**Response to review comments**

**Reviewer #1:**

The revised manuscript is well modified. In the revised version, the suggestions are answered and modified.

The revised manuscript can be accepted if it can be considered as follows: Futher work should be improved in Conclusions.

**Reply to Reviewer #1:**

We are deeply grateful to the reviewers for their valuable feedback and thoughtful recommendations, which have been instrumental in strengthening our paper.

1. Futher work should be improved in Conclusions.

**Reply:** Thank you for the comments. Further improvements are listed in the conclusion from 3 perspectives: ensemble learning novelty, dataset and domain segmentation and visualization. For ensemble learning novelty, we explain how DSME can be extended within the ensemble learning framework. For dataset and domains segmentation, a self-collected private dataset and quantitative analysis among different domains may be necessary in the future. For visualization, latent features from hidden layers need visualizing to further validate the method is efficient. **The revised future improvements is on Page 10~11.**

“This study still has limitations and can be improved in the future. **Improvements of ensemble learning novelty:** DSME enables experts to learn from their responsible domains at the same time and gains a final prediction generated through joint efforts of experts and domain-sniff module. Firstly, instead of generating classifications from the same label space, the generation of DSME can be improved by having each expert learn a part of the label space. The first expert, for example, may only learn ``Running'' activity from all domains, thus DSME is extended to label specific generation besides domain specific generation. Secondly, the integration of DSME can be improved by integrating latent features such as features from convolution layers or fully connected layers instead of only integrating the classifications of experts. **Improvements of dataset and domain segmentation:** Firstly, DSME is tested on 4 public datasets and can also be tested on a self-collected and weakly labeled dataset to validate the robustness of DSME. Secondly, domain segmentation in DSME is done manually and can also be improved by introducing methods to conduct a quantitative analysis of the difference in the same activity across different domains, thus the real calculated number of domains can be reasonable for training and test. **Improvements of visualization:** Weights from different experts are visualized to show the relative importance among experts. Besides, latent features in terms of classification from each expert should also be visualized to assess if the expert has adequately fitted its responsible domain. ”

**Reviewer #2:**

The paper proposes a Domain-Specific Mixture of Experts (DSME) model, specifically designed to address the issue of generalizing sensor-based human activity recognition across individuals, aiming to tackle the variation in data distributions between different people. The topic is interesting. However, the following issues need improvement:

1. The paper's proposed model only provides information on the number of parameters but does not include a comparison of the model's inference time.

2. The paper compares domain generalization methods (GILE) for HAR from three years ago, but the comparison with methods from the last three years does not focus on domain generalization techniques for HAR. There is insufficient comparison with recent domain generalization (DG) methods specifically applied to sensor-based activity recognition in the past three years

3. The "Backbones and Compared Works" section lacks proper citations.

4. There are instances of non-standard fonts: non-variable and meaningful subscripts should not be italicized in the formulation.

**Reply to Reviewer #2:**

Thank you for your comments. We particularly appreciate the suggestion to “recent domain generalization (DG) methods specifically applied to sensor-based activity recognition”. We have seriously checked the four points and revised our manuscript according to your suggestion.

1. The paper's proposed model only provides information on the number of parameters but does not include a comparison of the model's inference time.

**Reply:** Thanks for your comments. Table 1 shows the average inference time between baseline and DSME. Table 1 indicates DSME does have more computations and is time consuming compared to baseline so as to guarantee higher average classification performance so more inference time may be necessary. Besides, DSME does not intend to position lower computational performance as a core competitive advantage. As a result, Table III on Page 7 deletes the number of trainable parameters and keeps the average classification performance.

Table 1 Average inference time between baseline and DSME

表格

描述已自动生成

**Table III on Page 7：**

“

Table III Average weighted F1 score (%) on DS methods compared with baselines. “\” indicates the method perform by itself. “+DS” means the method serves as the backbone of DSME.

**表格

描述已自动生成**

”

1. The paper compares domain generalization methods (GILE) for HAR from three years ago, but the comparison with methods from the last three years does not focus on domain generalization techniques for HAR. There is insufficient comparison with recent domain generalization (DG) methods specifically applied to sensor-based activity recognition in the past three years

**Reply:** We appreciate the comments. SDMIX[1] and DIVERSIFY[2] are added for comparison.

[1] W. Lu, J. Wang, Y. Chen, S. J. Pan, C. Hu, and X. Qin, “Semantic-discriminative mixup for generalizable sensor-based cross-domain activity recognition,” Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 6, no. 2, pp. 1–19, 2022.

[2] W. Lu, J. Wang, X. Sun, Y. Chen, X. Ji, Q. Yang, and X. Xie, “Diversify: A general framework for time series out-of-distribution detection and generalization,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024.

SDMIX is a data augmentation method designed for domain generalization in sensor-based HAR. It considers the activity semantic ranges to overcome the semantic inconsistency brought by domain differences through aggregating two samples into a new one with calculated coefficients. DIVERSIFY is a general framework designed for domain generalization in sensor-based HAR. DIVERSIFY is an overall framework including domain detection and domain generalization designed for sensor-based HAR. DIVERSIFY obtains the pseudo number of domains through quantitative analysis and then encourages the model to learn domain-invariant features to enhance generalization capability.

Table 1 Respective and average weighted F1 score (\%) of compared works.

Revised result analysis on Page 8~9:

图片包含 图形用户界面

描述已自动生成

Table 1 shows respective and average weighted F1 score (\%) of compared works. In DSA, DSME with MTSDNet backbone has a 1.56% average F1 score gain compared with GRU-INC. In addition, DIVERSIFY outperforms DSME with MTSDNet backbone by 0.95% average weighted F1 score. This may be because DIVERSIFY considers the quantitative difference among domains and assumes the pseudo number of domains instead of manually dividing domains into a fixed number as DSME has done. For each test domain in DSA, DSME with DanHAR backbone gets the highest F1 score on domain 4, while DSME with MTSDNet backbone receives the highest F1 score on domain 7. In PAMAP2 and OPP, DSME with MTSDNet backbone achieves 1.61%, 1.67% and 0.29% average performance gain compared with TCCSNet, GRU-INC and DIVERSIFY, respectively. Especially in OPP, DSME with MTSDNet backbone gets the best performance on domain 1, 2 and 4. In UCIHAR where domains are similar with each other, DSME with Tri-Res backbone gets 0.54%, 2.79% and 1.56% performance gain compared with TCCSNet, GRU-INC and DIVERSIFY. DIVERSIFY treats the pseudo number of domains as a hyperparameter, thus requiring more fine-tuning of parameter combinations compared to DSME to achieve optimal performance. The results again prove that DSME is efficient in aggregating the source domains and relieve domain shift on target domain. **The reasons for selecting SDMIX and DIVERSIFY is added to the top left corner of Page 7**. **The result analysis is also revised on Page 8~9.**

**Reasons for selecting SDMIX and DIVERSIFY on Page 7:**

“SDMIX is a data augmentation method designed for domain generalization in sensor-based HAR. It takes into account the activity semantic ranges to overcome the semantic inconsistency brought by domain differences through aggregating two samples into a new one with calculated coefficients. DIVERSIFY is a general framework designed for domain generalization in sensor-based HAR. DIVERSIFY is an overall framework including domain detection and domain generalization designed for sensor-based HAR. DIVERSIFY obtains the pseudo number of domains through quantitative analysis and then encourages the model to learn domain-invariant features to enhance generalization capability.”

**Revised analysis on Page 8~9:**

“Table IV shows the comparison between DSME with three backbones (DanHAR, Tri-Res, MTSDNet) and five state-of-the-art works (CORAL, DANN, ConvLSTM, RSC, GILE, TCCSNet, GRU-INC, ELK, SDMIX, DIVERSIFY). The LODO strategy test every domain and then calculates the average weighted F1 score. ``DS'' prefix means DSME with a specific backbone. As shown in Table IV, DSME reaches a significant increase on all the domains. In DSA, DSME with MTSDNet backbone has a 1.56% average F1 score gain compared with GRU-INC. In addition, DIVERSIFY outperforms DSME with MTSDNet backbone by 0.95% average weighted F1 score. This may be because DIVERSIFY considers the quantitative difference among domains and assumes the pseudo number of domains instead of manually dividing domains into a fixed number as DSME has done. For each test domain in DSA, DSME with DanHAR backbone gets the highest F1 score on domain 4, while DSME with MTSDNet backbone receives the highest F1 score on domain 7. In PAMAP2 and OPP, DSME with MTSDNet backbone achieves 1.61%, 1.67% and 0.29% average performance gain compared with TCCSNet, GRU-INC and DIVERSIFY, respectively. Especially in OPP, DSME with MTSDNet backbone gets the best performance on domain 1, 2 and 4. In UCIHAR where domains are similar with each other, DSME with Tri-Res backbone gets 0.54%, 2.79% and 1.56% performance gain compared with TCCSNet, GRU-INC and DIVERSIFY. DIVERSIFY treats the pseudo number of domains as a hyperparameter, thus requiring more fine-tuning of parameter combinations compared to DSME to achieve optimal performance. The results again prove that DSME is efficient in aggregating the source domains and relieve domain shift on target domain.”

1. The "Backbones and Compared Works" section lacks proper citations.

**Reply:** Thanks for your suggestions. The citations in the “Backbones and Compared Works” section are added. **The revised citation is on Page 6~7.** **All the compared works are also cited as follows**:

For backbones, DanHAR [3], Tri-Res [4] and MTSDNet [5] are selected as the backbone of the proposed method. For compared works, CORAL [6], DANN [7], ConvLSTM [8], RSC [9], GILE [10], TCCSNet [11], GRU-INC[12], ELK [13], SDMIX [1] and DIVERSIFY [2] are picked.

[1] W. Lu, J. Wang, Y. Chen, S. J. Pan, C. Hu, and X. Qin, “Semantic-discriminative mixup for generalizable sensor-based cross-domain activity recognition,” Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 6, no. 2, pp. 1–19, 2022.

[2] W. Lu, J. Wang, X. Sun, Y. Chen, X. Ji, Q. Yang, and X. Xie, “Diversify: A general framework for time series out-of-distribution detection and generalization,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024.

[3] W. Gao, L. Zhang, Q. Teng, J. He, and H. Wu, “Danhar: Dual attention network for multimodal human activity recognition using wearable sensors,” Applied Soft Computing, vol. 111, p. 107728, 2021.

[4] Y. Tang, L. Zhang, Q. Teng, F. Min, and A. Song, “Triple cross-domain attention on human activity recognition using wearable sensors,” IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 6, no. 5, pp. 1167–1176, 2022.

[5] J. Pan, Z. Hu, L. Zhang, and X. Cai, “Multi-channel time series decomposition network for generalizable sensor-based activity recognition,”IEEE Transactions on Automation Science and Engineering, pp. 1–12, 2024.

[6] B. Sun and K. Saenko, “Deep coral: Correlation alignment for deep domain adaptation,” in Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III 14. Springer, 2016, pp. 443–450.

[7] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. March, and V. Lempitsky, “Domain-adversarial training of neural networks,” Journal of machine learning research, vol. 17, no. 59, pp. 1–35, 2016.

[8] F. J. Ordonez and D. Roggen, “Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition,” Sensors, vol. 16, no. 1, p. 115, 2016.

[9] Z. Huang, H. Wang, E. P. Xing, and D. Huang, “Self-challenging improves cross-domain generalization,” in Computer Vision–ECCV 2020:16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16. Springer, 2020, pp. 124–140.

[10] H. Qian, S. J. Pan, and C. Miao, “Latent independent excitation for generalizable sensor-based cross-person activity recognition,” in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, no. 13, 2021, pp. 11 921–11 929.

[11] E. Essa and I. R. Abdelmaksoud, “Temporal-channel convolution with self-attention network for human activity recognition using wearable sensors,” Knowledge-Based Systems, vol. 278, p. 110867, 2023.

[12] T. R. Mim, M. Amatullah, S. Afreen, M. A. Yousuf, S. Uddin, S. A. Alyami, K. F. Hasan, and M. A. Moni, “Gru-inc: An inception-attention based approach using gru for human activity recognition,” Expert Systems with Applications, vol. 216, p. 119419, 2023

[13] M. Yao, L. Zhang, D. Cheng, L. Qin, X. Liu, Z. Fu, H. Wu, and A. Song,“Revisiting large-kernel cnn design via structural re-parameterization for sensor-based human activity recognition,” IEEE Sensors Journal, 2024.

**Revised citation on Page 6:**

“For backbones, DanHAR [9], Tri-Res [10] and MTSDNet [31] are selected as the backbone of the proposed method.”

**Revised citation on Page 7:**

“For compared works, CORAL [32], DANN [15], ConvLSTM [28], RSC [37], GILE [35], TCCSNet [11], GRU-INC [13], ELK [29], SDMIX [34] and DIVERSIFY [46] are picked.”

4. There are instances of non-standard fonts: non-variable and meaningful subscripts should not be italicized in the formulation.

**Reply:** We appreciate the comments. We have revised the paper thoroughly to change the subscripts of constants from italic to regular font. **The revised subscripts is on Page 4~6.**

**Revised formulations on Page 4~5:**

“

图片包含 公司名称

描述已自动生成

手机屏幕的截图

中度可信度描述已自动生成



”

**Revised algorithms on Page 5~6:**

文本

描述已自动生成

文本

描述已自动生成