CoSS: Co-optimizing Sensor and Sampling Rate for Data-Efficient AI in Human Activity Recognition

Anonymous submission

Abstract

Recent advancements in Artificial Neural Networks have significantly improved human activity recognition using multiple time-series sensors. While employing numerous sensors with high-frequency sampling rates usually improves the results, it often leads to data inefficiency and unnecessary expansion of the ANN, posing a challenge for their practical deployment on edge devices. Addressing these issues, our work introduces a pragmatic framework for data-efficient utilization in HAR tasks, considering the optimization of both sensor modalities and sampling rate simultaneously. Central to our approach are the designed trainable parameters, termed 'Weight Scores,' which assess the significance of each sensor modality and sampling rate during the training phase. These scores guide the sensor modalities and sampling rate selection. The pruning method allows users to make a trade-off between computational budgets and performance by selecting the sensor modalities and sampling rates according to the weight score ranking. We tested our framework's effectiveness in optimizing sensor modality and sampling rate selection using three public HAR benchmark datasets. The results show that the sensor and sampling rate combination selected via CoSS achieves similar classification performance to configurations using the highest sampling rate with all sensors, but at a reduced hardware cost.

Introduction

Human Activity Recognition (HAR) based on Artificial Neural Networks (ANN) has achieved great success in the past years in many application areas like healthcare (Karar, Shehata, and Reyad 2022), fitness (Phukan et al. 2022), smart home (Bianchi et al. 2019) and smart manufacturing (Suh et al. 2023). The concurrent use of multiple sensors for human activity recognition provides a practical solution for complex activity recognition while improving recognition accuracy, robustness, and reducing noise (Aguileta et al. 2019). In addition, human activities consist of complex sequences and motor movements, whereas capturing these temporal dynamics is fundamental for successful human activity recognition (HAR) (Ordóñez and Roggen 2016). The high sampling rates can often contribute to better results, especially for complex activity recognition. However, the multiple-modality sensor-based solution and unnecessarily high sampling rate often require a larger Artificial Neural Networks (ANN) model for information fusion and higher power consumption for data acquisition and transmission, resulting in a challenging task towards practical deployment on edge devices where power conservation is crucial. While advancements in neural network compression, ranging from pruning (Han et al. 2015) and quantization (Nagel et al. 2021) to knowledge distillation (Hinton, Vinyals, and Dean 2015), have facilitated lightweight inference models on edge devices, the efficient utilization of sensor data for comprehensive AI model optimization, including sensor modalities and sample rates, remains a gap in current literature.

The purpose of sensor optimization is to remove sensors with minimal impact on classification results. Thus, the sensors' optimization can produce the benefit to reduce the input and computational complexity, while preserving the essential accuracy for HAR (Espinilla et al. 2017). Although reduced sampling rates imply more efficient resource utilization in real-world wearable HAR systems, existing works have predominantly overlooked the co-optimization of sensor modality selection and sampling rates.

When multiple sensors are used simultaneously, the sampling rate may vary between sensors depending on their nature and the type of data they capture. Optimizing these sampling rates involves not only considering each sensor individually but also how their data complements each other. Thus, we propose CoSS, a general framework focused on co-optimizing sensor modality and sampling rate for data-efficient AI in HAR tasks simultaneously requiring only training once. With CoSS, the trainable parameters "weight score" were designed to evaluate the importance of each sensor and sampling rate during training. The weight score ranking directs the selection of sensor modalities and sampling rates after training according to the trade of performance and hardware cost.

The main contributions of this paper can be summarized as follows.

- We propose a general efficient framework CoSS to address co-optimizing sensor and sampling rate for data-efficient AI in HAR tasks simultaneously requiring only training once.
- We demonstrate our framework's effectiveness in optimizing sensor modality and sampling rate selection using three public HAR benchmark datasets.

Related work

Sampling rate optimization

Research in HAR has explored various methods to optimize sampling rates, including adaptive frameworks (Yang et al. 2023), deep learning models (Qi et al. 2013), and sensor-specific algorithms (Anish et al. 2019). These approaches aim to achieve a balance between accuracy and efficiency in HAR systems. For instance, FreqSense (Yang et al. 2023), an adaptive resolution network, employed a subsampling strategy with conditional early exits, allowing variable resolution processing for different sample complexities. Although effective in balancing computational efficiency and accuracy, these networks are larger than typical neural networks, posing challenges for deployment on memory-limited edge devices. Additionally, while they reduce network complexity for simpler activities, their power consumption during data acquisition remains unchanged due to the need for high sampling rates for complex activities. Another proposed solution involves sensor-classifier co-optimization (Anish et al. 2019), but this adds hardware complexity and potential power usage. Hence, there is a gap in research regarding optimal overall sampling rates without requiring additional sensors.

Sensor optimization

The sensor optimization solution in the existing works includes the exhaustive search (Ertuğrul and Kaya 2017; Aziz, Robinovitch, and Park 2016), feature or classification selection (Leite and Xiao 2021; Tian and Zhang 2020) and adaptive-context-aware (Zappi et al. 2008), these methods often suffer from high computational cost and time-consuming at training phase. For instance, an exhaustive search method was applied in the work (Aziz, Robinovitch, and Park 2016) to identify the number and location of body-worn sensors for classifying walking, transferring, and sedentary activities; the computational cost of such a solution will be increased exponentially along with the number of the sensors to be optimized. Besides, this proposed method can not provide a ranking of the importance of the sensor modalities to guide users in selecting the sensor flexibly according to the hardware budgets. The proposed CoSS framework, while similar to Leite's work in sensor selection, introduces a more efficient approach. Leite's method uses a lightweight neural network and a sensor channel selection algorithm to rank and eliminate less important channels, requiring multiple training sessions, which is time and energy-intensive. In contrast, CoSS needs only a single training session, leveraging pruning technology to directly achieve sensor-pruned results. Additionally, while Leite's method evaluates sensor channel importance using $N \times (N-1)/2$ trainable weights, CoSS simplifies this with just N trainable weights, enhancing efficiency. Last but not least, the proposed CoSS framework focuses on the co-optimization of sensors and sampling rate together.

Proposed Framework

Fig. 1 shows the architecture of the proposed framework CoSS for sensor and sampling rate selection. The architecture of CoSS is designed based on the feature-level fusion architecture. Compared to the general feature-level fusion ANN model, three additional layers are integrated into the ANN model in CoSS architecture, such as resampling layers, sampling rate selection layers, and sensor selection layers, by which the importance ranking of sensor and sampling rate can be obtained by comparing the weight scores after training.

Resampling layer

The resampling layers in the system are engineered to produce multiple candidates at varied sampling rates. Each sensor node features a dedicated resampling layer that processes input data from sensors at the original sampling rate. This layer cycles through a set list of resampling rates, yielding several branches of down-sampled data. Notably, the sampling rates are highly flexible, allowing for arbitrary configurations, including fractional values. The resampling step size, S, is calculated based on the original sampling rate, $f_{original}$, and target sampling rate f_{target} as shown in Eq. (1).

$$S = \frac{f_{original}}{f_{target}} \tag{1}$$

However, since the step size (S) could be a float value, the generated index of original data for resampling can result in a non-integer value, which is illegal. To address this, we use linear interpolation to sample the target data (TD) from the original data (OD) according to Eq. (2).

$$\begin{aligned} TD[i] = OD[\lfloor i \times S \rfloor] + (i \times S - \lfloor i \times S \rfloor) \times \\ (OD[\lceil i \times S \rceil] - OD[\lfloor i \times S \rfloor]) \end{aligned} \tag{2}$$

where i is the new index of the target data range from zero to the length of the new target data.

Since the window size of the target data and original data is different after resampling layers, an adaptive kernel size (KS) was used in each feature extraction branch to guarantee that filters from different branches process the temporal information with equal time length. The adaptive kernel size is computed as follows.

$$KS_{target} = \lfloor KS_{original} \times \frac{f_{target}}{f_{original}} \rfloor \tag{3}$$

Thus, the sliding windows with a lower sampling rate have a smaller kernel size, leading to a smaller model size and lower computational load.

Sampling rate selection layer

The resampling rate selection layer is designed to rank the importance of different downsampling rates for classification tasks. We consider there are M sensor modalities and n_i downsampling rate candidates for $sensor_i$. Each $sensor_i$'s sampling rate selection layer comprises n_i trainable weights α, which indicates the importance of specific downsampling sampling rates to the final classification result. The extracted feature Fsr_i from the i_{th} sensor with different sampling rates is calculated according to Eq. (4).

$$Fsr_i = \sum_{j=1}^{n_i} F_{i,j} \times \frac{exp(\alpha_j)}{\sum_{k=1}^{n_i} exp(\alpha_k)} \tag{4}$$

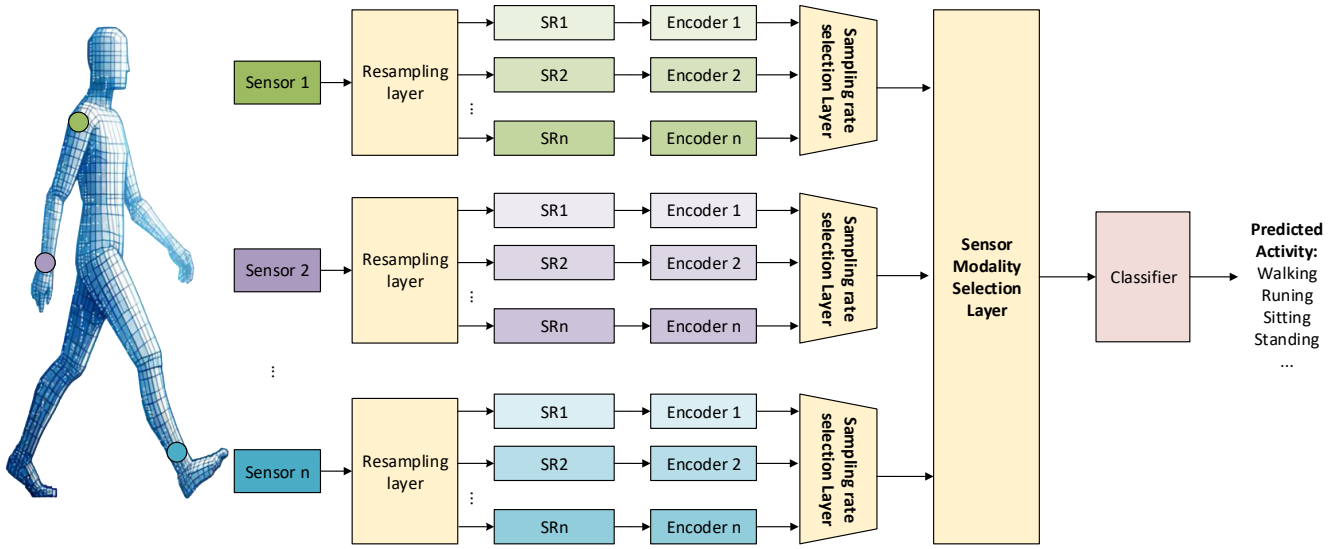


Figure 1: Proposed CoSS framework, the framework is based on multi-branch feature fusion architecture, where three additional layers, such as resampling layers, sampling rate, and sensor selection layers, are embedded. **SRn** is the downsampling data from the resampling layer

In the current literature, most of the results demonstrate that the higher sampling rate often leads to a better result. Thus, importance ranking and sampling rate could be closely related: candidates with higher sampling rates are supposed to have higher weight scores. During the sampling rate selection, sensors that are less sensitive to changes in the sampling rate are configured at a lower sampling frequency, by which the power consumption could be decreased compared to a uniform sampling rate configuration for all sensors.

Sensor selection layer

Similar to sampling rate selection, we assign M trainable weights α to M sensors, and the mixed features F_{mix} from the sensor selection layer are calculated as Eq. (5) shows.

$$F_{mix} = \sum_{i=1}^{M} Fsr_i \times \frac{exp(\alpha_i)}{\sum_{k=1}^{M} exp(\alpha_k)} \tag{5}$$

Thus, there are $M * N$ trainable weights α for both sampling rate and sensor selections in the proposed CoSS framework. The output of the sensor selection layer is fed to the fully connected layers for the final classification.

All trainable weights α can be obtained directly after model training. A progressive pruning method similar to the work (Han et al. 2015) is applied to remove the sensors with lower weights. The feature branch with the selected sampling rate data input is kept while the branches, including the rest of the sampling rate data, are pruned. The model's performance with the chosen sensors and sampling rate can be assessed without retraining the model from the beginning. In addition, fine-tuning can be used to improve the performance after sensor and sampling rate selection. In this work, we conducted an additional 10 training epochs for fine-tuning.

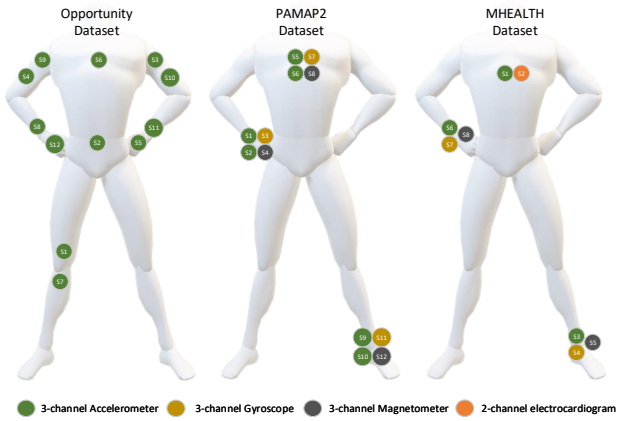


Figure 2: Sensor configurations of the three datasets, every sensor is annotated with an identifying number in this work

Evaluation

To evaluate our framework's effectiveness in optimizing sensor modality and sampling rate selection, the experiments on three public HAR benchmark datasets were conducted.

Dataset

Fig. 2 shows the sensor configuration of the three datasets. The Opportunity dataset is a collection of sensor data used for human activity recognition and ambient intelligence research. It features a range of sensors with original sampling rate of 30 Hz, like accelerometers and gyroscopes, placed on multiple subjects and in their environment. The dataset includes diverse daily activities. There are 5 classes for the locomotion recognition task and 18 classes for the gesture

recognition task. Only gesture recognition is tested in this work. The PAMAP2 dataset is designed for physical activity monitoring research. It comprises sensor data from accelerometers, gyroscopes, and magnetometers placed on the body of participants. The original sampling frequency of this dataset is 100 Hz. These participants, a diverse group, engaged in a variety of physical activities ranging from simple actions like walking and sitting to more vigorous exercises like running and cycling. The MHEALTH dataset is aimed at mobile health monitoring. It contains sensor data from accelerometers, gyroscopes, and ECG monitors attached to subjects. These subjects performed a series of activities, including both basic daily movements and fitness exercises. The original sampling rate of this dataset is 50 Hz.

Since the purpose of this work is not to discuss the neural network's performance across subjects or sessions, the leave-one-out cross-validation method was not utilized in this work. During the experiment, we divided all data from each dataset into three groups randomly: 70 % training dataset, 15 % validation dataset, and 15 % test dataset.

ANN Model

The neural network used for the classification tasks consists of two blocks: encoders and a classifier. The number of encoders equals $M * N$, M is the number of sampling rates to be evaluated and N is the number of sensors. The encoders consist of Convolution (CNN), Batch Normalization(BN), and ReLU activation function; the architecture of the encoder is as follows: $CNN - ReLU - BN - CNN - ReLU - BN - CNN - ReLU - Pooling - Flatten$. The filter number of each convolutional layer is 100. The original kernel sizes of the three experiments are 9 (Opportunity), 20 (PAMAP2) and 10 (MHEALTH). The classifier block consists of two dense layers. The cross-entropy loss function and SGD optimizer with a learning rate of 1e-2, a weight decay of 0.9, and a momentum of 1e-4, as well as a batch size of 512, were selected to train this model. The model was trained for 300 epochs with an early stopping using the patience of 30 to avoid overfitting. The ANN model was implemented in PyTorch. The experiments conducted on each dataset were repeated ten times to eliminate the possibility of randomness.

Experiment Result

The importance rankings of sensor modalities and sampling rate are obtained after training, according to which we can progressively prune unimportant sensor modalities and select the optimal sampling rate. Fig. 3 shows the weight score of the sensors in four experiments with different datasets, from which it can be observed that the weight of different sensors has different scores; the weight score of the sensor at the very end of the ranking are only around 0.05, while all the highest scores from the four experiments are above 0.15. Then, a progressive prune removed the sensors based on weight ranking from the neural networks, and the performance of the sensor-pruned model can be directly obtained by model inference. Fig. 4 shows the pruned result. It can be found that removing four sensors from the PAMAP2 and MHEALTH datasets has a negligible effect on the results, while the results are lightly affected by pruned sensors in the Opportunity dataset. However, the results are significantly improved after fine-tuning. We found that four, seven, and five sensors can be removed from the Opportunity-gestures, PAMAP2, and MHEALTH datasets accordingly, and the classification result has no obvious degradation after fine-tuning.

Fig. 5 shows weight scores of sampling rate candidates in every sensor. The result demonstrates that the candidates with higher sampling rates have a tendency to be assigned larger values, which is consistent with existing research. The weight score change rate with the decreasing sampling rate differs among the sensors in the same task. In addition, we found that the sensitivity of the sensors to the sampling rate is related to their importance for the classification task as Fig. 3 and Fig. 5 show. To select the optimal sampling rate combination among the sensors, we studied the performance sensitivity to the sampling rate of the selected sensors by keeping the feature branch of the studied sampling rate while pruning the rest of the sampling rate branches. The classification result can be obtained directly without the need to train again after pruning. Fig. 6 shows the result of the sampling rate optimization with the selected sensors. Overall, the performance has a very obvious degradation when the sampling rate of the most important sensor (with the highest weight score) is decreased, while the performance degradation of the sensors with less important weight is relatively smaller. For example, only the sensors $S4$ and $S9$ are sensitive to the sampling rate in the Opportunity-gesture task as Fig. 6a. Thus, we can keep these two sensors at the highest sampling rate and the rest of the sensors at the lowest sampling rate to save both hardware and power consumption. The classification performance in PAMAP2 and MHEALTH dataset has obvious degradation after the sampling rate of all sensors has been down to 60 Hz and 30 Hz accordingly as Fig. 6b and Fig. 6c shown.

The optimized sensors and sampling rate were selected based on the weight score ranking of sensors and their sampling rate, as well as classification results with the progressive prune method. Table 1 shows the result summary of the co-optimizing sensors and sampling rate. The most efficient result of sensor optimization is from the experiment on the MHEALTH dataset; five of eight sensors were removed, and around 62 % of the model size was reduced, while the performance was just reduced by around 3.19 % compared to the original model. It can be found from the sampling rate optimization result that the performance degradation is negligible in the PAMAP2 and MHEALTH datasets when the sampling rate of most of the reserved sensors was decreased. Overall, the optimized sensors and their sampling rate by CoSS can preserve the classification performance while reducing the memory (smaller model size), computational (smaller kernel and window size), and power (less and lower speed data transaction) costs.

Discussion

With the CoSS framework, the sensor and sampling rate selection can be guided by ranking the weight scores of the sensor and sampling rate as well as the progressive pruned

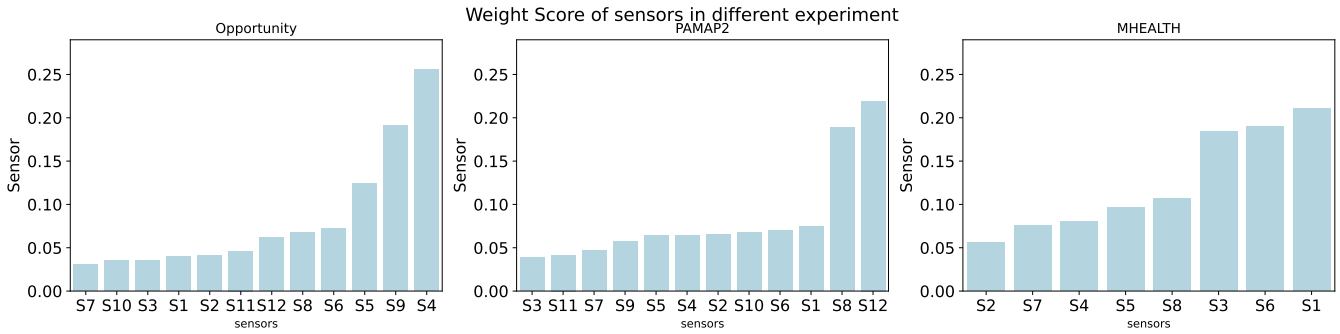
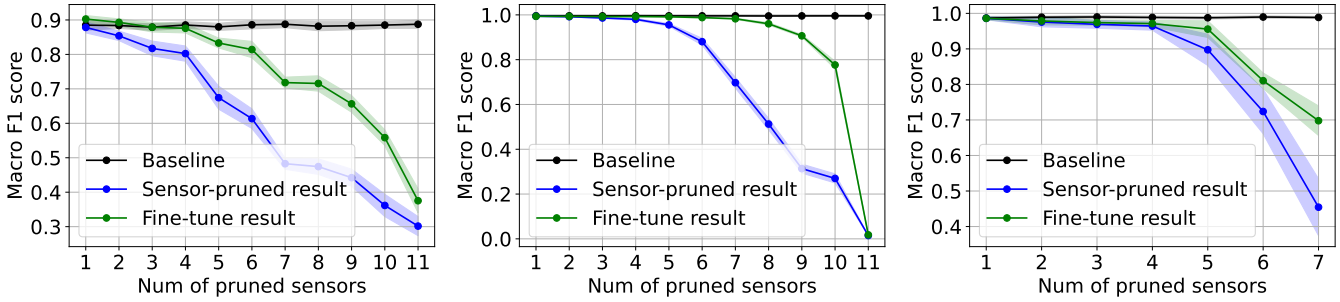


Figure 3: Weight score of sensors after training in different experiments



(a) Opportunity-gesture (b) PAMAP2 (c) MHEALTH

Figure 4: Sensor optimization result

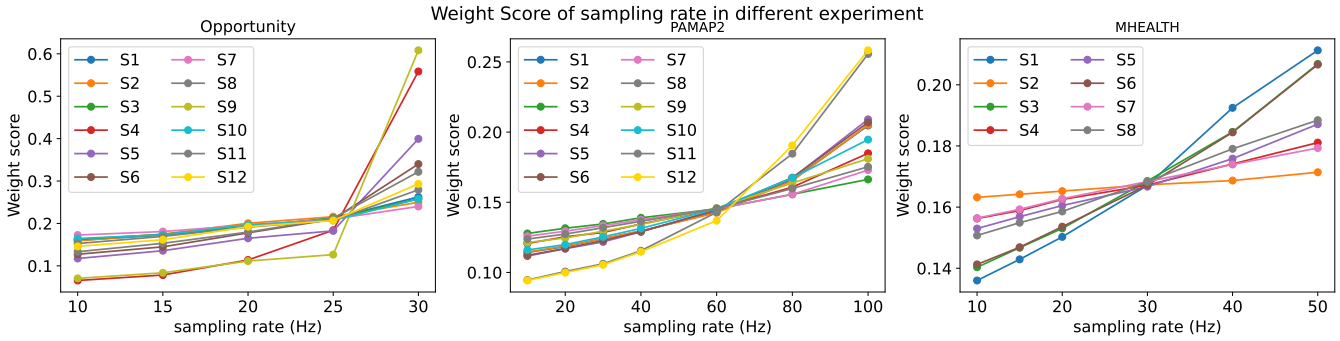
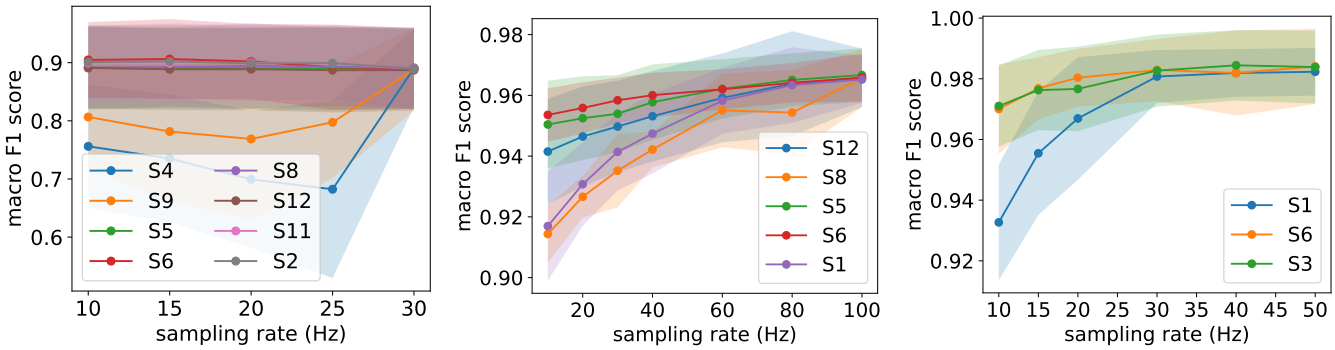


Figure 5: Weight score of sampling rate after training in different experiments

result, which demonstrated a more effective solution for selecting both sensor and sampling rate than an exhaustive search method. In addition, we found that the weight scores of sensors and their sampling rates are closely related during the co-optimization process. As shown in Fig. 7, the standard deviation was calculated from the weight scores of the sampling rate from each sensor to represent their sensitivity to the sampling rate, it can be found that the sensors with higher weight scores are more sensitive to the sampling rate in this work. Thus, assigning a higher sampling rate to these sensors with higher weight scores could be an optimal combination in multi-sensor-based HAR tasks. However, two major limitations of the proposed CoSS are also observed: Firstly, the number of feature extraction branches could be very large if there are many sensors and sampling rate candidates, which leads to the training process being more complex and time-consuming, although it just requires training once. Secondly, only the classification performance is considered in the framework during the sampling rate selection, which results in a higher sampling rate having a higher ranking; however, the hardware and computational cost of the CNN channel with different sampling rates is not considered, which should be integrated into the loss function and optimized during training in the future.



(a) Opportunity-gesture (b) PAMAP2 (c) MHEALTH

Figure 6: The classification performance of the sampling rate selection of the selected sensors (without fine-tuning)

Table 1: Result Summary

Experiment	Opportunity(gestures)	PAMAP2	MHEALTH
#Sensor	12	12	8
Sampling rates (Hz)	30,25,20,15,10	100,80,60,40,30,20,10	50,40,30,20,15,10
Window size (s)	3	2	4
Sensor pruned result (with original sampling rate)			
Pruned sensor	S7,S10,S3,S1	S3, S11,S7,S9,S5,S4,S2	S2,S7,S4,S5,S8
Performance reduction(%)	0.95 ± 1.09	1.30 ± 0.23	3.19 ± 2.23
Model size reduction(MB)	2.81/8.47(33%)	4.90/8.46 (58%)	**3.51/5.66 (62%)**
Sampling rate selection result (with selected sensors)			
Selected sampling rate (Hz)	30(S4),30(S9),10(S5),10(S6), 10(S8),10(S12),10(S11),10(S2)	80(S12),100(S8), 60(S5),40(S6), 80(S1)	30(S1), 20(S6), 30(S3)
Performance reduction(%)	1.85 ± 1.01	0.42 ± 0.22	0.29 ± 0.50

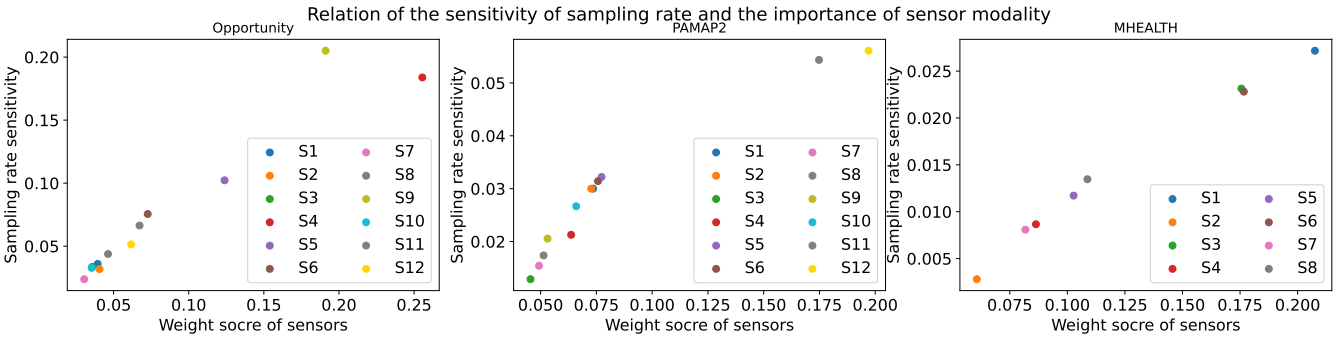


Figure 7: The relation of sampling rate sensitivity and the importance of sensor modality

Conclusion

In this work, we introduced the CoSS framework, designed to co-optimize sensors and sampling rates in HAR tasks for efficient utilization of sensor data. We proposed two types of trainable weight scores, integrated into a feature-level fusion artificial neural network (ANN). These scores assess the importance of both sensors and their sampling rates. By training the model once, we can rank these weight scores, guiding the progressive pruning of sensors and selecting the optimal sampling rate. To validate our framework, we tested its effectiveness in optimizing sensor modality and sampling rate selection on three public HAR benchmark datasets. The results show that the sensor and sampling rate combination selected via CoSS achieves similar classification performance to configurations using the highest sampling rate with all sensors but at a reduced hardware cost. In future work, a neural network architecture search method will be considered to be integrated into the framework, by which both input complexity and neural network complexity could be reduced in HAR task.

References

Aguileta, A. A.; Brena, R. F.; Mayora, O.; Molino-Minero-Re, E.; and Trejo, L. A. 2019. Multi-sensor fusion for activity recognition—A survey. *Sensors*, 19(17): 3808.

Anish, N.; Bhat, G.; Park, J.; Lee, H. G.; and Ogras, U. Y. 2019. Sensor-classifier co-optimization for wearable human activity recognition applications. In *2019 IEEE International Conference on Embedded Software and Systems (ICESS)*, 1–4. IEEE.

Aziz, O.; Robinovitch, S. N.; and Park, E. J. 2016. Identifying the number and location of body worn sensors to accurately classify walking, transferring and sedentary activities. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 5003–5006. IEEE.

Bianchi, V.; Bassoli, M.; Lombardo, G.; Fornacciari, P.; Mordonini, M.; and De Munari, I. 2019. IoT wearable sensor and deep learning: An integrated approach for personalized human activity recognition in a smart home environment. *IEEE Internet of Things Journal*, 6(5): 8553–8562.

Ertuğrul, Ö. F.; and Kaya, Y. 2017. Determining the optimal number of body-worn sensors for human activity recognition. *Soft Computing*, 21(17): 5053–5060.

Espinilla, M.; Medina, J.; Calzada, A.; Liu, J.; Martínez, L.; and Nugent, C. 2017. Optimizing the configuration of an heterogeneous architecture of sensors for activity recognition, using the extended belief rule-based inference methodology. *Microprocessors and Microsystems*, 52: 381–390.

Han, S.; Pool, J.; Tran, J.; and Dally, W. 2015. Learning both weights and connections for efficient neural network. *Advances in neural information processing systems*, 28.

Hinton, G.; Vinyals, O.; and Dean, J. 2015. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*.

Karar, M. E.; Shehata, H. I.; and Reyad, O. 2022. A survey of IoT-based fall detection for aiding elderly care: Sensors, methods, challenges and future trends. *Applied Sciences*, 12(7): 3276.

Leite, C. F. S.; and Xiao, Y. 2021. Optimal sensor channel selection for resource-efficient deep activity recognition. In *Proceedings of the 20th International Conference on Information Processing in Sensor Networks (co-located with CPS-IoT Week 2021)*, 371–383.

Nagel, M.; Fournarakis, M.; Amjad, R. A.; Bondarenko, Y.; Van Baalen, M.; and Blankevoort, T. 2021. A white paper on neural network quantization. *arXiv preprint arXiv:2106.08295*.

Ordóñez, F. J.; and Roggen, D. 2016. Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. *Sensors*, 16(1): 115.

Phukan, N.; Mohine, S.; Mondal, A.; Manikandan, M. S.; and Pachori, R. B. 2022. Convolutional neural network-based human activity recognition for edge fitness and context-aware health monitoring devices. *IEEE Sensors Journal*, 22(22): 21816–21826.

Qi, X.; Keally, M.; Zhou, G.; Li, Y.; and Ren, Z. 2013. AdaSense: Adapting sampling rates for activity recognition in body sensor networks. In *2013 IEEE 19th Real-Time and Embedded Technology and Applications Symposium (RTAS)*, 163–172. IEEE.

Suh, S.; Rey, V. F.; Bian, S.; Huang, Y.-C.; Rožanec, J. M.; Ghinani, H. T.; Zhou, B.; and Lukowicz, P. 2023. Worker Activity Recognition in Manufacturing Line Using Near-body Electric Field. *IEEE Internet of Things Journal*.

Tian, Y.; and Zhang, J. 2020. Optimizing sensor deployment for multi-sensor-based HAR system with improved glowworm swarm optimization algorithm. *Sensors*, 20(24): 7161.

Yang, G.; Zhang, L.; Bu, C.; Wang, S.; Wu, H.; and Song, A. 2023. FreqSense: Adaptive Sampling Rates for Sensor-Based Human Activity Recognition Under Tunable Computational Budgets. *IEEE Journal of Biomedical and Health Informatics*.

Zappi, P.; Lombriser, C.; Stiefmeier, T.; Farella, E.; Roggen, D.; Benini, L.; and Tröster, G. 2008. Activity recognition from on-body sensors: accuracy-power trade-off by dynamic sensor selection. In *Wireless Sensor Networks: 5th European Conference, EWSN 2008, Bologna, Italy, January 30-February 1, 2008. Proceedings*, 17–33. Springer.