A survey of attention mechanisms for wearable sensor-based human activity recognition

Xin Wang[1], Yan Wang[1,*], Yingrui Geng[1], Hongnian Yu[2], Hongmei Yang[1], Xiaoxu Wen[1] and Aihui Wang[1]

[1]*School of Electric and Information, Zhongyuan University of Technology, Zhengzhou 450007, China*

[2]*School of Computing, Engineering and the Built Environment, Edinburgh Napier University, Edinburgh EH10 5DT, UK*

Abstract

Attention mechanisms, widely used in many fields such as computer vision (CV) and natural language processing (NLP), enable deep learning networks to extract more important information from the input, thereby improving performance and efficiency. Recently, attention mechanisms are introduced to wearable sensor-based human activity recognition (WSHAR) for learning more robust feature representations. This paper investigates the attention mechanisms in WSHAR with a special focus on the principles of computing attention and the targets on which the attention works in a network and future directions. The aim is to provide readers with a clearer understanding of attention mechanisms in WSAHR and motivate more diverse work in the future.

Keywords

Attention mechanisms, deep learning, human activity recognition, wearable sensors

1. Introduction

Recently, attention mechanisms have become enormously popular in deep learning [1, 2, 3]. Researchers apply attention mechanisms to networks, allowing models to dynamically focus on key parts of the input to perform specific tasks more effectively. The basic principle of the attention mechanism is to weigh the input information so that parts of input with higher weights are considered more relevant to the task and have a greater impact on the model's decisions [4]. Attention mechanisms were first introduced into the encoder-decoder network for natural language processing (NLP) [5], enabling the decoder to selectively access the input sequence parts that are important to the context. Vaswani et al. [6] performed an innovative development on attention mechanisms, where they relied solely on a self-attention mechanism to model the global dependencies of the input. The mechanism overcame the difficulties such as performance degradation and computational inefficiency caused by recurrent neural networks

(RNNs) as the length of the input sequence increased, and showed state-of-the-art results on NLP tasks when it was proposed. Another mainstream application scenario of attention mechanisms is in computer vision (CV) for image classification [7, 8], anomaly detection [9, 10] or semantic segmentation [11, 12]. Attention mechanisms in CV can effectively solve the problem of image information overload, which contributes to saving computational resources [13].

In the past decade, with the advances in Artificial Intelligence (AI) and sensor technologies, human activity recognition (HAR) is gradually playing an important role in many cross-cutting areas, such as smart healthcare [14], motion monitoring [15] and human-robot interaction [16]. Depending on the type of sensors arranged, HAR can be classified into three categories: 1) vision sensor-based HAR (VSHAR), 2) ambient sensor-based HAR (ASHAR) and 3) wearable sensor-based HAR (WSHAR) [17]. The WSHAR systems are more portable, require less computational resources. They overcome the limitation of working only in specified areas which cameras and ambient sensors have, thus becoming a research hotspot in the field of HAR recently. Learning robust feature representations from raw wearable sensor data is critical for WSHAR tasks. As an effective and powerful performance-enhancing network, attention mechanisms have also been introduced into WSHAR to learn more valuable features from raw signals [18, 19]. Applying attention mechanisms to raw data from different wearable sensors enables the model to focus on the data that contribute more to activity recognition, avoiding the effect of noise caused by individual faulty sensors. For example, Mahmud et al. [20] proposed the sub-module named Sensor Modality Attention to weigh the data captured from different sensor modalities according to their varying contribution levels. The weighted representations obtained showed better performance than the raw data. Combining attention mechanisms with some typical neural networks, such as convolutional neural networks (CNNs) and long short-term memory network (LSTM), can also help improve the performance in handling HAR tasks. Sun et al. [21] introduced an attention layer into the LSTM to automatically capture the important temporal dependencies. Singh et al. [22] utilized self-attention to select and learn important time points of spatial-temporal features captured by the combination of CNN and LSTM, which achieved a significant enhancement over the existing methods.

Although several papers have reviewed the attention mechanisms or their applications in natural language processing [23, 24] and computer vision [25, 26], few can be found in WSHAR. The use of attention mechanisms in WSHAR tasks differs in the calculation principle and the target that the attention works on in a network. This paper thus provides a concise review and discussion on the applications of attention mechanisms in WSHAR. The main contributions of this work are as follows.

- Identifying the attention mechanisms in WSHAR according to the principles of computing attention and presenting the theoretical explanations accordingly;
- Exploring the applications of attention mechanisms in WSHAR tasks according to the targets that the attention works on in a network, and mining the corresponding motivations and reasons behind;
- Providing the future directions of attention mechanisms in WSHAR, i.e., exploring the feasibility of introducing the cross-attention mechanism to perform WSHAR tasks.

The rest of the paper is organized as follows: Section 2 provides theoretical principles of the generic attention mechanisms in WSHAR. Section 3 identifies and summarizes the three

attention methods in WSHAR tasks. Section 4 discusses the future directions of attention-based WSHAR. Finally, the conclusion is presented in Section 5.

2. Attention mechanism theoretical principles

Thanks to the high compatibility, and excellent performance in feature and model optimization, attention mechanisms have been acquiring great success in WSHAR and a large number of attention-based HAR models have emerged in a short period of time. According to the principles of computing attention, there are two basic attention mechanisms in WSHAR: weight-based attention and self-attention. Before looking into the attention methods in WSHAR, we detail the two types of attention mechanisms in this section.

2.1. Weight-based attention

Extracting important features in the input elements helps further enhance the performance of HAR models. The weight-based attention learns the importance of the input elements. It assigns different weight coefficients to each element, thus increasing the proportion of important elements in feature extraction and reducing or eliminating unimportant elements. As illustrated in Figure 1, the input elements x^i are first fed into a specific network for features learning to obtain the attention representations a^i. Then a softmax operation is performed on a^i to produce the attention scores α^i. Finally, the obtained attention scores α^i are fused with the input elements to derive the weighted attention output. Usually, there are two ways of fusion, one is to directly fuse the attention scores with the original input elements x^i to obtain the weighted input vectors c^i, and the other is to apply the attention scores to the attention representations a^i to get the weighted attention representations o^i.
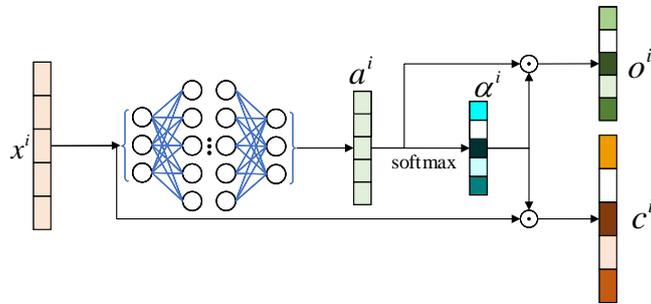

Figure 1: Wight-based attention

2.2. Self-attention

The weight-based attention calculation process is shown as:

$$a^i = \sigma \cdot \left(W x^i + b \right) \tag{1}$$

$$\alpha^i = \text{softmax}\left(a^i\right) = \frac{\exp\left(a_k^i\right)}{\sum\limits_{k=1}^{n} \exp\left(a_k^i\right)} \tag{2}$$

$$c^i = \alpha^i \cdot x^i \tag{3}$$

$$o^i = \alpha^i \cdot a^i \tag{4}$$

where W and b in (1) are the parameters learned by the network when linear transformations are applied to x^i, and σ denotes the nonlinear activation function that help extract the nonlinear features of inputs.

The weight-based attention has various variants in HAR due to different element objects, action ranges and network structures. According to the action element object, it can be divided into spatial attention, temporal attention, etc. The networks designed to capture attention representations are diverse, such as fully connected layers, CNNs, RNNs.

Self-attention is a mechanism proposed by Vaswani et al. [6] in Transformer. Unlike weight-based attention, self-attention contains both attention score calculation and feature extraction, i.e., self-attention enables the extraction of features from inputs independently, as shown in Figure 2.
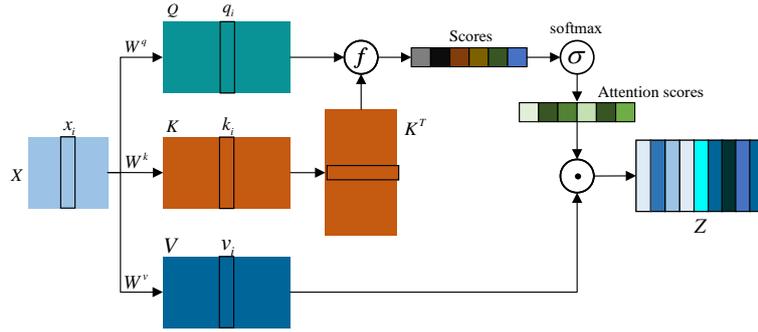

Figure 2: Self-attention [6]

The calculation of self-attention consists of three primary operations. Firstly, applying linear transformations to the input X to obtain the transformation matrices Query (Q), Key (K) and Value (V). x_i, q_i, k_i and v_i are the elements of X, Q, K and V, respectively. W^q, W^k and W^v are the trainable parameter matrices. Secondly, comparing q_i with each k_i, which means conducting matrix operations on Q and the transpose of K to get the scores as follows:

$$f = \frac{Q \cdot K^T}{\sqrt{d_k}} \tag{5}$$

where d_k represents the dimension of K and the scaled operation leads to having more stable gradients. Then the results are normalized by softmax to obtain the attention scores. Finally,

multiply the attention scores with V to give the final output Z. The entire formula for self-attention can be expressed as:

$$Z(Q, K, V) = \text{softmax}(\frac{Q \cdot K^T}{\sqrt{d_k}}) \cdot V \tag{6}$$

The multi-head attention proposed in [6] performs multiple self-attention operations on the input in parallel. Then it stitches the output of each self-attention together to extract the information obtained by multiple self-attention through an additional parameter matrix W^o, as defined in (7).

$$MZ(Q, K, V) = W^o \cdot \text{Concat}(Z_1, Z_2, \dots, Z_h) \tag{7}$$

where Z_i indicates the output of the i-th self-attention and h is the number of self-attention. Multi-head attention enables the extraction of more valuable features from different views, which enhances model's performance and thus is widely used in WSHAR.

3. Attention methods in WSHAR

Section 2 detials the attention mechanisms based on computational principles. Both attentions can enhance the information processing ability of neural networks. Factors such as the target objects or the binding networks contribute to the diversity of applications of both attention mechanisms in WSHAR. In this section, we review the attention methods in HAR according to the targets that the attention works on in a network and give explanations of the function of each attention applied in a specific WSHAR task, with a summary of the related works in Table 1.

3.1. Temporal attention

Activity data from wearable sensors are time series signals with high time dependence [41]. Extracting temporal features from raw sensor data is crucial to improve the recognition performance. Although RNNs and their variants, such as LSTM and gated recurrent unit (GRU), specialize in learning the sequential dependencies [42, 43, 44], they treat all time steps of sensor data equally. It means that noise or unimportant signals are also fed into models for training, which may lead to a degradation of recognition accuracy. The temporal attention mechanism weights the time steps in the input or obtained representations to make the temporal models mentioned above that focus on the information extraction on important time steps [45]. Therefore, a model with temporal attention is able to learn more robust temporal features from the raw sensor data or representations and exhibits high efficiency [46]. Figure 3 presents the generic structure of temporal attention in WSHAR, where S and C represent the number of samples and sensor channels, respectively.

Haque et al. [27] introduced a temporal attention named hierarchical context-based attention to the two-layer GRU model to learn the hierarchy of temporal features. The attention generates diverse importance for temporal features learned bu GRU, effectively capturing the context of relevant time steps. Similarly, Betancourt et al. [29] added a two-layer self-attention after an LSTM layer to improve the model's performance. Self-attention layer compares each time

Table 1
Typical studies in attention-based WSHAR

| Category	Method description	Reference
Temporal attention	Conducting weight-based attention on the temporal features generated by GRU	[27], [28]
	Conducting self-attention or weight-based attention on the temporal features learned by LSTM	[29], [30], [31]
	Performing self-attention on the representations of ConvLSTM model from the temporal dimension	[22], [32]
Spatial attention	Executing weight-based attention on the spatial features generated by different CNN three times in succession	[33]
	Performing weight-based attention on the 3D spatial features learned by CNN	[34]
Temporal & spatial attention	Executing self-attention on the temporal features generated by GRU and spatial features generated by CNN, respectively	[35]
	Conducting weight-based attention on the spatial or temporal features generated by CNN	[36], [37], [38]
	Executing weight-based attention on the input and model-generated representations from the spatial dimension or the temporal dimension, respectively	[39], [40]
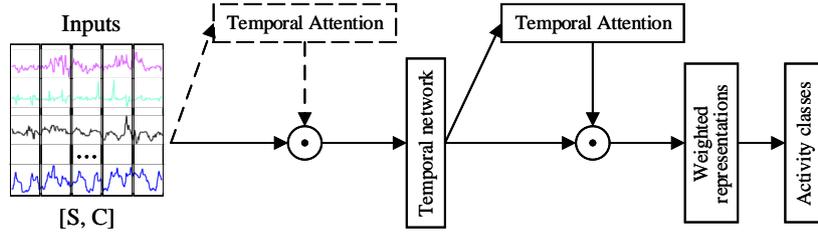
Figure 3: Temporal attention-based model

step in the representations learned by the last LSTM layer with all time steps to determine the more relevant time steps. The obtained relevance helps to extract robust time dependencies. The LSTM network with the attention module improved the accuracy by 4.0% and 4.2% on the UCI-HAR and MTUT-HAR datasets, respectively. Singh et al. [22] used self-attention to further learn the spatial-temporal representations generated by CNN and LSTM. The ConvLSTM model with self-attention shows better performance on all six publicly available HAR datasets than those without self-attention.

3.2. Spatial attention

Wearable sensors such as accelerometers, gyroscopes or magnetometers capture different information about the same activity and bring different contributions to activity recognition [47].

The involvement of different body parts during the execution of an activity is diverse, so the contribution of data collected from different positions to activity recognition also varies [48]. Hence, activity data from wearable sensors are not only rich in time dependencies but can have complex spatial features. Spatial attention assigns varying weights to data from different sensors or wearing positions in the spatial dimension according to their importance for activity recognition. Thus spatial attention-based models can learn more robust spatial feature representations, as shown in Figure 4.
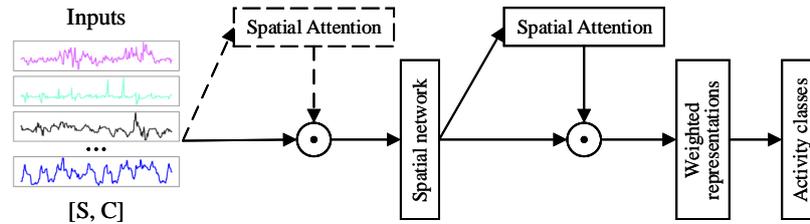

Figure 4: Spatial attention-based model

Wang et al. [33] proposed an attention-based CNN architecture that separately adds attention submodules after the third, fourth and fifth CNN layers, respectively. The attention submodule weighs local features of the total spatial features learned by a CNN to enhance the noticeable parts and weaken the less significant parts. The CNN model with attention submodules shows a more efficient backpropagation and improves recognition accuracy. Sarkar et al. [34] applied continuous wavelet transform to convert the time series from sensors with the size of $c \times t$ to 2D frequency-time domain scalograms with the size of $c \times t \times t$, here c is the number of timestamps and t denotes the number of sensor channels. Then they conducted the spatial attention to c to improve the features maps in CNN, thus enhancing CNN's ability to learn deeper features.

3.3. Temporal & spatial attention

Temporal and spatial attention mechanisms have demonstrated their particular strengths in WSHAR tasks. Based on their advantages, the temporal & spatial attention can extract more robust spatial-temporal features from the raw sensor data or feature representations learned by the model, as shown in Figure 5.
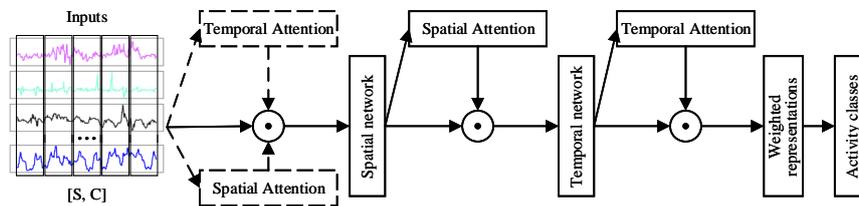

Figure 5: Temporal & spatial attention-based model

Ma et al. [35] proposed the AttnSense model that adds different attention mechanisms after CNN and GRU, respectively. The attention subnet after CNN takes self-attention to give varying weights for different sensor modalities from the spatial dimension, aiming to prioritize the important modalities. The attention subnet after GRU uses self-attention to increase the feature extraction of time steps with important contributions, while weakening the impact of unimportant time steps. The comparison results on three publicly available HAR datasets show that the model with two attention subnets achieves higher accuracy than using either one or neither. Gao et al. [36] introduced channel attention and temporal attention in the DanHAR model for feature re-extraction on the representations generated by CNN from the spatial and temporal dimensions, respectively. The channel attention they proposed for HAR time series signals is essentially a kind of spatial attention that uses max-pooling to combine the important representations generated through multiple filters in the convolutional layer. In the temporal attention, they applied global average-pooling and max-pooling to aggregate important global contextual information. Similarly, Zheng [39] also used spatial & temporal attention in the LGSTNet model, with the difference that Zheng applies temporal attention to the original input windows and spatial attention to the representations generated by 2D CNN. The probabilities generated by temporal and spatial attention can enhance the contributions of the important segments of the local spatial-temporal features. The ablation experiment results indicate that the attention mechanisms can improve the recognition performance of the LGSTNet model, especially in recognizing some similar activities.

4. Future directions

The studies mentioned above have demonstrated the potential of the self-attention mechanism in human activity recognition [29, 22, 35]. The Q, K and V in WSHAR tasks are almost from the same data source, which is effective in capturing sequential features and global information. Recently, some studies have improved self-attention to propose a mechanism called cross-attention, where Q and K come from the same data source and V comes from another data source. Thus, cross-attention enables to learn the inter-dependencies of different data sources. Lin et al. [49] proposed a cross-attention mechanism that alternates attention to information within image slices to acquire local information and captures attention information between local image slices to obtain global information. Their model improves the recognition rate in vision tasks and reduces the computational effort of the self-attention mechanism in the standard Transformer. Bhatti et al. [50] established attention-based cross-modality for connecting different locations of wearable information for emotion recognition.

For multi-position wearable HAR, how to fuse the data provided by multiple heterogeneous sensors and learn the data relationships within and between sensors is also one of the main challenges. The sensor data from different wearing positions are either similar or different. Therefore, introducing a cross-attention mechanism to establish the interaction of sensor information from different wearing positions can be further explored.

5. Conclusion

This paper summarizes attention mechanisms used in wearable sensor-based human activity recognition. Firstly, based on the principles of computing attention, we divide the attention mechanisms commonly used in WSHAR into two categories, i.e., weight-based attention and self-attention, and theoretically demonstrate their principles. Secondly, targeting the objects on which the attention mechanism acts, we survey and discuss the applications of attention mechanisms in WSHAR in terms of the attention methods: temporal attention, spatial attention and spatial & temporal attention. Finally, we point out that introducing the cross-attention mechanisms to WSHAR can benefit learning global and cross correlations in related works. Future work should consider running the codes of works mentioned above and further mine the attention principles in specific WSHAR tasks.

Acknowledgments

References

[1] H. Liu, Z. Dai, D. So, Q. V. Le, Pay attention to mlps, Advances in Neural Information Processing Systems 34 (2021) 9204–9215.

[2] M.-H. Guo, Z.-N. Liu, T.-J. Mu, S.-M. Hu, Beyond self-attention: External attention using two linear layers for visual tasks, IEEE Transactions on Pattern Analysis and Machine Intelligence 45 (2022) 5436–5447. doi:10.1109/TPAMI.2022.3211006.

[3] Y. Wang, C. Jing, S. Xu, T. Guo, Attention based spatiotemporal graph attention networks for traffic flow forecasting, Information Sciences 607 (2022) 869–883. doi:10.1016/j.ins.2022.05.127.

[4] S. Serrano, N. A. Smith, Is attention interpretable?, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019, pp. 2931–2951. doi:10.18653/v1/P19-1282.

[5] D. Bahdanau, K. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translate, arXiv preprint arXiv:1409.0473 (2014). doi:10.48550/arXiv.1409.0473.

[6] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Advances in neural information processing systems 30 (2017). URL: https://dl.acm.org/doi/10.5555/3295222.3295349.

[7] Y. Peng, X. He, J. Zhao, Object-part attention model for fine-grained image classification, IEEE Transactions on Image Processing 27 (2017) 1487–1500. doi:10.1109/TIP.2017.2774041.

[8] A. M. Obeso, J. Benois-Pineau, M. S. G. Vázquez, A. Á. R. Acosta, Visual vs internal attention

mechanisms in deep neural networks for image classification and object detection, Pattern Recognition 123 (2022) 108411. doi:`10.1016/j.patcog.2021.108411`.

[9] S. Chang, Y. Li, S. Shen, J. Feng, Z. Zhou, Contrastive attention for video anomaly detection, IEEE Transactions on Multimedia 24 (2021) 4067–4076. doi:`10.1109/TMM.2021.3112814`.

[10] Q. Li, R. Yang, F. Xiao, B. Bhanu, F. Zhang, Attention-based anomaly detection in multi-view surveillance videos, Knowledge-Based Systems 252 (2022) 109348. doi:`10.1016/j.knosys.2022.109348`.

[11] M. Wu, C. Zhang, J. Liu, L. Zhou, X. Li, Towards accurate high resolution satellite image semantic segmentation, IEEE Access 7 (2019) 55609–55619. doi:`10.1109/ACCESS.2019.2913442`.

[12] R. Li, S. Zheng, C. Duan, J. Su, C. Zhang, Multistage attention resu-net for semantic segmentation of fine-resolution remote sensing images, IEEE Geoscience and Remote Sensing Letters 19 (2021) 1–5. doi:`10.1109/LGRS.2021.3063381`.

[13] Z. Niu, G. Zhong, H. Yu, A review on the attention mechanism of deep learning, Neurocomputing 452 (2021) 48–62. doi:`10.1016/j.neucom.2021.03.091`.

[14] F. Serpush, M. B. Menhaj, B. Masoumi, B. Karasfi, Wearable sensor-based human activity recognition in the smart healthcare system, Computational intelligence and neuroscience 2022 (2022). doi:`10.1155/2022/1391906`.

[15] Z. Wang, J. Wang, H. Zhao, S. Qiu, J. Li, F. Gao, X. Shi, Using wearable sensors to capture posture of the human lumbar spine in competitive swimming, IEEE Transactions on Human-Machine Systems 49 (2019) 194–205. doi:`10.1109/THMS.2019.2892318`.

[16] A. Anagnostis, L. Benos, D. Tsaopoulos, A. Tagarakis, N. Tsolakis, D. Bochtis, Human activity recognition through recurrent neural networks for human–robot interaction in agriculture, Applied Sciences 11 (2021) 2188. doi:`10.3390/app11052188`.

[17] Y. Wang, S. Cang, H. Yu, A survey on wearable sensor modality centred human activity recognition in health care, Expert Systems with Applications 137 (2019) 167–190. doi:`10.1016/j.eswa.2019.04.057`.

[18] S. Chaudhari, V. Mithal, G. Polatkan, R. Ramanath, An attentive survey of attention models, ACM Transactions on Intelligent Systems and Technology (TIST) 12 (2021) 1–32. doi:`10.1145/3465055`.

[19] K. Chen, L. Yao, D. Zhang, X. Wang, X. Chang, F. Nie, A semisupervised recurrent convolutional attention model for human activity recognition, IEEE transactions on neural networks and learning systems 31 (2019) 1747–1756. doi:`10.1109/TNNLS.2019.2927224`.

[20] S. Mahmud, M. Tanjid Hasan Tonmoy, K. Kumar Bhaumik, A. Mahbubur Rahman, M. Ashraful Amin, M. Shoyaib, M. Asif Hossain Khan, A. Ahsan Ali, Human activity recognition from wearable sensor data using self-attention, in: Twenty-fourth European Conference on Artificial Intelligence (ECAI), IOS Press, 2020, pp. 1332–1339. doi:`10.3233/FAIA200236`.

[21] B. Sun, M. Liu, R. Zheng, S. Zhang, Attention-based lstm network for wearable human activity recognition, in: 2019 Chinese Control Conference (CCC), 2019, pp. 8677–8682. doi:`10.23919/ChiCC.2019.8865360`.

[22] S. P. Singh, M. K. Sharma, A. Lay-Ekuakille, D. Gangwar, S. Gupta, Deep convlstm with self-attention for human activity decoding using wearable sensors, IEEE Sensors Journal 21 (2020) 8575–8582. doi:`10.1109/JSEN.2020.3045135`.

[23] D. Hu, An introductory survey on attention mechanisms in nlp problems, in: Intelli-

gent Systems and Applications: Proceedings of the 2019 Intelligent Systems Conference (IntelliSys) Volume 2, Springer, 2020, pp. 432–448. doi:10.1007/978-3-030-29513-4_31.

[24] N. Zhang, J. Kim, A survey on attention mechanism in nlp, in: 2023 International Conference on Electronics, Information, and Communication (ICEIC), IEEE, 2023, pp. 1–4. doi:10.1109/ICEIC57457.2023.10049971.

[25] M.-H. Guo, T.-X. Xu, J.-J. Liu, Z.-N. Liu, P.-T. Jiang, T.-J. Mu, S.-H. Zhang, R. R. Martin, M.-M. Cheng, S.-M. Hu, Attention mechanisms in computer vision: A survey, Computational Visual Media 8 (2022) 331–368. doi:10.1007/s41095-022-0271-y.

[26] X. Yang, An overview of the attention mechanisms in computer vision, in: Journal of Physics: Conference Series, volume 1693, IOP Publishing, 2020, p. 012173. doi:10.1088/1742-6596/1693/1/012173.

[27] M. N. Haque, M. T. H. Tonmoy, S. Mahmud, A. A. Ali, M. A. H. Khan, M. Shoyaib, Gru-based attention mechanism for human activity recognition, in: 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), IEEE, 2019, pp. 1–6. doi:10.1109/ICASERT.2019.8934659.

[28] T. R. Mim, M. Amatullah, S. Afreen, M. A. Yousuf, S. Uddin, S. A. Alyami, K. F. Hasan, M. A. Moni, Gru-inc: An inception-attention based approach using gru for human activity recognition, Expert Systems with Applications 216 (2023) 119419. doi:10.1016/j.eswa.2022.119419.

[29] C. Betancourt, W.-H. Chen, C.-W. Kuan, Self-attention networks for human activity recognition using wearable devices, in: 2020 IEEE international conference on systems, man, and cybernetics (SMC), IEEE, 2020, pp. 1194–1199. doi:10.1109/SMC42975.2020.9283381.

[30] L. Liu, J. He, K. Ren, J. Lungu, Y. Hou, R. Dong, An information gain-based model and an attention-based rnn for wearable human activity recognition, Entropy 23 (2021) 1635. doi:10.3390/e23121635.

[31] X. Yin, Z. Liu, D. Liu, X. Ren, A novel cnn-based bi-lstm parallel model with attention mechanism for human activity recognition with noisy data, Scientific Reports 12 (2022) 1–11. doi:10.1038/s41598-022-11880-8.

[32] M. A. Khatun, M. A. Yousuf, S. Ahmed, M. Z. Uddin, S. A. Alyami, S. Al-Ashhab, H. F. Akhdar, A. Khan, A. Azad, M. A. Moni, Deep cnn-lstm with self-attention model for human activity recognition using wearable sensor, IEEE Journal of Translational Engineering in Health and Medicine 10 (2022) 1–16. doi:10.1109/JTEHM.2022.3177710.

[33] K. Wang, J. He, L. Zhang, Attention-based convolutional neural network for weakly labeled human activities' recognition with wearable sensors, IEEE Sensors Journal 19 (2019) 7598–7604. doi:10.1109/JSEN.2019.2917225.

[34] A. Sarkar, S. S. Hossain, R. Sarkar, Human activity recognition from sensor data using spatial attention-aided cnn with genetic algorithm, Neural Computing and Applications 35 (2023) 5165–5191. doi:10.1007/s00521-022-07911-0.

[35] H. Ma, W. Li, X. Zhang, S. Gao, S. Lu, Attnsense: multi-level attention mechanism for multimodal human activity recognition, in: Proceedings of the 28th International Joint Conference on Artificial Intelligence, 2019, pp. 3109–3115. doi:10.5555/3367471.3367473.

[36] W. Gao, L. Zhang, Q. Teng, J. He, H. Wu, Danhar: Dual attention network for multimodal

human activity recognition using wearable sensors, Applied Soft Computing 111 (2021) 107728. doi:10.1016/j.asoc.2021.107728.

[37] Y. Tang, L. Zhang, Q. Teng, F. Min, A. Song, Triple cross-domain attention on human activity recognition using wearable sensors, IEEE Transactions on Emerging Topics in Computational Intelligence 6 (2022) 1167–1176. doi:10.1109/TETCI.2021.3136642.

[38] C. Gao, Y. Chen, X. Jiang, L. Hu, Z. Zhao, Y. Zhang, Bi-stan: bilinear spatial-temporal attention network for wearable human activity recognition, International Journal of Machine Learning and Cybernetics (2023) 1–17. doi:10.1007/s13042-023-01781-1.

[39] G. Zheng, A novel attention-based convolution neural network for human activity recognition, IEEE Sensors Journal 21 (2021) 27015–27025. doi:10.1109/JSEN.2021.3122258.

[40] M. Zeng, H. Gao, T. Yu, O. J. Mengshoel, H. Langseth, I. Lane, X. Liu, Understanding and improving recurrent networks for human activity recognition by continuous attention, in: Proceedings of the 2018 ACM international symposium on wearable computers, 2018, pp. 56–63. doi:10.1145/3267242.3267286.

[41] Y. Tang, L. Zhang, F. Min, J. He, Multiscale deep feature learning for human activity recognition using wearable sensors, IEEE Transactions on Industrial Electronics 70 (2022) 2106–2116. doi:10.1109/TIE.2022.3161812.

[42] Y. Zhao, R. Yang, G. Chevalier, X. Xu, Z. Zhang, Deep residual bidir-lstm for human activity recognition using wearable sensors, Mathematical Problems in Engineering 2018 (2018) 1–13. doi:10.1155/2018/7316954.

[43] L. Wang, R. Liu, Human activity recognition based on wearable sensor using hierarchical deep lstm networks, Circuits, Systems, and Signal Processing 39 (2020) 837–856. doi:10.1007/s00034-019-01116-y.

[44] S. Zhao, H. Wei, K. Zhang, Deep bidirectional gru network for human activity recognition using wearable inertial sensors, in: 2022 3rd International Conference on Electronic Communication and Artificial Intelligence (IWECAI), IEEE, 2022, pp. 238–242. doi:10.1109/IWECAI55315.2022.00054.

[45] R. A. Hamad, M. Kimura, L. Yang, W. L. Woo, B. Wei, Dilated causal convolution with multi-head self attention for sensor human activity recognition, Neural Computing and Applications 33 (2021) 13705–13722. doi:10.1007/s00521-021-06007-5.

[46] J. Pan, Z. Hu, S. Yin, M. Li, Gru with dual attentions for sensor-based human activity recognition, Electronics 11 (2022) 1797. doi:10.3390/electronics11111797.

[47] K. Chen, D. Zhang, L. Yao, B. Guo, Z. Yu, Y. Liu, Deep learning for sensor-based human activity recognition: Overview, challenges, and opportunities, ACM Computing Surveys (CSUR) 54 (2021) 1–40. doi:10.1145/3447744.

[48] T. Sztyler, H. Stuckenschmidt, W. Petrich, Position-aware activity recognition with wearable devices, Pervasive and mobile computing 38 (2017) 281–295. doi:10.1016/j.pmcj.2017.01.008.

[49] H. Lin, X. Cheng, X. Wu, D. Shen, Cat: Cross attention in vision transformer, in: 2022 IEEE International Conference on Multimedia and Expo (ICME), IEEE, 2022, pp. 1–6. doi:10.1109/ICME52920.2022.9859720.

[50] A. Bhatti, B. Behinaein, P. Hungler, A. Etemad, Attx: Attentive cross-connections for fusion of wearable signals in emotion recognition, arXiv preprint arXiv:2206.04625 (2022). doi:10.48550/arXiv.2206.04625.