutional(Kernels)	60	
	5-6	LSTM(cells)	27	
		Output of LSTM cells	600	

3. 3. Model Setting Table 2 shows the settings of the hyperparameters related to the four deep learning methods used in this article. The network parameters are optimized by minimizing the cross-entropy loss function. This optimization is done using mini-batch gradient descent with the RMSProp update rule. In Table 2, the fully connected layers were followed by a Rectified Linear Unit (ReLU) activation layer to prepare the non-linear presentation. Each of the convolution blocks in Table 2 includes convolution layers and ReLU, and in the hybrid method, in addition to them, it consists of the max-pooling layer. The size of the sliding window defined the number of cells in the LSTM layers. In the LSTM cells, sigmoid and hyperbolic functions are used for gate and other activations, respectively. After the last fully connected layer is used, a classifier layer is employed to provide predictions for each class. All parameters of the deep learning methods were randomly initialized and trained using the ADADELTA optimizer [38] with default parameters (initial learning rate of 1) for 50 epochs. The batch size is set to 100. In this article, the sliding time window size (T) and the number of sensor channels (n) were fixed at 64 and 107, respectively.

4. EXPERIMENTAL RESULTS

In this section, the performance of the proposed method is evaluated and analyzed on two benchmark datasets. This section is organized as follows: the analysis of the results obtained from the proposed method and its comparison with other known methods are presented in Sub-section 4.1; a comparison between the proposed method and three machine learning algorithms is provided in Sub-section 4.2. Finally, the performance of the Soft-Max and hybrid loss functions are examined in Sub-section 4.3.

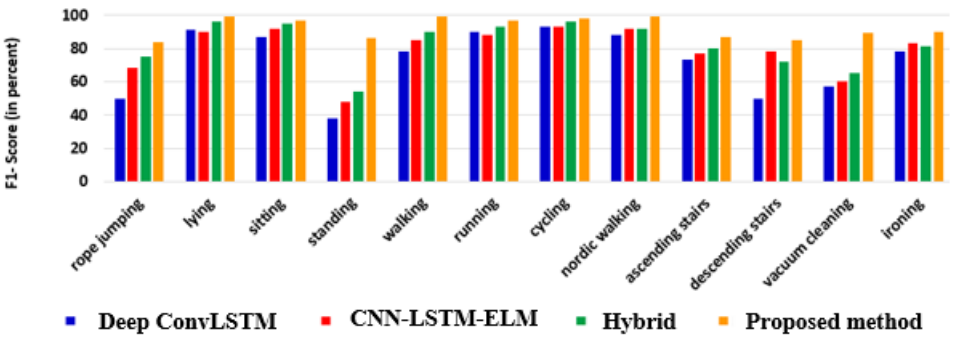
4. 1. Performance Comparison A comparison between the deep learning methods and the proposed method based on the performance parameters is shown in Table 3. The performance parameters include the accuracy (ACC), weighted F1-score ($F1_W$), and average F1-score ($F1_A$). According to Table 3, the proposed method has the highest score in terms of overall performance on the two defined datasets.

Specifically, the accuracy, weighted F1-score, and average F1-score of the proposed method increased to 93.13% and 93.28% and 78.12%, respectively, on the OPPORTUNITY dataset. The OPPORTUNITY data set is imbalanced [36]; this leads to insufficient training in all tested methods. Thus, the average F1- score has improved less compared to the other performance parameters. There is a significant difference between the average F1-score in CNN-LSTM-ELM [19] and hybrid method [1] with the Deep ConvLSTM method [17], which is due to the use of a suitable classifier in the last layer and the appropriate number of LSTM layers. One of the drawbacks of the CNN-LSTM-ELM method, which has led to its low efficiency compared to the proposed method, is its strong dependence on the number of hidden nodes. According to Table 3, the proposed method has increased the average accuracy by 6.75% compared to the hybrid method on the two datasets. This improvement is due to the extracted discriminative features, which significantly enhance intra-class compression and inter-class diversity. For a more detailed analysis, the results of different deep learning methods were compared on each class of the PAMAP2 dataset. Figure 6(a) illustrates the F1-score of each class for our proposed method and the deep learning methods. According to Figure 6(a), the lowest efficiency is attributed to the "standing" activity. The values of F1-score in "standing" activity with the Deep ConvLSTM, CNN-LSTM-ELM, hybrid, and proposed methods are 38%, 48%, 54%, and 86%, respectively. Thus, the value of the F1-score achieves an increase of 48% by the proposed method, compared with the Deep ConvLSTM. The highest difference of F1-score between the hybrid and proposed methods is related to the "standing" and "vacuum cleaning" activities. This significant increase in F1-score (related to the "standing" and "vacuum cleaning" activities) indicates the high ability of the proposed method to determine discriminative features because it is a key parameter in identifying activities. The proposed method reached an F1-score of 84% in "rope jumping"; thus, the F1-score increased by 9% compared with the hybrid method. Figure 6(b) illustrates the confusion matrixes of all deep learning methods on the PAMAP2 dataset. Confusion matrixes contain detailed information about the actual and predicted classifications conducted by the system; therefore, it determines the nature of the classification error. According to Figure 6(b), the number of classes correctly predicted by the proposed method was more than other deep methods. The Deep ConvLSTM method only recognized the activities of 3 classes correctly, and the most errors in the CNN-LSTM-ELM and hybrid methods were related to classes 4, 11, and 1.

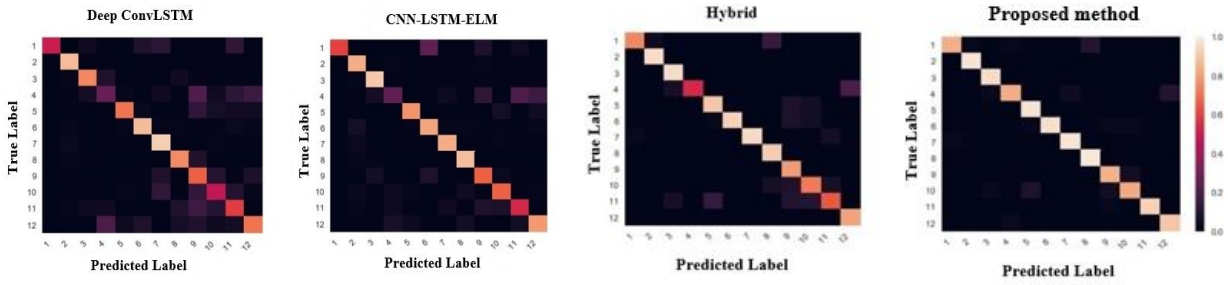
4. 2. Comparison with Machine Learning Algorithms In order to show the effectiveness of the proposed method, a comparison has been made between it and machine learning algorithms on the OPPORTUNITY dataset. Table 4 shows the results of this comparison. A total of 18 hand-crafted features were used, including 15 simple statistical values and three frequency values, which were calculated on each sensor channel independently [1]. In addition, we trained machine learning algorithms with extracted features from the first layer fully connected to the LSTM network.

TABLE 3. Classification performance results (in percent) of the various deep learning methods on the OPPORTUNITY and the PAMAP2 datasets

	OPPORTUNITY			PAMAP2		
Method	**ACC**	**$F1_W$**	**$F1_A$**	**ACC**	**$F1_W$**	**$F1_A$**
Deep ConvLSTM [17]	87.47	87.23	55.49	67.54	66.12	58.76
CNN-LSTM-ELM [19]	91.34	90.85	70.38	85	83.12	76
Hybrid [1]	91.76	91.56	70.86	85.12	83.73	76.10
Proposed method	**94.15**	**94.05**	**79.12**	**96.23**	**96.11**	**95.78**



(a)



(b)

Figure 6. (a) The F1-score of each class of various deep learning methods on the PAMAP2 dataset. (b) Confusion matrix of the deep learning methods on the PAMAP2 dataset: 1: rope jumping; 2: lying; 3: sitting; 4: standing; 5: walking; 6: running; 7: cycling; 8: Nordic walking; 9: ascending stairs; 10: descending stairs; 11: vacuum cleaning and 12: ironing.

TABLE 4. Classification performance results (in percent) of proposed method and three machine learning classifiers on the OPPORTUNITY dataset

Method	**ACC**	**$F1_W$**	**$F1_A$**
SVM	89.96	89.53	63.76
Random Forest	89.21	87.08	52.45
Naive Bayes	44.79	52.61	32.81
LSTM-SVM	91.81	91.62	70.24
LSTM- Random Forest	91.84	91.63	70.24
LSTM- Naive Bayes	91.15	91.29	69.03
Proposed method	**94.13**	**94.28**	**79.12**

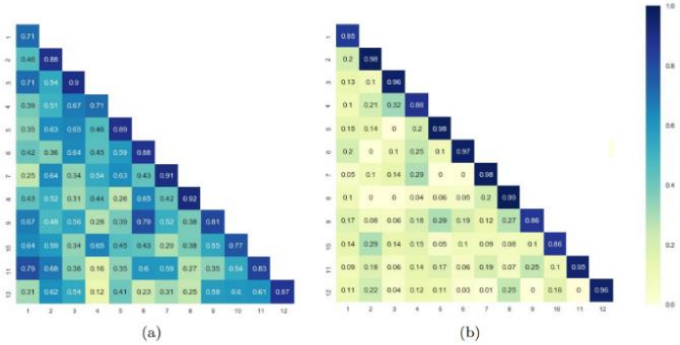


Figure 7. Heatmaps of cosine similarity of all classes using different loss functions on the PAMAP2 testing dataset: (a) Soft-Max loss; (b) hybrid loss

There was a significant efficiency improvement for the three machine learning algorithms, especially Naive Bayes. The highest efficiency was related to the proposed method; its average F1-score reached up to 79.12%. This high efficiency indicates the ability of the proposed method to improve the discrimination capability of extracted features.

4. 3. Comparison Between Soft-max and Hybrid Losses The effectiveness of the hybrid loss was assessed by comparing its performance with the Soft-Max loss function. This evaluation was done by calculating the cosine similarity (i.e., intra-class and inter-class). The heatmap of these similarities on the PAMAP2 dataset is presented in Figure 7. The proposed method, which has the highest performance on the PAMAP2 dataset, is used in this experiment. Features are extracted by the hybrid deep learning method for these two loss functions. Figure 7(a) and 7(b) show the heatmaps of cosine similarity related to the proposed method with Soft-Max and hybrid loss functions, respectively. As shown in Figure 7(a) the Soft-Max function resulted in the least intra-class similarity. Based on the results obtained, the hybrid loss can capture more discriminative features than the Soft-Max loss function. The hybrid loss effectively improved intra-class compactness and inter-class variety.

5. CONCLUSIONS

This research work applied the combined method based on the fuzzy centralized coordinate learning (FCCL) and hybrid loss function to deal with two well-known issues in sensor-based HAR. The first one is to extract discriminative features. For this reason, the extracted features are dispersed in the coordinate space by using the FCCL. In this case, intra-class diversity and inter-class similarity are significantly decreased. The second problem is the separability of different user activities by using the appropriate classifier. We have used the hybrid loss function as the classifier. Two benchmark datasets, OPPORTUNITY, and PAMAP2 were used to evaluate the performance and compare the proposed method with three deep learning methods, i.e., CNN-LSTM-ELM, Deep ConvLstm, and hybrid methods. The results showed that the proposed method outperformed all the three deep learning methods. In addition, a comparison was made between the proposed method and the three machine learning algorithms on the OPPORTUNITY dataset. The proposed method can improve the classification performance compared to the machine learning algorithms. Performance was significantly improved when the LSTM network was used in the machine learning algorithms for feature extraction. However, the proposed method could increase the discrimination capability of extracted features compared with machine learning algorithms. For future research, we aim to study the effectiveness of the proposed method on the open-set human activity recognition and the use of methods to enhance the sequential learning adaptive capability. A potential working direction could be using of a transfer learning approach (which reuses learned previous knowledge) to identify the activities carried out by various types of users in different environmental situations.

6. REFERENCES

1. Li, Frédéric, Kimiaki Shirahama, Muhammad Adeel Nisar, Lukas Köping, and Marcin Grzegorzek. "Comparison of feature learning methods for human activity recognition using wearable sensors." ***Sensors*** 18, No. 2, (2018), 679, DOI: 10.3390/s18020679 .
2. Ogbuabor, Godwin, and Robert La. "Human activity recognition for healthcare using smartphones." Proceedings of the 2018 10th International Conference on Machine Learning and Computing. (2018), https://doi.org/10.1145/3195106.3195157.
3. Khan, Adil Mehmood, Y-K. Lee, Seok-Yong Lee, and T-S. Kim. "Human activity recognition via an accelerometer-enabled-smartphone using kernel discriminant analysis." In 2010 5th international conference on future information technology, IEEE, (2010), 1-6. DOI: 10.1109/FUTURETECH.2010.5482729.

4. Shoaib, Muhammad, Stephan Bosch, Ozlem Durmaz Incel, Hans Scholten, and Paul JM Havinga. "A survey of online activity recognition using mobile phones." ***Sensors*** 15, No. 1, (2015), 2059-2085, https://doi.org/10.3390/s150102059.
5. Mobark, Mohammed, Suriayati Chuprat, and Teddy Mantoro. "Improving the accuracy of complex activities recognition using accelerometer-embedded mobile phone classifiers." In 2017 Second International Conference on Informatics and Computing (ICIC), IEEE, (2017), 1-5. DOI: 10.1109/IAC.2017.8280606.
6. Chen, Zhenghua, Qingchang Zhu, Yeng Chai Soh, and Le Zhang. "Robust human activity recognition using smartphone sensors via CT-PCA and online SVM." IEEE Transactions on Industrial Informatics 13, No. 6, (2017), 3070-3080, DOI: 10.1109/TII.2017.2712746.
7. Uddin, Md Taufeeq, Md Muttlaleb Billah, and Md Faisal Hossain. "Random forests based recognition of human activities and postural transitions on smartphone." In 2016 5th International Conference on Informatics, Electronics and Vision (ICIEV), pp. 250-255. IEEE, (2016), DOI: 10.1109/ICIEV.2016.7760005.
8. Fan, Liwei, Kim-Leng Poh, and Peng Zhou. "A sequential feature extraction approach for naïve bayes classification of microarray data." ***Expert Systems with Applications*** 36, No. 6, (2009), 9919-9923.
9. Bustoni, I. A., I. Hidayatulloh, A. M. Ningtyas, A. Purwaningsih, and S. N. Azhari. "Classification methods performance on human activity recognition." In Journal of Physics: Conference Series, vol. 1456, No. 1, p. 012027. IOP Publishing, (2020).
10. Georgiou, Theodoros, Yu Liu, Wei Chen, and Michael Lew. "A survey of traditional and deep learning-based feature descriptors for high dimensional data in computer vision." ***International Journal of Multimedia Information Retrieval*** 9, No. 3, (2020), 135-170.
11. Dargan, Shaveta, Munish Kumar, Maruthi Rohit Ayyagari, and Gulshan Kumar. "A survey of deep learning and its applications: a new paradigm to machine learning." ***Archives of Computational Methods in Engineering*** 27, No. 4, (2020), 1071-1092.
12. Feizi, A. "Convolutional gating network for object tracking." ***International Journal of Engineering, Transactions A: Basics*** 32, No. 7, (2019), 931-939, DOI: 10.5829/ije.2019.32.07a.05.
13. Hassanpour, M., and H. Malek. "Learning Document Image Features With SqueezeNet Convolutional Neural Network." ***International Journal of Engineering*** 33, No. 7, (2020), 1201-1207, DOI: 10.5829/ije.2020.33.07a.05.
14. Chikhaoui, Belkacem, and Frank Gouineau. "Towards automatic feature extraction for activity recognition from wearable sensors: a deep learning approach." In 2017 IEEE International Conference on Data Mining Workshops (ICDMW), 693-702. IEEE, (2017), DOI: 10.1109/ICDMW.2017.97.
15. Panwar, Madhuri, S. Ram Dyuthi, K. Chandra Prakash, Dwaipayan Biswas, Amit Acharyya, Koushik Maharatna, Arvind Gautam, and Ganesh R. Naik. "CNN based approach for activity recognition using a wrist-worn accelerometer." In 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2438-2441. IEEE, (2017), DOI: 10.1109/EMBC.2017.8037349.
16. Cruciani, Federico, Anastasios Vafeiadis, Chris Nugent, Ian Cleland, Paul McCullagh, Konstantinos Votis, Dimitrios Giakoumis, Dimitrios Tzovaras, Liming Chen, and Raouf Hamzaoui. "Feature learning for human activity recognition using convolutional neural networks." ***CCF Transactions on Pervasive Computing and Interaction*** 2, No. 1, (2020), 18-32.
17. Ordóñez, Francisco Javier, and Daniel Roggen. "Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition." ***Sensors*** 16, No. 1, (2016), 115, https://doi.org/10.3390/s16010115.
18. Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. "Softmax units for multinoulli output distributions. Deep Learning." (2018), 180-184.
19. Sun, Jian, Yongling Fu, Shengguang Li, Jie He, Cheng Xu, and Lin Tan. "Sequential human activity recognition based on deep convolutional network and extreme learning machine using wearable sensors." ***Journal of Sensors,*** (2018), https://doi.org/10.1155/2018/8580959.
20. Huang, Guang-Bin, Hongming Zhou, Xiaojian Ding, and Rui Zhang. "Extreme learning machine for regression and multiclass classification." ***IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*** 42, No. 2, (2011), 513-529, DOI: 10.1109/TSMCB.2011.2168604.
21. Liu, Weiyang, Yandong Wen, Zhiding Yu, and Meng Yang. "Large-margin softmax loss for convolutional neural networks." In International Conference on Machine Learning, vol. 2, No. 3, p. 7. (2016).
22. Liu, Weiyang, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song. "Sphereface: Deep hypersphere embedding for face recognition." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 212-220. (2017).
23. Qi, Xianbiao, and Lei Zhang. "Face recognition via centralized coordinate learning." arXiv preprint arXiv:1801.*05678* (2018).
24. Huang, Gary B., Marwan Mattar, Tamara Berg, and Eric Learned-Miller. "Labeled faces in the wild: A database forstudying face recognition in unconstrained environments." In Workshop on faces in'Real-Life'Images: detection, alignment, and recognition. (2008).
25. Chen, Bor-Chun, Chu-Song Chen, and Winston H. Hsu. "Cross-age reference coding for age-invariant face recognition and retrieval." In European conference on computer vision, pp. 768-783. Springer, Cham, (2014).
26. Deng, Weihong, Jiani Hu, Nanhai Zhang, Binghui Chen, and Jun Guo. "Fine-grained face verification: FGLFW database, baselines, and human-DCMN partnership." *Pattern Recognition* 66 (2017), 63-73, https://doi.org/10.1016/j.patcog.2016.11.023.
27. Zheng, Tianyue, Weihong Deng, and Jiani Hu. "Cross-age lfw: A database for studying cross-age face recognition in unconstrained environments." arXiv preprint arXiv:1708.08197 (2017).
28. Wolf, Lior, Tal Hassner, and Itay Maoz. "Face recognition in unconstrained videos with matched background similarity." In in Computer Vision and Pattern Recognition (CVPR) 2011, pp. 529-534. IEEE, (2011), DOI: 10.1109/CVPR.2011.5995566
29. Kemelmacher-Shlizerman, Ira, Steven M. Seitz, Daniel Miller, and Evan Brossard. "The megaface benchmark: 1 million faces for recognition at scale." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4873-4882. (2016).
30. Lara, Oscar D., and Miguel A. Labrador. "A survey on human activity recognition using wearable sensors." ***IEEE Communications Surveys & Tutorials*** 15, No. 3, (2012), 1192-1209, DOI: 10.1109/SURV.2012.110112.00192.
31. Lin J, Keogh E, Lonardi S, Chiu B. "A symbolic representation of time series, with implications for streaming algorithms". InProceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery 2003; 2-11, https://doi.org/10.1145/882082.882086.
32. Cook, Diane J., and Narayanan C. Krishnan. "Activity learning: discovering, recognizing, and predicting human behavior from sensor data". ***John Wiley & Sons***, (2015).
33. Lawal, Isah A., and Sophia Bano. "Deep human activity recognition with localisation of wearable sensors." ***IEEE Access*** 8, (2020), 155060-155070, DOI: 10.1109ACCESS.2020.3017681.
34. Zohrevand, A., Imani, Z. and Ezoji, M.. "Deep Convolutional Neural Network for Finger-knuckle-print Recognition". ***International Journal of Engineering, Transactions A: Basics***,

2021, Vol. 34, No. 7, 1684-1693, DOI: 10.5829/ije.2021.34.07a.12.

35. Gers, Felix A., Nicol N. Schraudolph, and Jürgen Schmidhuber. "Learning precise timing with LSTM recurrent networks." ***Journal of Machine Learning Research*** 3, (2002), 115-143.
36. Chavarriaga, Ricardo, Hesam Sagha, Alberto Calatroni, Sundara Tejaswi Digumarti, Gerhard Tröster, José del R. Millán, and Daniel Roggen. "The Opportunity challenge: A benchmark database for on-body sensor-based activity recognition**." *Pattern Recognition Letters*** 34, No. 15, (2013), 2033-2042, https://doi.org/10.1016/j.patrec.2012.12.014.
37. Reiss, Attila, and Didier Stricker. "Introducing a new benchmarked dataset for activity monitoring." In 2012 16th international symposium on wearable computers, IEEE, (2012), 108-109. DOI: 10.1109/ISWC.2012.13.
38. Zeiler, Matthew D. "Adadelta: an adaptive learning rate method." arXiv preprint arXiv:1212.5701 (2012).
39. Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems 25 (2012), 1097-1105.

Persian Abstract

چکیده

تشخیص فعالیتهای انسان، یکی از موضوعات تحقیقاتی رایج در سالهای اخیر بوده است .توسعه سریع تکنیک های یادگیری عمیق به محققان در دستیابی موفقیت در این زمینه بسیار کمک کرده است. اما محققان معمولاً از توزیع ویژگی ها در فضای مختصات با وجود تأثیر قابل توجه آن بر وضعیت همگرایی شبکه و طبقه بندی فعالیت ها چشم پوشی می کنند. این مقاله یک روش ترکیبی مبتنی بر یادگیری مختصات متمرکز فازی و تابع هزینه ترکیبی، برای غلبه بر محدودیت توضیح داده شده پیشنهاد می کند. یادگیری مختصات متمرکز فازی باعث می شود که ویژگی ها به صورت پراکنده در تمام چهاربخش از فضای مختصات پخش شوند. به همین دلیل، زاویه بین بردارهای ویژگی کلاسهای فعالیت به میزان قابل توجهی افزایش می یابد. علاوه بر این ، یک تابع هزینه ترکیبی برای افزایش قدرت تشخیص در روش پیشنهادی، ارائه شده است. آزمایشات ما بر روی مجموعه داده های OPPORTUNITY و PAMAP2 انجام شده است. روش پیشنهادی با شش روش یادگیری ماشین و سه روش یادگیری عمیق برای تشخیص فعالیت مقایسه شده است. نتایج تجربی نشان داده است که روش پیشنهادی به دلیل شناسایی ویژگی های تبعیض آمیز بهتراز تمامی روش های مقایسه ای عمل می کند. روش پیشنهادی با موفقیت میانگین دقت را تا ۱۷.۰۱ درصد و ۳.۹۶ درصد درمقایسه با روشهای یادگیری عمیق، به ترتیب بر روی مجموعه داده های PAMAP2 و OPPORTUNITY بهبود بخشیده است.