Final Literature Review

Hithesh Shanmugam

Citations:

1. Tasnim, Nusrat, and Joong-Hwan Baek. "Dynamic Edge Convolutional Neural Network for SkeletonBased Human Action Recognition." *Sensors* 23, no. 2 (2023): 778.

2. Zeng, Ailing, Lei Yang, Xuan Ju, Jiefeng Li, Jianyi Wang, and Qiang Xu. "Smoothnet: a plug-and-play network for refining human poses in videos." In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part V*, pp. 625-642. Cham: Springer Nature Switzerland, 2022.

3. Kim, Taehwan, Jeongho Park, Juwon Lee, and Jooyoung Park. "Predicting human motion signals using modern deep learning techniques and smartphone sensors." *Sensors* 21, no. 24 (2021): 8270.

4. Cui, Zhicheng, Wenlin Chen, and Yixin Chen. "Multi-scale convolutional neural networks for time series classification." *arXiv preprint arXiv:1603.06995* (2016).

5. Zhang, Ye, Yi Hou, Shilin Zhou, and Kewei Ouyang. "Encoding time series as multi-scale signed recurrence plots for classification using fully convolutional networks." *Sensors* 20, no. 14 (2020): 3818.

Applied Methods:

[1] The paper proposes a new deep learning model for human action recognition (HAR) that integrates criss-cross attention and edge convolution to extract discriminative features from the skeleton sequence. The attention mechanism is applied in both spatial and temporal directions to capture intra- and inter-frame relationships. Several edge convolutional layers are then conducted to explore the geometric relationships among neighboring joints in the human body. The proposed model is dynamically updated after each layer by recomputing the graph based on k-nearest joints for learning local and global information in action sequences.

The proposed method was evaluated on two publicly available benchmark skeleton datasets, UTD-MHAD and MSR-Action3D. The authors also investigated the proposed method with different configurations of network architectures to assure effectiveness and robustness.

[2] This paper proposes a novel method called SmoothNet, which addresses the problem of highly unbalanced jittery outputs from existing pose estimators when analysing human motion videos. SmoothNet is a dedicated temporal-only refinement network that attaches to existing pose estimators to mitigate jitters. It models the natural smoothness characteristics in body movements by learning the long-range temporal relations of every joint without considering the noisy correlations among joints.

[**3]** This paper proposes a new method for predicting human motion signals obtained from sensors attached to individuals. The method involves converting the motion signals into image formats using the recurrence plot method, which is then used as an input into a deep learning model. The deep learning model used is a Fourier neural operator, which combines neural networks and the Fourier transform.

[4] The proposed method in the paper is Multi-scale Convolutional Neural Network (MCNN), which is an end-to-end neural network model for time series classification (TSC). MCNN incorporates feature extraction and classification in a single framework and leverages a multi-branch layer and learnable convolutional layers to automatically extract features at different scales and frequencies.

The authors compared the performance of MCNN with various traditional approaches that involve ad-hoc feature extraction using dynamic time warping (DTW) or shapelet transformation. The empirical evaluation was conducted on many benchmark datasets.

[5] This paper proposes a novel approach for time series classification (TSC) by leveraging Multi-scale Signed Recurrence Plots (MS-RP) and Fully Convolutional Networks (FCN). The method first encodes time series as RP images with phase space dimension and time delay embedding, then constructs MS-RP images with designed sign masks to remove the tendency confusion. Finally, FCN is trained with MS-RP images to perform classification.

<u>Overview:</u>

[1] Human action recognition (HAR) is a critical research area in computer vision for providing efficient and accessible remote access in various domains such as rehabilitation, virtual games, and healthcare. The paper proposes a new deep learning model that integrates criss-cross attention and edge convolution to extract discriminative features from the skeleton sequence for action recognition.

The proposed method is evaluated on two publicly available benchmark skeleton datasets, UTD-MHAD and MSR-Action3D, and is compared with state-of-the-art methods. The authors also investigate the proposed method with different configurations of network architectures to ensure effectiveness and robustness.

[2] The proposed method aims to improve the temporal smoothness of existing pose estimators and enhance the estimation accuracy of challenging frames. SmoothNet achieves this by learning the natural smoothness characteristics in body movements, which allows it to effectively mitigate jitters in the output of existing pose estimators. Unlike existing solutions, SmoothNet is a temporal-only model, which makes it highly transferable across various types of estimators, modalities, and datasets.

[3] The study focuses on the use of wearable sensors and sensor signals that operate in the form of time series for healthcare applications. Most studies convert the sensor signals into an image format for analysis. The proposed method uses the recurrence plot method for converting motion signals into image formats and utilizes the Fourier neural operator for predicting subsequent motion signals.

[4] The paper focuses on the problem of time series classification (TSC), which has been around for decades within the community of data mining and machine learning. Traditional approaches to TSC involve ad-hoc feature extraction using dynamic time warping (DTW) or shapelet transformation. However, these methods limit accuracy performance because they separate the feature extraction and classification parts, and most methods fail to account for the fact that time series often have features at different time scales.

To address these limitations, the authors propose an end-to-end neural network model, MCNN, which incorporates feature extraction and classification in a single framework and accounts for features at different time scales. The authors also conduct empirical evaluation to compare the performance of MCNN with various traditional approaches on benchmark datasets.

[5] The paper addresses the problem of variability in distinctive region scale and sequence length and the tendency confusion problem in TSC using MS-RP images and FCN. The proposed approach outperforms the state-of-the-art methods in terms of classification accuracy and visualization evaluation on 45 benchmark datasets.

Strengths:

[1] The proposed method offers several strengths. First, it integrates criss-cross attention and edge convolution to extract discriminative features from the skeleton sequence for action recognition, providing a novel approach to HAR. Second, the proposed model is dynamically updated after each layer by recomputing the graph based on k-nearest joints for learning local and global information in action sequences, which enables the model to adapt to different action sequences.

Finally, the proposed method achieves state-of-the-art performance on two publicly available benchmark skeleton datasets, UTD-MHAD and MSR-Action3D, demonstrating its effectiveness and robustness.

[2] SmoothNet is a novel approach to the problem of jittery outputs from existing pose estimators when analysing human motion videos. By modelling the natural smoothness characteristics in body movements, it achieves significant improvements in temporal smoothness and estimation accuracy. Its temporal-only design also makes it highly transferable across various types of estimators, modalities, and datasets.

[3] The proposed method offers better performance compared to the CNN model. The use of wearable sensors and sensor signals provides an efficient way to monitor human activities and characterize human motions for healthcare applications. The Fourier neural operator is a novel approach for predicting subsequent motion signals.

[4] The strength of the proposed method, MCNN, is that it is an end-to-end neural network model that incorporates feature extraction and classification in a single framework. This approach overcomes the limitations of traditional methods that separate the feature extraction and classification parts, which limits accuracy performance. Additionally, MCNN automatically extracts features at different scales and frequencies, leading to superior feature representation.

The authors conducted comprehensive empirical evaluation on benchmark datasets and showed that MCNN outperforms other leading methods in terms of accuracy performance.

[5] The proposed approach tackles the important problem of variability in distinctive region scale and sequence length and the tendency confusion problem in TSC, improving the state-of-the-art in terms of classification accuracy and visualization evaluation. By leveraging MS-RP images and FCN, the method achieves impressive results, demonstrating the potential of this approach for TSC.

Weaknesses:

[1] One potential weakness of the proposed method is that it may be computationally expensive due to the use of attention mechanisms and edge convolutional layers. Additionally, the paper does not provide a detailed analysis of the interpretability of the proposed method, which may limit its practical applications in certain scenarios where interpretability is important.

[2] The paper does not provide a detailed comparison of the proposed method with other state-of-the-art solutions. While the results of the experiments demonstrate the efficacy of the proposed solution, a more in-depth analysis of its strengths and weaknesses compared to other methods would have been beneficial. Additionally, the paper does not discuss potential limitations or challenges associated with the implementation of SmoothNet in real-world applications.

[3] The study is limited to predicting motion signals obtained from sensors attached to individuals. The proposed method has not been tested on other types of sensor signals or for other healthcare applications. The study does not provide an analysis of the limitations or potential drawbacks of using the recurrence plot method for converting time series data to image formats.

[4] One potential weakness of the proposed method is that it may not be easily interpretable because of the black-box nature of neural networks. Additionally, the paper does not provide a detailed analysis of the computational complexity of the proposed method, which may limit its practical applications in certain scenarios where computational resources are limited.

[5] The proposed approach relies heavily on the design of sign masks to remove the tendency confusion, which may limit its generalization to new datasets. Additionally, the paper does not provide a thorough comparison with other state-of-the-art methods in terms of computational efficiency, which may be an important consideration in practical applications.