Physiological
Measurement

PAPER

MAG-Res2Net: a novel deep learning network for human activity recognition

To cite this article: Hanyu Liu *et al* 2023 *Physiol. Meas.* **44** 115007

IOP Publishing Physiol. Meas. 44 (2023) 115007 https://doi.org/10.1088/1361-6579/ad0ab8

Physiological Measurement

IPEM Institute of Physics and Engineering in Medicine

PAPER

RECEIVED
20 May 2023

REVISED
21 October 2023

ACCEPTED FOR PUBLICATION
7 November 2023

PUBLISHED
28 November 2023

MAG-Res2Net: a novel deep learning network for human activity recognition

Hanyu Liu, Boyang Zhao[1], Chubo Dai, Boxin Sun, Ang Li and Zhiqiong Wang*

College of Medicine and Biological Information Engineering, Northeastern University, Shenyang 110819, People's Republic of China

[1] This author contributed equally to this work and should be considered co-first authors.

* Author to whom any correspondence should be addressed.

E-mail: wangzq@bmie.neu.edu.cn

Keywords: human activity recognition, borderline-SMOTE, multi-scale convolutional neural network, attention mechanism, metric learning

Abstract

Objective. Human activity recognition (HAR) has become increasingly important in healthcare, sports, and fitness domains due to its wide range of applications. However, existing deep learning based HAR methods often overlook the challenges posed by the diversity of human activities and data quality, which can make feature extraction difficult. To address these issues, we propose a new neural network model called MAG-Res2Net, which incorporates the Borderline-SMOTE data upsampling algorithm, a loss function combination algorithm based on metric learning, and the Lion optimization algorithm. *Approach.* We evaluated the proposed method on two commonly utilized public datasets, UCI-HAR and WISDM, and leveraged the CSL-SHARE multimodal human activity recognition dataset for comparison with state-of-the-art models. *Main results.* On the UCI-HAR dataset, our model achieved accuracy, F1-macro, and F1-weighted scores of 94.44%, 94.38%, and 94.26%, respectively. On the WISDM dataset, the corresponding scores were 98.32%, 97.26%, and 98.42%, respectively. *Significance.* The proposed MAG-Res2Net model demonstrates robust multimodal performance, with each module successfully enhancing model capabilities. Additionally, our model surpasses current human activity recognition neural networks on both evaluation metrics and training efficiency. Source code of this work is available at: https://github.com/LHY1007/MAG-Res2Net.

1. Introduction

Human activity recognition (HAR) is an important research area that aims to recognize and classify human activities based on the features extracted from collected data, using machine learning algorithms (Li *et al* 2021). HAR has wide-ranging applications in various fields, such as healthcare, sports and fitness, smart homes, autonomous driving, and security monitoring (Guan *et al* 2017). The proliferation and integration of smart devices and high-fidelity sensing modalities have facilitated advances in human activity recognition. The continuous iteration and refinement of technologies such as smart watches and human signal sensors, including inertial, EMG and ECG related sensors, have precipitated growing research emphasis on enhancing the accuracy and fidelity of human activity detection and classification.

In the healthcare field, HAR technology can help doctors monitor the daily activities of patients to evaluate their rehabilitation progress and provide personalized care. It can also help elderly and disabled people to live independently by monitoring their daily activities and alerting caregivers or family members in case of emergencies (Guan *et al* 2017). In the smart home field, HAR technology can help smart devices adapt to users' activities automatically to provide better services and experiences (Lu *et al* 2017). In the sports and fitness field, HAR technology can help athletes and fitness enthusiasts monitor their movement postures and techniques to improve their exercise effects and reduce the risk of injury (Host and Ivašić-Kos 2022).

There are two main technical approaches for human activity recognition: video-based systems and sensor-based systems. Video-based systems use cameras to capture images or videos to recognize people's behaviors;

sensor-based systems use body or environmental sensors to estimate people's movement details or record their activity trajectories (Dang *et al* 2020). Sensors can be embedded in portable mobile devices (such as phones, watches, and sports wristbands) and immovable objects (such as furniture and walls) to collect people's daily movement information continuously and non-invasively (Ferrari *et al* 2021). Currently, due to privacy and portability issues, sensor-based HAR systems have dominated our daily activities. This article mainly discusses the sensor-based HAR problem.

The development and application prospects of HAR technology are very broad. With the continuous development and innovation of artificial intelligence and machine learning technology, HAR technology has played an important role in many application scenarios (Khan and Ghani 2021). In the early stage, traditional machine learning methods such as decision trees (DT) (Fan *et al* 2013), support vector machines (SVM) (Vijayvargiya *et al* 2021), random forests (RF) (Nazari *et al* 2021), and naive Bayes (NB) (Gadebe *et al* 2020) have made considerable progress in sensor-based human activity recognition. Vijayvargiya *et al* (2021) studied and implemented the application of machine learning algorithms such as support vector machines and random forests in human activity recognition, and the results showed that using time and frequency domain features can improve the accuracy of the algorithm. Lin *et al* (Fan *et al* 2013) proposed a human activity recognition model based on decision tree algorithm and conducted experimental research and performance evaluation on the model, providing an effective method for achieving efficient and accurate human activity recognition. Gadebe *et al* (2020) proposed an intelligent smartphone naive Bayes human activity recognition method based on personalized data sets and achieved good recognition performance in experiments. It merits note that hidden Markov models (HMMs) significantly outperform alternative machine learning models on human activity recognition tasks (Xue and Liu 2022). This stems from their inherent aptitude for logical modelling. Additionally, HMMs can surpass certain deep learning models in interpretability, generalizability, efficiency, and scalability. Historically, feature extraction for machine learning has relied heavily upon expert knowledge. To address the limitations of traditional techniques regarding both feature extraction and classification performance, Hartmann *et al* (2023b) proposed an innovative combination of interpretable high-level features for human activity classification. Among examined models, hidden Markov models achieved the highest performance.

Unlike traditional machine learning, deep learning has powerful nonlinear modeling capabilities and can automatically extract features, which provides possibilities for the automation of activity recognition tasks (Almaslukh *et al* 2018). In recent years, many deep learning-based action recognition methods have been proposed (Li and Wang 2022a, Mekruksavanich *et al* 2022a, 2022b). Mekruksavanich *et al* (2022a) proposed a sensor-based human activity recognition (HAR) model called ResNet-SE that incorporates channel-wise attention through squeeze-and-excitation modules and shortcut connections to enhance recognition performance on complex HAR tasks, thereby improving the accuracy of recognizing complex human activities from smartwatch sensor data. Li *et al* (Li and Wang 2022a) developed a deep learning model predicated on residual blocks and bidirectional LSTM. It first leverages residual blocks to automatically extract spatial features from multi-dimensional signals from MEMS inertial sensors, followed by the application of a BiLSTM algorithm to capture sequential dependencies between the extracted features. Compared to existing models, this model achieves superior performance with fewer parameters. Furthermore, Mekruksavanich *et al* (2022b) devised a human activity recognition deep neural network called ResNeXt based on biosignals. By modulating the number of groups, it strikes a balance between plain convolutional kernels and depthwise separable convolutions, achieving state-of-the-art performance on the multimodal CSL-SHARE dataset.

Although deep learning has achieved significant success in the field of human activity recognition and has been applied to various real-life scenarios, there are still some technical challenges that need to be addressed, as demonstrated by experimental results (Han *et al* 2005, Wang *et al* 2019, Mekruksavanich *et al* 2022b). This article identifies the following three primary challenges faced by the HAR field:

- Class imbalance: class imbalance is a common issue in various fields, and in activity recognition, both minority and majority classes are equally important. This makes it difficult to extract accurate rules, and the model tends to become data-dependent, thereby reducing its overall performance (Han *et al* 2005).
- Difficulty in deep feature extraction: due to the diversity of human activities, data quality, and high similarity between similar actions, deep feature extraction in the HAR field is relatively challenging (Bulling *et al* 2014). This often leads to reduced accuracy, longer development and training times/costs, and poor model robustness when applied.
- Classification accuracy problem: high similarity and individual differences in action classification in the field of human activity recognition typically result in low classification accuracy, characterized by large intra-class

differences and high inter-class similarity (Wang *et al* 2019). This can lead to poor model generalization ability, difficulty in accurate classification, and overfitting.

To tackle the challenges in sensor-based human activity recognition, we investigate a more effective neural network structure called Multiscale-Attention-Gated Res2Net (MAG-Res2Net), which is capable of capturing multi-scale time-series features and assigning different weights to these features to enhance the model's robustness and accuracy. Additionally, we optimize the entire process from data preprocessing to model training and integrate all the optimization methods to address the challenges faced by the field of human activity recognition.

(1) During the data preprocessing stage, we applied the Borderline SMOTE oversampling algorithm to oversample the boundary samples of the minority class in the dataset, thereby effectively mitigating the negative effects of data imbalance. To be specific, we carefully tuned the relevant parameters and thresholds and executed this approach through a series of algorithmic steps. Experimental results demonstrate that this method outperforms both SMOTE and random oversampling methods (Han *et al* 2005).

(2) During the model construction stage, we proposed a novel neural network model that integrates multi-scale gated residual networks and attention mechanisms, which can effectively address the challenge of deep feature extraction, thereby enhancing the accuracy and robustness of the model (Yang *et al* 2022).

(3) During the model training stage, we integrated center loss and cross-entropy loss functions to minimize intra-class differences and inter-class similarities, improving the model's classification accuracy and effectively alleviating the overfitting issue that may exist in certain categories (Chen *et al* 2023). Furthermore, for large-scale data samples, we applied the Lion optimization algorithm and optimized the relevant parameters, which led to significantly faster and better performance than the current mainstream optimization algorithms.

This paper aims to investigate and employ an advanced deep learning optimization method for sensor-based human activity recognition to tackle the challenges in this field. The improvements proposed in this paper consist of three aspects: a data upsampling approach based on the Borderline-SMOTE algorithm, MAG-Res2Net that integrates multi-scale residual networks and adaptive attention mechanisms, and a hybrid loss function based on the Lion optimization algorithm. By conducting experiments on benchmark datasets such as UCI-HAR and WISDM, we demonstrate that our proposed method achieves remarkable performance in human activity recognition, outperforming some existing methods, which validates its effectiveness and generality.

2. Method

2.1. Overall process framework

We start by cleaning, denoising, and normalizing the sensor data. For datasets with class imbalance, we use the Borderline-SMOTE algorithm to perform upsampling and balance the number of samples between different classes. Then, we select Gated-Res2Net as the primary model structure and introduce the ECA module. In the model training phase, we use two loss functions, cross-entropy loss and center loss, to improve classification accuracy through two approaches. Finally, based on the chosen training batch size, we select either the AdamW or Lion optimization algorithm to train the model. The specific details are outlined in the following sections. Figure 1 shows the flow of our work.

2.2. Data preprocessing

In human activity recognition, class imbalance can lead to difficulties in extracting accurate patterns, resulting in data dependency and reduced model performance (Wang *et al* 2019). To address the issue of data imbalance, we utilized the Borderline-SMOTE algorithm for upsampling. This algorithm can generate new synthetic samples at the boundaries between classes while maintaining the original data distribution, thereby increasing the number of samples in the minority class.

First, we divided the minority class samples into three categories: safe, danger, and noise, and only oversampled the Danger samples. Then, we used two independent branch algorithms, Borderline-SMOTE1 and Borderline-SMOTE2, depending on the different situations. In Borderline-SMOTE1, we randomly selected minority class samples in the K nearest neighbors to generate new samples, while in Borderline-SMOTE2, we selected any sample in the K nearest neighbors without considering its class (Wang *et al* 2019). Furthermore, we assume S as the sample set, S_{min} as the minority class samples, $S_{max\,j}$ as the majority samples set of class j, m as

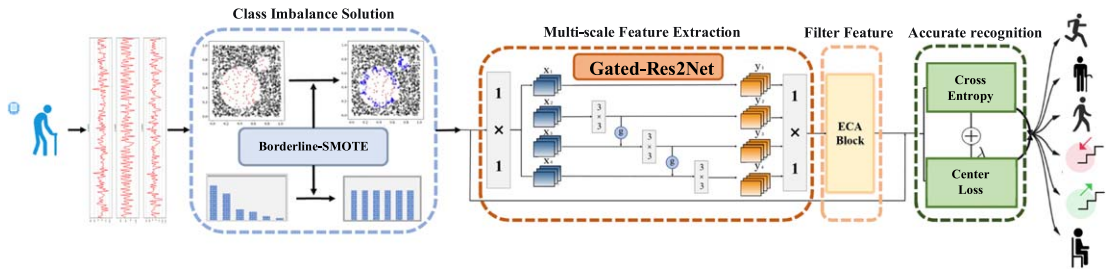

Figure 1. Diagram of the method flow.

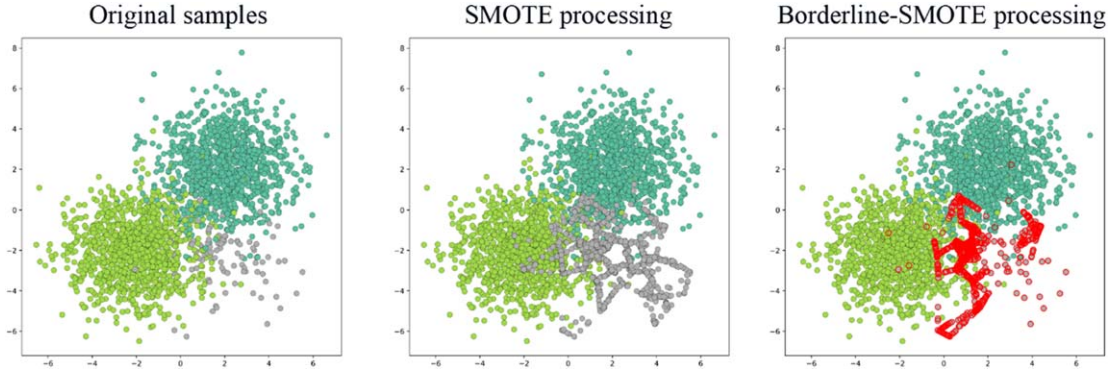

Figure 2. Comparison of samples before and after upsampling methods.

the number of neighboring samples, x_i as the attribute, x_{ij} as all attributes of neighboring samples, x_n as the neighbor sample, and R_{ij} takes values of 0.5 or 1. Then, for each $x_i \in S_{min}$, we dccetermine the nearest sample set, denoted as S_{NN}($S_{NN} \in S$). For each sample x_i, we count the number of samples belonging to the majority class sample set in the closest neighbor samples, which is $|S_{NN} \cap S_{max\,j}| < m$; If this condition is satisfied, we synthesize minority class samples. Additionally, we calculate the difference between x_i and the corresponding attribute j of the neighbor, denoted as $d_{ij} = x_i - x_{ij}$. Finally, we generate new minority class samples using the following formula:

$$h_{ij} = x_i + d_{ij} \times \text{rand}(0, R_{ij}).$$

Finally, we validated whether the number of samples belonging to each class in the validation dataset was balanced. If the dataset was balanced, we passed it to the training function. Figure 2 shows how the Borderline-SMOTE1 algorithm compares to other methods.

2.3. The proposed MAG-Res2Net

Deep feature extraction in the field of human activity recognition (HAR) can be challenging, leading to issues such as reduced accuracy, longer development and training time and costs, and poor model robustness during application (Bulling *et al* 2014). In this study, we utilized the Gated-Res2Net and ECA modules. Gated-Res2Net is an improvement of Res2Net that introduces a gating mechanism while maintaining its multi-scale feature extraction capability. Meanwhile, the ECA module enhances the model's ability to focus on important features.

We first established a ResNet neural network module and replaced a group of 3×1 convolutional kernels with a smaller set of filter groups consisting of w channels, using a hierarchical residual connection to connect different filter groups, thereby constructing Res2Net (Gao *et al* 2019). After each 3×1 convolution, we added a gating module to form Gated-Res2Net (Yang *et al* 2020): network model architecture: figure 3(a) shows ResNet, (b) shows Res2Net and (c) shows Gated-Res2Net.

Then, the upsampled data is divided into multiple data x_i, and the other data except x_i are fed into a 3×1 convolution. As the data is propagated downward, the gating module selects the effective feature maps for downward propagation.

$$f_2(x) = \text{concat}(a(X),\ a(y_i - 1),\ a(x_i))$$

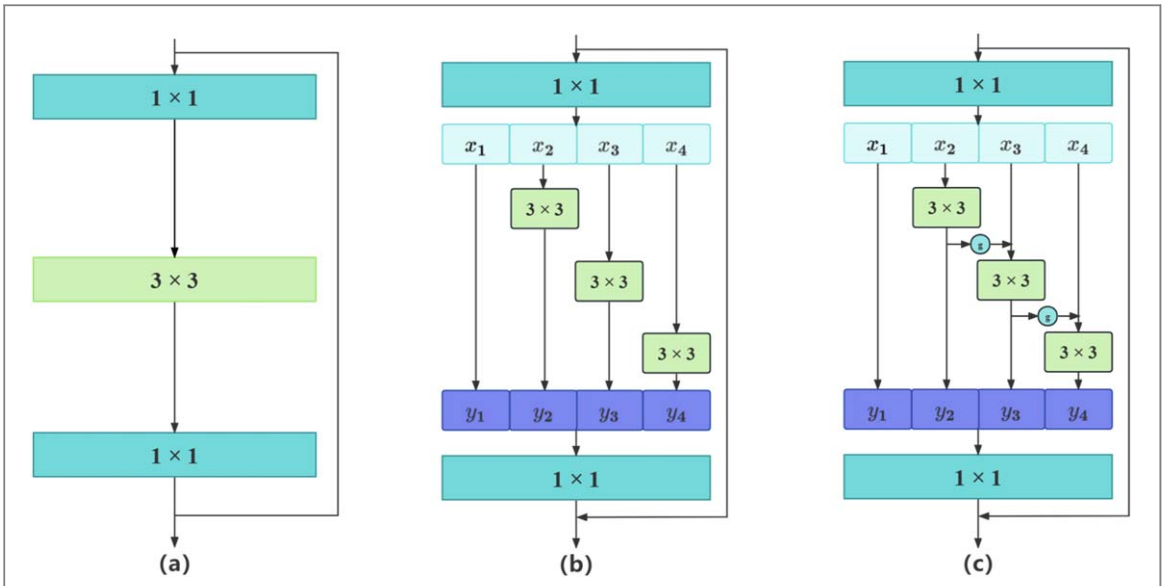

Figure 3. (a) Shows ResNet (b) shows Res2Net and (c) shows Gated-Res2Net.

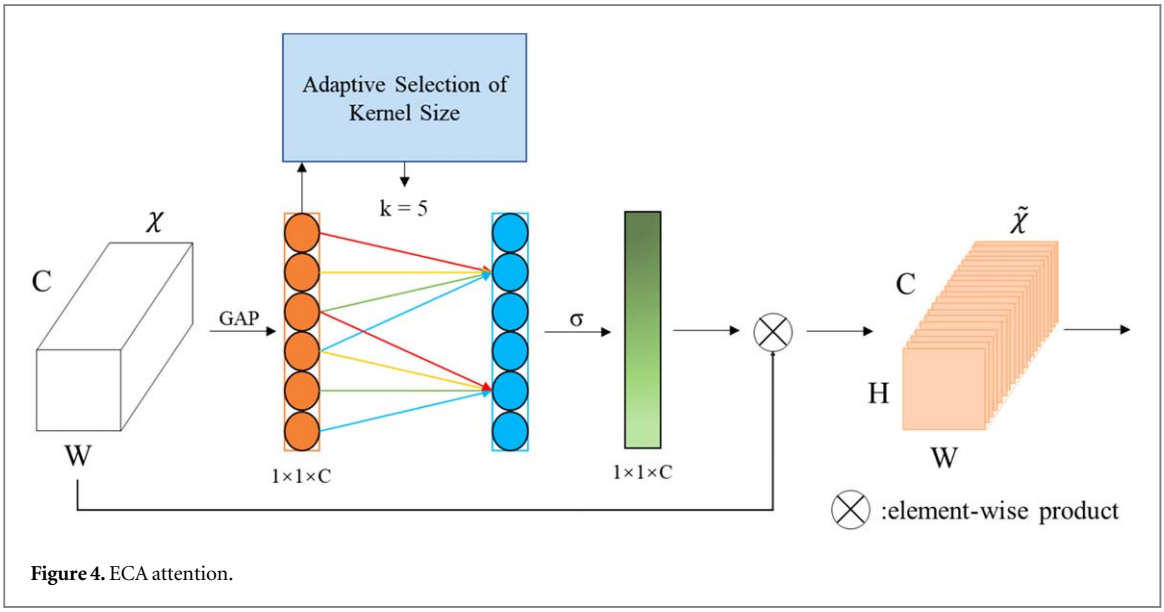

Figure 4. ECA attention.

$$f_1(x) = a(f_2(x))$$

$$g_i = \tanh\ (f_1(x))$$

The data is propagated downward to form multiple data y_i:

$$y_i = \begin{cases} x_i & i = 1 \\ K_i(x_i) & i = 2 \\ K_i(x_i + g_i \times y_{i-1}) & 2 < i \leqslant s \end{cases}.$$

After stacking, the output is sent to a 1×1 convolutional layer. Next, we pass the data through an ECA module after a 1×1 convolution. The ECA module performs global average pooling on the input feature data, resulting in a feature vector with a channel number of 1, which is then subjected to one-dimensional convolution (Wang *et al* 2020). Finally, we apply the sigmoid activation function to this vector, obtaining a vector with values ranging from 0 to 1, and multiply it with the input feature data to obtain weighted feature data. Figure 4 shows the processing flow of the ECA module.

2.4. Combination and optimization of loss functions

The issue of classification accuracy is often characterized by large within-class differences and high inter-class similarity (Wang *et al* 2019). This can result in poor model generalization, inaccurate classification, and overfitting. To tackle this issue, we utilized a combination of cross-entropy loss and center loss. Cross-entropy loss is used to optimize the accuracy of class differentiation and penalize classification errors, while center loss is used to optimize the tightness of within-class features by clustering feature vectors of the same class together (Wen *et al* 2016). We calculated cross-entropy loss and center loss separately and weighted them to obtain the total loss function.

We first selected the Lion optimizer or AdamW optimizer based on the chosen batch size according to the size of the dataset, and then trained the model.

The formula for the Lion optimization algorithm is:

$$\text{Lion:} = \begin{cases} \mu_t = \text{sign}\left(\beta_1 m_{t-1} + (1-\beta_1) g_t\right) + \lambda_t \theta_{t-1} \\ \theta_t = \theta_{t-1} - \eta_t \mu_t \\ m_t = \beta_2 m_{t-1} + (1-\beta_2) g_t \end{cases}$$

AdamW optimization algorithm formula is:

$$\text{AdamW:} = \begin{cases} m_t = \beta_1 m_{t-1} + (1-\beta_1) g_t \\ \nu = \beta_2 \nu_{t-1} + (1-\beta_2) g_t^2 \\ \hat{m}_t = m_t / (1-\beta_1^t) \\ \hat{\nu}_t = \nu_t / (1-\beta_2^t) \\ \mu_t = \hat{m}_t / \left(\sqrt{\hat{\nu}_t} + \epsilon\right) + \lambda_t \theta_{t-1} \\ \theta_t = \theta_{t-1} - \eta_t \mu_t \end{cases}$$

At the end of each batch training, we calculate the loss of the cross-entropy loss function and the weighted center loss function, which can be expressed as:

$$\text{Loss}_{ce} = -\sum_{i=1}^{n} p(x_i) \log q(x_i)$$

$$\text{Loss}_{cr} = \frac{1}{2N} \sum_{i=1}^{N} \| x_i - c \|_2^2$$

$$\text{Loss}_{\text{total}} = \text{Loss}_{ce} + \text{Loss}_{cr}.$$

After that, we update the model's weight parameters through backpropagation. Then, we evaluate the trained model using the validation set, calculating the model's classification accuracy and the value of the center loss function, to adjust the model parameters and improve the model's performance. Finally, we evaluate the final model using the test set, calculating the model's classification accuracy and the value of the center loss function.

3. Problem statement

3.1. Experimental design

In this study, we utilized the Multiscale-Attention-Gated Res2Net (MAG-Res2Net) as the network model and employed a comprehensive multi-scale metric learning optimization method for human activity recognition. We tested the model's performance on the UCI-HAR and WISDM datasets, where each sample in the dataset belongs to one of six different classes. To evaluate the model's prediction results, we input the test dataset into the model after transforming each sample into an array of shape [9, 128] (for the UCI-HAR dataset) or [3, 90] (for the WISDM dataset) to make it compatible with the model. The output of our model is an array of shape [6,], which we convert into a probability distribution. We then select the class with the highest probability as the predicted result.

3.2. Dataset used in the experiments

3.2.1. WISDM dataset

The WISDM (Wireless Sensor Data Mining) dataset is a public dataset widely used in the field of human activity recognition. The dataset was released in 2010 by a research team at Binghamton University, State University of New York, and contains sensor data from smartphones and tri-axial accelerometers (Kwapisz *et al* 2011). The dataset includes six common daily activities, including walking, going upstairs, going downstairs, sitting, standing, and lying down. The specific information is shown in table 1:

Table 1. Sample information of the WISDM dataset.

Number	Activity	Sample size
1	Walking	424 400
2	Jogging	342 177
3	Upstairs	122 869
4	Downstairs	121 757
5	Sitting	58 637
6	Standing	40 663

Table 2. Sample information of the UCI-HAR dataset.

Number	Activity	Sample size
1	Walking	1517
2	Walking Upstairs	1391
3	Walking Downstairs	1355
4	Sitting	1986
5	Standing	1965
6	Laying	2435

Table 3. Dataset partitioning.

Datasets	WISDM	UCI-HAR
Training samples	60%	80%
Validation samples	20%	20%
Test samples	20%	Separate

3.2.2. UCI-HAR Dataset

The UCI-HAR (UCI Human Activity Recognition) dataset is a public dataset widely used in the field of human activity recognition. The dataset was released in 2012 by a research team at the University of California, Irvine, and contains sensor data from smartphones and tri-axial accelerometers (Anguita *et al* 2013). The dataset includes six common daily activities, including walking, going upstairs, going downstairs, sitting, standing, and lying down. The specific information is shown in table 2:

3.2.3. Dataset partitioning

We will partition the dataset into training, validation, and test sets using the random partitioning method. During training, we will use repeated k-fold cross-validation ($K = 5$) to divide the training and validation sets into a 4:1 ratio. Specifically, the training set will be used for model training, the validation set will be used for adjusting model parameters, and the test set will be used for the final evaluation of the model's performance. Furthermore, we will not use any data upsampling methods in the test set. The specific information is shown in table 3:

We utilize this partitioning method to fully exploit the samples in the dataset while minimizing overfitting and selection bias issues. To avoid model performance bias caused by imbalanced sample distributions, we will ensure that the training, validation, and test sets have similar sample distributions. Before partitioning the dataset, we will perform necessary preprocessing steps, such as data cleaning, feature extraction, and feature normalization. To ensure that the resulting dataset adequately reflects real-world problems, we will preserve the target value distribution during dataset partitioning. Finally, we will use the partitioned dataset for model training and evaluation and record the model metrics.

3.3. Evaluation metrics

To evaluate the performance of the proposed model for HAR, the followed metrics (Yang *et al* 2019) were used for evaluation generally.

3.3.1. Accuracy

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}.$$

TP represents true positive, which is the number of samples correctly identified as positive; TN represents true negative, which is the number of samples correctly identified as negative; FP represents false positive, which is the number of samples incorrectly identified as positive; FN represents false negative, which is the number of samples incorrectly identified as negative.

3.3.2. $\text{F1} - \text{macro}$

$$\text{F1} - \text{macro} = 2 \times \frac{\text{Precision}_{\text{macro}} \times \text{Recall}_{\text{macro}}}{\text{Precision}_{\text{macro}} + \text{Recall}_{\text{macro}}}$$

$\text{Precision}_{\text{macro}}$ represents the average precision for all labels and is used to measure the proportion of samples correctly identified as positive in all samples predicted as positive. $\text{Recall}_{\text{macro}}$ represents the average recall for all labels and is used to measure the coverage of positive samples by the model, that is, the proportion of samples correctly identified as positive to all true positives.

3.3.3. $\text{F1} - \text{weighted}$

$$\text{F1} - \text{weighted} = 2 \times \frac{\text{Precision}_{\text{weighted}} \times \text{Recall}_{\text{weighted}}}{\text{Precision}_{\text{weighted}} + \text{Recall}_{\text{weighted}}}$$

$\text{Precision}_{\text{weighted}}$ used to measure the predicted results of the model, which is the weighted average of the proportion of samples correctly identified as positive in all samples predicted as positive. $\text{Recall}_{\text{weighted}}$ is used to measure the coverage of positive samples by the model, which is the weighted average of the proportion of samples correctly identified as positive to all true positives.

3.3.4. $\text{G} - \text{mean}$

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{TNR} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

$$\text{G} - \text{mean} = \sqrt{\text{TPR} \times \text{TNR}}$$

TPR stands for true positive rate, which is the proportion of positive samples that are correctly predicted by a classifier. TNR stands for True Negative Rate, which is the proportion of negative samples that are correctly predicted by a classifier. To calculate the $\text{G} - \text{mean}$ metric, first calculate TPR and TNR, then multiply them and take the square root of the result.

3.4. Hyperparameter tuning

Experimental environment: the experiments in this paper were conducted on the Kaggle platform. We used two types of graphic processing units (GPUs) available on the platform (NVIDIA P100 GPU 16 GB and NVIDIA Tesla T4 GPU 16 GB), with a server processor Intel(R) Xeon(R) @ 2.00 GHz, and default configurations for memory, storage, and other devices. We implemented the experiments using Python 3.7.12 on the Pytorch 1.3.0 deep learning framework, and conducted the experiments on Ubuntu 20.04.5 LTS, as shown in table 4.

3.4.1. Optimization method

We will use grid search method to find the optimal hyperparameter combination. Specifically, we will perform a grid search over the following hyperparameters, and the search range is shown in table 5:

Finally, as shown in table 6, we obtained different optimal hyperparameter combinations in two different datasets.

Table 4. Experimental environment.

Platform	Kaggle
OS	Ubuntu 20.04.5 LTS
GPU1	NVIDIA P100 GPU 16GB
GPU 2	NVIDIA Tesla T4 GPU 16GB
CPU	Intel(R) Xeon(R) @ 2.00 GHz
Framework	PyTorch1.13.0

Table 5. The range of hyperparameter tuning.

Datasets	UCI-HAR and WISDM
Model	18–152
Batch size	32–2048
Learning rate 1	0.0001–0.1 (AdamW/Lion)
Learning rate 2	0.0001–0.1 (AdamW/Lion)
Weight	0.0001–0.1
Epochs	15–100

Table 6. The optimal combination of hyperparameters.

Datasets	UCI-HAR	WISDM
Model	50(3-4-6-3)	42 (3-4-5-1)
Batch size	256	2048
Learning rate 1	AdamW, 5.0E-03	Lion, 3.0E-03
Learning rate 2	AdamW, 5.0E-03	Lion, 3.0E-03
Loss weight	0.0005	0.0007
Epochs	30	40

Table 7. Model performance.

Datasets	UCI-HAR	WISDM
ACC	94.44%	98.32%
F1-macro	94.38%	97.26%
F1-weight	94.26%	98.42%
G-mean	0.9709	0.9874

3.5. Expected results

By utilizing hyperparameter tuning methods, we can identify the optimal hyperparameter combination to enhance the accuracy and F1 score of the model and minimize training time and memory usage. The findings of this study will aid in improving the performance and efficiency of human activity recognition technology, as well as provide guidance for designing and optimizing deep learning classification models.

4. Results analysis

4.1. Model performance on datasets

In this study, we conducted experiments on two datasets: the UCI-HAR dataset was used to test the efficacy of our method's single module, while the WISDM dataset was used for final evaluation. We trained multiple deep learning models and conducted a comprehensive analysis and interpretation of the experimental results. The confusion matrix format we used was designed and displayed according to the format proposed in literature (Jannat *et al* 2023).

The overall results are shown in table 7, and specific results are described in the following text:

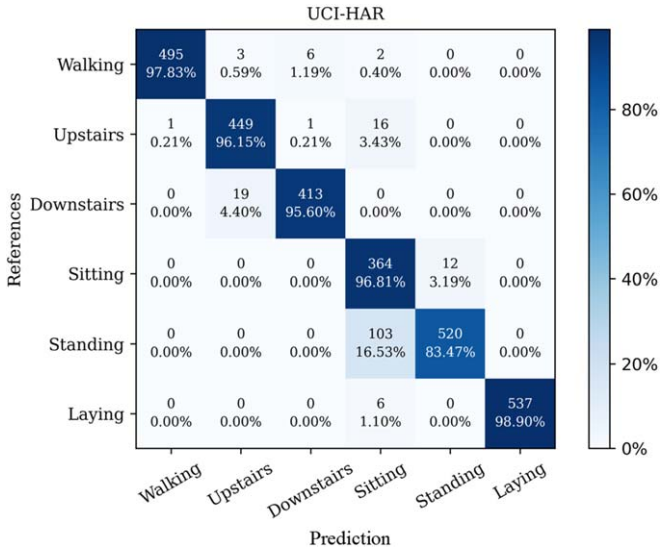

Figure 5. Confusion matrix (UCI-HAR).

4.1.1. Performance on UCI-HAR dataset

On the UCI-HAR dataset, MAG-Res2Net utilized a network with 50 residual blocks, with varying numbers of residual blocks at different scales [3-4-6-3]. Given the limited sample size in UCI-HAR, we configured a batch size of 256 and a combined cross-entropy and center loss function with the AdamW optimizer. The learning rates were set to 5.0E-03 and 5.0E-03, weight to 0.005, trained for 30 epochs. This yielded final accuracy of 94.44%, F1-macro score of 94.38%, F1-weight score of 94.26%, and G-mean of 0.9709. In contrast to the balanced category distribution of WISDM, UCI-HAR categories exhibit high class imbalance. Consequently, applying upsampling algorithms on such balanced datasets can degrade model performance. For the imbalanced WISDM dataset, this issue was effectively mitigated. This demonstrates the advantage of our Borderline-SMOTE algorithm applies primarily to imbalanced dataset categories. Additionally, the predefined training and test splits of UCI-HAR potentially introduce data inconsistencies. Our experiments using the training set for further training, validation and testing sans the original test set yielded performance approaching that on WISDM. This implies inherent differences between the UCI-HAR training and test distributions. Figure 5 shows the confusion matrix of MAG-Res2Net on UCI-HAR.

4.1.2. Performance on WISDM dataset

On the WISDM dataset, due to severe overfitting on a network with 50 residual blocks, we used a lighter MAG neural network with 42 residual blocks, with a different number of residual blocks at different scales (3-4-5-1). The WISDM dataset with data upsampling algorithm has many samples, so we set the batch size to 2048 and used the Lion optimizer, which has advantages in large-scale training. We set the learning rate to 3.0E-03 and 3.0E-03, weight to 0.007, and trained for 40 epochs. The final accuracy was 98.32%, F1-macro score was 97.26%, and F1-weight score was 98.42% and G-mean was 0.9874. In addition, for datasets like WISDM that have large differences in sample sizes between classes, Borderline-SMOTE generates minority class samples that are nearly ten times the original amount for some subclasses (Sitting, Standing). This results in the model relying too heavily on the synthesized data, However, our model still achieved decent performance with Borderline-SMOTE. We believe there would be better results if we could effectively constrain the upsampling of the algorithm. Therefore, we suggest exercising caution when applying Borderline-SMOTE on datasets with highly imbalanced class distributions, and avoid aggressive upsampling that equalizes sample sizes across all classes. Figure 6 shows the confusion matrix of MAG-Res2Net on WISDM.

4.1.3. Summary of model performance

Leveraging the scalability of the Lion optimizer for large-batch training, we substantially increased the batch size to accelerate training. Comparative experiments on WISDM with the AdamW and Lion optimizers demonstrated nearly identical model performance when specifying batch sizes of 1024 and 2048, respectively. However, the training speed was markedly enhanced with Lion. Moreover, owing to factors such as sample distribution divergence, the reduced 42 layer MAG-Res2Net architecture attained superior performance on WISDM compared to the 50 layer version on UCI-HAR. According to the confusion matrix, the model has excellent discriminative ability for general actions. For similar actions, such as standing and sitting, the

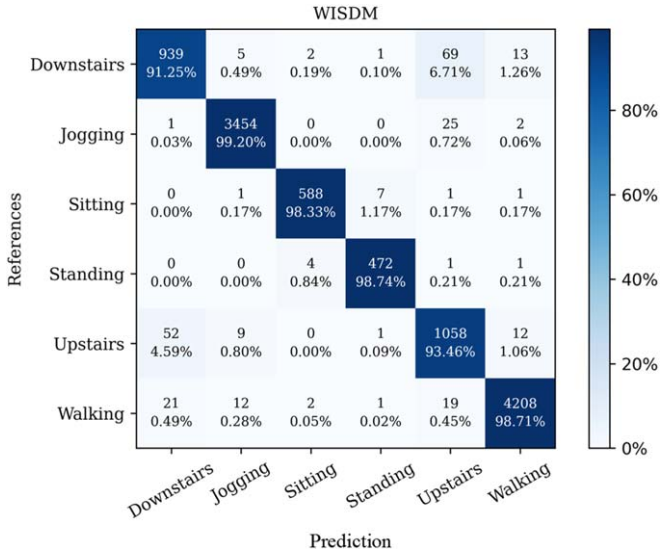

Figure 6. Confusion matrix (WISDM).

Table 8. Ablation experiment.

Model		Dataset: UCI-HAR				Dataset: WISDM			
Network	Extra Addition	ACC	F1-m	F1-w	G-mean	ACC	F1-m	F1-w	G-mean
Res2Net Gao *et al* (2019)	\	92.01%	92.15%	91.78%	0.9615	95.00%	92.71%	94.81%	0.9562
	Gated model	92.43%	92.40%	92.32%	0.9598	96.40%	94.91%	96.42%	0.974
Gated-Res2Net Yang *et al* (2020)	SE Model	93.42%	93.47%	93.39%	0.9663	97.02%	95.83%	97.02%	0.9793
	ECA Model	94.03%	93.97%	93.95%	0.9689	97.50%	96.32%	97.49%	0.9811
MAG-Res2Net	Loss combined	93.89%	93.94%	93.91%	0.9679	97.52%	96.74%	97.53%	0.9817
	SMOTE	92.74%	92.68%	92.73%	0.963	95.98%	94.78%	96.03%	0.9758
	Loss combined + SMOTE	93.61%	93.25%	93.21%	0.9653	97.39%	96.26%	97.39%	0.9812
	Borderline	93.91%	93.95%	93.89%	0.9644	96.92%	95.11%	96.76%	0.9743
	Loss combined + Borderline	94.44%	94.38%	94.26%	0.9709	98.32%	97.26%	98.42%	0.9874

recognition ability still needs improvement, but the discriminative ability for these actions has been significantly enhanced compared to previous models. Regarding the parameter selection of the Lion optimizer, we found that, under the same other hyperparameters, the optimal learning rate of the Lion optimizer is often one-third of the optimal learning rate of the AdamW optimizer. If the learning rate is adjusted to one-tenth of the AdamW learning rate, the results are usually very poor.

In addition, the MAG-Res2Net model requires slightly longer training time and more memory usage during training than other simple neural network models. However, in the experiment exploring the network depth, we found that reducing a certain number of network layers can effectively reduce training time and memory usage without significantly affecting the model's performance. Therefore, in practical applications, we can balance the model's performance and training cost by reducing the network depth.

Overall, MAG-Res2Net has shown remarkable performance in human action recognition problems and is expected to play a crucial role in practical applications.

4.2. Ablation experiment

In this study, we conducted ablation experiments on the UCI-HAR and WISDM datasets to systematically evaluate the impact of various deep learning architectural configurations and techniques on classification performance. F1-m in the table 8 stands for F1-macro and F1-w for F1-weighted. Specifically, we utilized Res2Net as the baseline model and incrementally incorporated methods including gated modules, squeeze-and-excitation (SE) modules, and efficient channel attention (ECA) modules to quantify performance

Table 9. Comparison with existing work.

Dataset	Model	Accuracy	F1-macro	F1-weighted
WISDM	CNN Ignatov (2019)	93.31%	91.83%	93.51%
	LSTM (Dang *et al* 2020)	96.71%	95.50%	96.68%
	LSTM-CNN Xia *et al* (2020)	95.90%	94.33%	95.97%
	TSE-CNN Huang *et al* (2020)	94.57%	92.33%	94.78%
	SC-CNN Alemayoh *et al* (2021)	94.87%	94.33%	95.00%
	CNN-GRU Dua *et al* (2021)	94.95%	95.00%	96.21%
	Self-Attention Mahmud *et al* (2020)	96.23%	95.03%	96.13%
	CNN-Attention Gao *et al* (2021)	96.16%	95.07%	96.31%
	ResNet He *et al* (2016)	94.74%	92.84%	94.35%
	Res2Net Gao *et al* (2019)	95.00%	92.91%	94.81%
	SE-Res2Net Gao *et al* (2019)	95.52%	94.73%	95.56%
	Gated-Res2Net Yang *et al* (2020)	96.40%	94.91%	96.42%
	Res-BLSTM Li and Wang (2022b)	97.25%	96.64%	97.33%
	ResNeXt Mekruksavanich *et al* (2022b)	97.52%	96.40%	97.45%
	Gated-Res2Net (SE) Yang *et al* (2020)	97.02%	95.83%	97.02%
	MAG-Res2Net	97.50%	96.32%	97.49%
	MAG-Res2Net*	98.32%	97.26%	98.42%

improvements. We additionally attempted to further enhance model capabilities by Loss Combined with SMOTE and Borderline-SMOTE oversampling techniques.

Our results demonstrate that integrating gated mechanisms within the Res2Net architecture led to minor performance gains over vanilla Res2Net, with statistically significant augmentation of F1 metrics. Building upon the gated multi-scale residual backbone, attention modules were evaluated, with SE modules conferring considerable improvements to both accuracy and F1 scores, while ECA modules yielded slightly lower gains compared to SE.

Moreover, on top of the attention-gated multi-scale residual base network, we assessed supplementary deep learning techniques. The data oversampling methods proved ineffective on the balanced UCI-HAR dataset, where oversampling degraded model performance. In contrast, Borderline-SMOTE alone hindered performance on imbalanced WISDM, but Loss Combined with Borderline-SMOTE sampling substantially boosted model capabilities, also reflected in UCI-HAR albeit with smaller gains due to its balanced class structure.

Taken together, our results demonstrate that incorporating techniques such as gated modules, SE modules, and ECA modules into deep neural architectures can effectively enhance classification performance on UCI-HAR. Moreover, Loss Combined with Borderline-SMOTE surpasses vanilla SMOTE, with Borderline-SMOTE models exhibiting more robust performance, although struggling to discriminate similar classes without Loss Combined, evident upon comparing standalone Borderline-SMOTE versus the hybrid approach. Furthermore, Borderline-SMOTE appears best suited for imbalanced datasets, conferring negligible benefits on balanced distributions like UCI-HAR. Our systematic evaluation provides significant insights to inform the design and optimization of deep learning systems.

4.3. Comparison with existing work

We compared the proposed MAG-Res2Net architecture to 17 related neural network models from recent literature, comprising early traditional networks like CNNs and LSTMs, as well as emerging attention-based models and state-of-the-art sensor-based HAR approaches. All models were reimplemented using standardized methodology and evaluated on the WISDM dataset without optimization techniques to enable direct comparison of model capabilities. Three metrics were assessed - accuracy, F1-macro, and F1-weighted.

The results are shown in table 9. The results demonstrate MAG-Res2Net attained accuracy of 98.32%, F1-macro of 97.26%, and F1-weighted of 98.42% even without optimizations, surpassing existing models to some degree. Among early models, CNNs and variants displayed mediocre performance, while LSTM-related architectures achieved good results but substantially slower inference than CNNs. Additionally, CNNs including ResNet and Res2Net exhibited low F1-macro scores compared to LSTMs and attention models, which was ameliorated but not resolved by gating mechanisms. This issue was mitigated by incorporating attention modules like SE and ECA. By integrating multiscale networks, gating, and attention, MAG-Res2Net outperformed individual usage of these modules. Furthermore, MAG-Res2Net surpassed generic attention-based models in accuracy and reliability (Gao *et al* 2021).

Table 10. Two-way SOTA comparison.

Model	Dataset	ACC	F1-m	F1-w
MAG-Res2Net	WISDM	97.50%	96.32%	97.49%
	UCI-HAR	94.03%	93.97%	93.95%
	CSL-SHARE	92.38%	93.48%	92.37%
ResNeXt Mekruksavanich *et al* (2022b)	WISDM	97.52%	96.40%	97.45%
	UCI-HAR	93.57%	93.15%	93.35%
	CSL-SHARE	91.60%	92.13%	\
Uncertainty Quantification Folgado *et al* (2023)	CSL-SHARE	83.5%	\	\
HLF Hartmann *et al* (2023a)		89.7%	\	\
HMM Hartmann *et al* (2022)		96.1%	\	\
HMM Liu (2021)		97.5%	\	\
Motion Units Liu *et al* (2021b)		93.9%	\	\
LDA-HMM Hartmann *et al* (2021)		$93.7 \pm 1.4\%$ $97.8 \pm 0.2\%$	\	\

Relative to most other neural architectures, MAG-Res2Net demonstrates superior accuracy, albeit with numerically modest gains. This is largely attributable to the intrinsic difficulty of discriminating between similar activity categories, a ubiquitous challenge in sensor-based human activity recognition. In the present datasets, the standing and sitting categories exemplified this issue, resulting in less pronounced improvements in overall accuracy and F1 metrics, despite MAG-Res2Net conferring substantially larger gains on these specific categories. Nevertheless, certain proposed techniques like Loss Combined and multi-scale networks still conferred solid performance boosts on these categories without considerably increasing computational or memory costs, albeit accompanied by marginal reductions in accuracy on other categories. Such trade-offs are acceptable but will be targeted for minimization in future work to more substantially elevate overall model performance.

4.4. Comparison with SOTA models

We further benchmarked MAG-Res2Net against ResNeXt, the current state-of-the-art ResNet architecture for this domain. Specifically, we evaluated both models on the WISDM and UCI-HAR datasets utilized in this work. Additionally, we assessed MAG-Res2Net on the multimodal CSL-SHARE dataset from the ResNeXt study, which incorporates diverse sensing modalities and rigorous segmentation/annotation, enabling assessment of generalizability to varied modalities (Liu *et al* 2021a, 2023). For CSL-SHARE, we adopted the top-performing sensor combination of EMG, ECG, and IMU from ResNeXt. We unified all the data and extracted the time periods containing valid actions for each user ID. The extracted time periods were aggregated by user ID. Missing values were imputed using linear interpolation. After standardizing the data, we applied a sliding window approach to divide the data into segments. Each segment contained 2000 time steps with 50% overlap between segments. The data was split into roughly 70% for training, 10% for validation and 20% for testing using 5-fold cross-validation. In the model training part, we conducted 30 times of training with a batch size of 64, chose to use the AdamW optimization algorithm, and set the initial value of the learning rate to 0.008 to obtain the optimal performance. The same data processing pipeline was applied consistently across all datasets in this study.

As shown in table 10, MAG-Res2Net demonstrated consistent advantages over ResNeXt, particularly on CSL-SHARE, and the MAG-Res2Net model training here is based on the premise of not using the optimization method mentioned in this article. Employing these methods imparted further performance gains, validating the versatility of MAG-Res2Net for diverse sensing modalities and contexts. In order to ensure the rigor of the experiment, we plotted the confusion matrix (figure 7) of the model on the data set, and additionally added the performance of some existing work on CSL-SHARE as a reference. It can be seen that our model is still in the leading position compared to traditional deep learning methods.

4.5. MAG-Res2Net training process analysis

The model training process with the highest F1-weight score over 40 runs was analyzed. The training and validation loss decreased as the number of iterations increased, with both curves converging after 10 iterations. The training accuracy plateaued at 98.5% by iteration 25, reaching 99% at iteration 30 and stabilizing thereafter. The validation accuracy culminated at 97.75%, demonstrating strong model performance on the WISDM dataset. Figure 8 shows the Loss and accuracy curves of the model when it is trained.

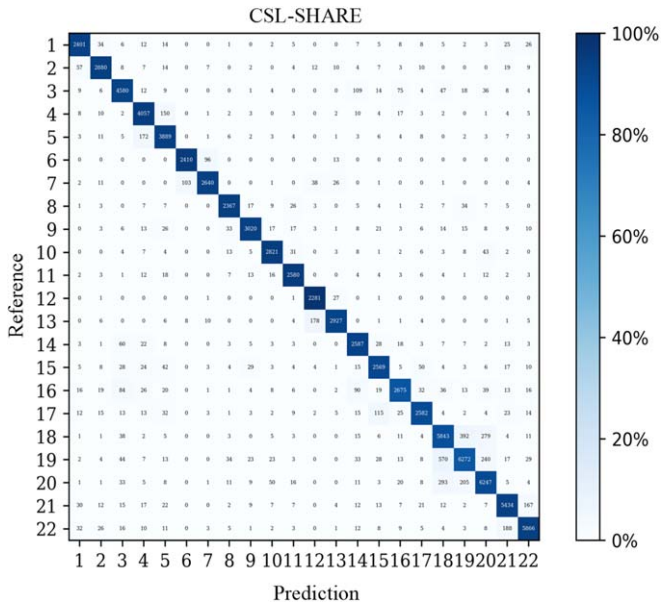

Figure 7. Confusion matrix (CSL-SHARE).

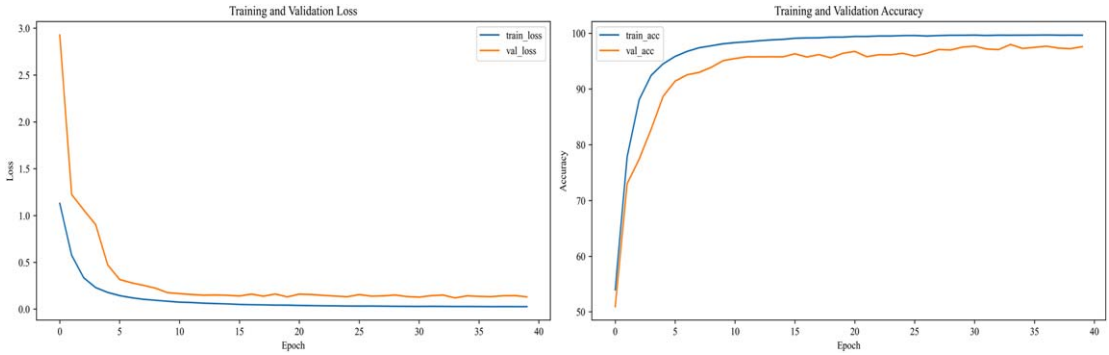

Figure 8. The Loss and Acc of the MAG-Res2Net model on WISDM.

4.5.1. Efficiency and time cost

Borderline-SMOTE substantially augments the sample count, particularly for WISDM, increasing computational costs. However, the Lion optimizer alleviated this large-scale training burden. Relative to the AdamW baseline, Borderline-SMOTE inflated the 40-iteration WISDM training time four-fold, mitigated to a two-fold slowdown with Lion, owing to doubled batch size. Thus, Borderline-SMOTE imposed minimal overhead on balanced datasets such as UCI-HAR. To quantify computational costs, we conducted comparative experiments on representative model architectures from the ablation study. The models were benchmarked against MAG-Res2Net configured with different methed influencing computational load, as shown in table 11.

The comparative experiments demonstrated the following trends regarding model training time:

(1) Incorporating gated modules incurred negligible computational overhead.

(2) The ECA attention mechanism was more efficient than SE modules while achieving better performance.

(3) Borderline-SMOTE improved model capabilities but substantially inflated training time.

(4) The Lion optimizer reduced Borderline-SMOTE's overhead by 50% versus AdamW while maintaining similar or higher performance.

In summary, gated modules and ECA attention were highly efficient techniques to enhance model performance. Borderline-SMOTE conferred major accuracy gains but with significant computational costs. However, Lion optimizer integration effectively amortized these expenses and enabled scalable training.

Table 11. Comparison of model training time and efficiency.

Model	Optimizer	Method	Training time	Performance ordering
Res2Net	AdamW (1024)	None	113 s	6
G-Res2Net	AdamW (1024)	None	125 s	5
G-Res2Net+SE	AdamW (1024)	None	156 s	4
MAG	AdamW (1024)	None	146 s	3
MAG	AdamW (1024)	Borderline	335 s	2
MAG	Lion (2048)	Borderline	214 s	1

These quantitative analyses provide valuable insights into the trade-offs between model accuracy versus efficiency for various architectural configurations and training strategies. The evidence-based findings can inform cost-aware model design and hyperparameter tuning tailored to specific applications and hardware constraints.

5. Conclusion

In this work, we have proposed a novel deep learning framework and MAG-Res2Net architecture for human activity recognition (HAR) tasks. Our methodology effectively addresses class imbalance while improving classification accuracy through Loss Combined. Moreover, the introduced model adaptively learns salient spatiotemporal features from input sequences, further enhancing performance. Rigorous benchmarking on UCI-HAR and WISDM datasets demonstrated accuracies of 94.44% and 98.32% respectively, outperforming prior arts in terms of accuracy, F1 metrics, generalization, and robustness. Our key contributions comprise the proposal of a set of optimization techniques, including new network architectures, upsampling algorithms, and loss function combinations. This framework effectively mitigates data imbalance issues while boosting classification accuracy. Our approach holds substantial potential for applications in disease monitoring, exercise tracking, rehabilitation, and beyond. This work provides significant insights to inform future research on optimizing HAR systems. Moving forward, we aim to further improve the model and explore deployments in practical contexts to maximize societal impact.

Acknowledgments

This work was supported by National Natural Science Foundation of China (62072089); Fundamental Research Funds for the Central Universities of China (N2116016, N2104001 and N2019007); National College Students Innovation and Entrepreneurship Training Program, Research on Human Behavior Recognition Based on Multi-scale Metric Learning Neural Network (202310145023).

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary information files).

ORCID iDs

Hanyu Liu https://orcid.org/0009-0002-0344-681X
Zhiqiong Wang https://orcid.org/0000-0002-0095-0378

References

Alemayoh T T, Lee J H and Okamoto S 2021 New sensor data structuring for deeper feature extraction in human activity recognition *Sensors* **21** 2814

Almaslukh B, Al Muhtadi J and Artoli A M 2018 A robust convolutional neural network for online smartphone-based human activity recognition *J. Intell. Fuzzy Syst.* **35** 1609–20

Anguita D, Ghio A, Oneto L, Parra X and Reyes-Ortiz J L 2013 A public domain dataset for human activity recognition using smartphones *ESANN* **3** 437–442

Bulling A, Blanke U and Schiele B 2014 A tutorial on human activity recognition using body-worn inertial sensors *ACM Comput. Surv.* **46** 1–33

Chen X *et al* 2023 Symbolic discovery of optimization algorithms arXiv;1–13

Dang L M, Min K, Wang H, Piran M J, Lee C H and Moon H 2020 Sensor-based and vision-based human activity recognition: a comprehensive survey *Pattern Recognit.* **108** 1–37
Dua N, Singh S N and Semwal V B 2021 Multi-input CNN-GRU based human activity recognition using wearable sensors *Computing* **103** 1461–78
Fan L, Wang Z and Wang H 2013 Human activity recognition model based on decision tree *2013 Int. Conf. on Advanced Cloud and Big Data* pp 64–8
Ferrari A, Micucci D, Mobilio M and Napoletano P 2021 Trends in human activity recognition using smartphones *J. Reliable Intell. Environ.* **7** 189–213
Folgado D, Barandas M, Famiglini L, Santos R, Cabitza F and Gamboa H 2023 Explainability meets uncertainty quantification: Insights from feature-based model fusion on multimodal time series *Inf. Fusion.* **100** 101955
Gadebe M L, Kogeda O P and Ojo S O 2020 Smartphone naïve Bayes human activity recognition using personalized datasets *J. Adv. Comput. Intell. Intell. Inf.* **24** 685–702
Gao S, Yao X and Chen Y 2019 Res2Net: a new multi-scale backbone architecture *IEEE Trans. Pattern Anal. Mach. Intell.* **43** 652–62
Gao W, Zhang L, Teng Q, He J and Wu H 2021 DanHAR: dual attention network for multimodal human activity recognition using wearable sensors *Appl. Soft Comput.* **111** 107728
Guan K, Shao M and Wu S 2017 A remote health monitoring system for the elderly based on smart home gateway *J. Healthcare Eng.* **2017** 1–11
Han H, Wang W-Y and Mao B-H 2005 Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning *Int. Conf. on Advances in Intelligent Computing* pp 878–87
Hartmann Y, Liu H, Lahrberg S and Schultz T 2022 Interpretable high-level features for human activity recognition *Proc. of the 15th Int. Joint Conf. on Biomedical Engineering Systems and Technologies* (https://doi.org/10.5220/0010840500003123)
Hartmann Y, Liu H, Lahrberg S and Schultz T 2023b Interpretable high-level features for human activity recognition *Presented at: 15th Int. Conf. on Bio-inspired Systems and Signal Processing* (https://doi.org/10.5220/0010840500003123)
Hartmann Y, Liu H and Schultz T 2021 Feature Space Reduction for Human Activity Recognition based on Multi-channel Biosignals
Hartmann Y, Liu H and Schultz T 2023a High-level features for human activity recognition and modeling *Commun. Comput. Inf. Sci.* **1814** 141–63
He K, Zhang X, Ren S and Sun J 2016 Deep residual learning for image recognition *Presented at: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition* (https://doi.org/10.1109/CVPR.2016.90)
Host K and Ivašić-Kos M 2022 An overview of human action recognition in sports based on computer vision *Heliyon* **8** e09633
Huang J, Lin S, Wang N, Dai G, Xie Y and Zhou J 2020 TSE-CNN: a two-stage end-to-end cnn for human activity recognition *IEEE J. Biomed. Health Inf.* **24** 292–9
Ignatov A 2019 Real-time human activity recognition from accelerometer data using convolutional neural networks *Appl. Soft. Comput.* **62** 6611–21
Jannat M K A, Islam M S, Yang S H and Liu H 2023 Efficient wi-fi-based human activity recognition using adaptive antenna elimination *IEEE Access.* **11** 105440–54
Khan N S and Ghani M S 2021 A survey of deep learning based models for human activity recognition *Wirel. Pers. Commun.* **120** 1593–635
Kwapisz J R, Weiss G M and Moore S A 2011 Activity recognition using cell phone accelerometers *SIGKDD Explorations* **12** 74–82
Li X, Zhao P, Wu M, Chen Z and Zhang L 2021 Deep learning for human activity recognition *Neurocomputing* **444** 214–6
Li Y and Wang L 2022a Human activity recognition based on residual network and BiLSTM *Sensors* **22** 635
Li Y and Wang L 2022b Human activity recognition based on residual network and BiLSTM *Sensors* **22** 1–13
Liu H 2021 Biosignal processing and activity modeling for multimodal human activity recognition [Internet]. Bremen: Suub; Nov [cited 2023 Aug 25]. Available from
Liu H, Gamboa H and Schultz T 2023 Sensor-based human activity and behavior research: where advanced sensing and recognition technologies meet *Sensors* **23** 125
Liu H, Hartmann Y and Schultz T 2021a CSL-SHARE: a multimodal wearable sensor-based human activity dataset *Front Comput. Sci.* **3** 759136 Internet]. [cited 2023 Aug 25]
Liu H, Hartmann Y and Schultz T 2021b Motion units: generalized sequence modeling of human activities for sensor-based activity recognition *2021 29th European Signal Processing Conf. (EUSIPCO), Dublin, Ireland* pp 1506–10
Lu L, Qing-ling C and Yi-Ju Z 2017 Activity recognition in smart homes *Multimedia Tools Appl.* **76** 24203–20
Mahmud S *et al* 2020 Human Activity Recognition from Wearable Sensor Data Using Self-Attention
Mekruksavanich S, Jantawong P and Jitpattanakul A 2022b A deep learning-based model for human activity recognition using biosensors embedded into a smart knee bandage *Proc. Comput. Sci.* **214** 621–7
Mekruksavanich S, Jitpattanakul A, Sitthithakerngkiet K, Youplao P and Yupapin P 2022a ResNet-SE: channel attention-based deep residual network for complex activity recognition using wrist-worn wearable sensors *IEEE Access.* **10** 51142–54
Nazari F, Nahavandi D, Mohajer N and Khosravi A 2021 Human activity recognition from knee angle using machine learning techniques *2021 IEEE Int. Conf. on Systems, Man, and Cybernetics (SMC)* pp 295–300
Vijayvargiya A, Kumari N, Gupta P and Kumar R 2021 Implementation of machine learning algorithms for human activity recognition *2021 3rd Int. Conf. on Signal Processing and Communication (ICPSC)* pp 440–4
Wang J, Chen Y, Hao S, Peng X and Hu L 2019 Deep learning for sensor-based activity recognition: a survey *Pattern Recognit. Lett.* **119** 3–11
Wang Q, Wu B, Zhu P, Li P, Zuo W and Hu Q 2020 ECA-Net: efficient channel attention for deep convolutional neural networks *IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)* (https://doi.org/10.1109/CVPR42600.2020.01155)
Wen Y, Zhang K, Li Z and Qiao Y 2016 A discriminative feature learning approach for deep face recognition *European Conf. on Computer Vision (ECCV)* pp 499–515
Xia K, Huang J and Wang H 2020 LSTM-CNN architecture for human activity recognition *IEEE Access.* **8** 56855–66
Xue T and Liu H 2022 Hidden markov model and its application in human activity recognition and fall detection: a review *Communications, Signal Processing, and Systems. Lecture Notes in Electrical Engineering* ed Q Liang *et al* (Springer) pp 863–9
Yang C, Jiang M, Guo Z and Liu Y 2020 Gated Res2Net for multivariate time series analysis *Int. Joint Conf. on Neural Networks (IJCNN)* pp 1–7
Yang C, Wang X, Yao L, Long G, Jiang J and Xu G 2022 Attentional gated Res2Net for multivariate time series classification *Neural Processing Letters* pp 1–14
Yang D, Huang J, Tu X, Ding G, Shen T and Xiao X 2019 A wearable activity recognition device using air-pressure and IMU sensors *IEEE Access.* **7** 6611–21