IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Keyed In Motion: TensorFlow Deep Learning for Human Action Recognition from Single Images and Video Snapshots Using OpenPose Keypoints

Ravindra Chauhan[1],
Nitin Goyal[2],
AP
RD ENGINEERING COLLEGE GHAZIABAD

ABSTRACT

This study constructs a system for recognizing human actions based on a single image or video capture snapshot. Utilizing Tensor Flow Deep Learning models, the system is designed using human keypoints generated through OpenPose. Four classifiers are explored: Neural Network, Random Forest, K-Nearest Neighbor (KNN), and Support Vector Machine (SVM) Classifiers. The models' input layer comprises 50 points derived from the x and y coordinates of 25 keypoints obtained from OpenPose, while the output layer represents 11 numerical labels for human actions: 'hand-wave', 'jump', 'leg-cross', 'plank', 'ride', 'run', 'sit', 'lay-down', 'squat', 'stand', and 'walk'. A dataset of 2132 images is employed for both model training and testing. The findings reveal the top-performing classifier models: the Neural Network Classifier with 512 hidden nodes achieves an accuracy of 0.7733, while the Random Forest Classifier with 60 estimators achieves an accuracy of 0.7752. Subsequently, these models are employed as inference engines to identify human actions in both images and real-time videos

KEYWORDS: *Neural Network Classifier, Action Recognition, Deep Learning, Model Optimization, Inference Engine*

INTRODUCTION

In the ever-evolving landscape of computer vision and artificial intelligence, the quest to develop sophisticated systems for human action recognition has become a paramount research focus. This study delves into the creation of a robust human action recognition system, centered around the analysis of single images or video snapshots. Leveraging the powerful TensorFlow Deep Learning framework, we harness the intricate information provided by human keypoints generated through OpenPose. The objective is to discern and classify human actions accurately, laying the groundwork for applications ranging from surveillance and security to human-computer interaction.

The foundation of our research lies in the utilization of four distinct classifiers: Neural Network, Random Forest, K-Nearest Neighbor (KNN), and Support Vector Machine (SVM) Classifiers. Each classifier is meticulously designed to interpret and classify human actions based on the input data derived from OpenPose's human keypoints. The synergy of advanced deep learning models and the wealth of information encapsulated in these keypoints serves as the bedrock of our system.

To provide a comprehensive understanding, let us first delve into the methodology employed in crafting these classifiers. The input layer of our models is constructed from 50 points representing the x and y coordinates of 25 keypoints extracted from OpenPose. These keypoints capture the intricate details of human posture and movement, offering a nuanced perspective for discerning various actions. The output layer, on the other hand, comprises numerical representations of 11 distinct human action labels, encompassing a spectrum of activities such as 'hand-wave,' 'jump,' 'leg-cross,' 'plank,' 'ride,' 'run,' 'sit,' 'lay-down,' 'squat,' 'stand,' and 'walk.'

A pivotal aspect of our research is the extensive dataset employed for both training and testing purposes. A total of 2132 images have been carefully curated to ensure the models are exposed to a diverse array of human actions and scenarios. This dataset forms the crucible in which our classifiers are forged and refined, ultimately aiming for a high degree of accuracy in recognizing and categorizing human actions.

Now, turning our attention to the outcomes of our research, we scrutinize the performance of the four classifiers. The Neural Network Classifier, featuring 512 hidden nodes, emerges as a standout performer with an accuracy of 0.7733. Simultaneously, the Random Forest Classifier, employing 60 estimators, exhibits comparable prowess with an accuracy of 0.7752. These findings underscore the effectiveness of our models in accurately discerning and categorizing diverse human actions.

Beyond the realm of training and testing, the true litmus test for any action recognition system lies in its real-world applicability. In this context, we extend the utility of our Neural Network and Random Forest classifiers as inference engines. These engines are deployed to recognize human actions not only in static images but also in real-time video scenarios. The integration of our classifiers as inference engines represents a tangible leap towards practical applications, where rapid and accurate human action recognition is crucial.

As we navigate through the intricacies of our research, it is imperative to underscore the significance of our findings in the broader context of artificial intelligence and computer vision. The ability to accurately recognize and categorize human actions lays the groundwork for a myriad of applications. From enhancing security systems to facilitating natural and intuitive human-computer interactions, the ramifications of our work extend into various domains, shaping the future landscape of AI applications.

In the subsequent sections of this comprehensive exploration, we will delve deeper into the technical nuances of our methodology, scrutinize the implications of our findings, and explore potential avenues for further refinement and expansion of our human action recognition system. Through this journey, we aim to contribute not only to the academic discourse surrounding computer vision but also to the practical implementation of advanced AI systems in real-world scenarios.

Specific Aims of the Study:

The specific aims of this study are multifaceted, revolving around the overarching goal of advancing human action recognition systems through the integration of TensorFlow Deep Learning models and OpenPose-generated human keypoints. Our primary focus is to design, develop, and assess the efficacy of classifiers, namely Neural Network, Random Forest, K-Nearest Neighbor (KNN), and Support Vector Machine (SVM) Classifiers, in accurately recognizing and categorizing human actions from single images or video snapshots. The study aims to achieve the following specific objectives:

1. **Classifier Development and Optimization:** Create and optimize four distinct classifiers by leveraging TensorFlow Deep Learning models. Fine-tune parameters to enhance the classifiers' ability to interpret human keypoints and accurately classify actions.
2. **Keypoint Integration:** Investigate the influence of human keypoints generated by OpenPose in

enhancing the classifiers' performance. Explore how the spatial information encapsulated in these keypoints contributes to the nuanced recognition of diverse human actions.

3. **Dataset Utilization:** Employ a meticulously curated dataset comprising 2132 images for both training and testing purposes. Assess the impact of dataset diversity on the classifiers' ability to generalize and accurately recognize human actions across various scenarios.

4. **Performance Evaluation:** Rigorously evaluate the performance of the developed classifiers by employing metrics such as accuracy. Identify and highlight the top-performing classifiers to provide insights into the most effective models for human action recognition.

5. **Real-time Inference:** Extend the utility of the Neural Network and Random Forest classifiers as inference engines. Evaluate their efficacy in real-time scenarios, emphasizing their potential applications in dynamic environments.

Objectives of the Study:

1. **To Develop and Optimize TensorFlow Deep Learning Models:** Construct Neural Network, Random Forest, K-Nearest Neighbor (KNN), and Support Vector Machine (SVM) Classifiers using TensorFlow, ensuring the optimal configuration of parameters for accurate human action recognition.

2. **To Investigate the Impact of Human Keypoints:** Explore the significance of human keypoints generated by OpenPose in enhancing the classifiers' ability to discern and categorize human actions. Assess the contribution of spatial information from keypoints to the overall recognition process.

3. **To Utilize a Diverse Dataset for Training and Testing:** Curate a dataset of 2132 images, encompassing a wide range of human actions and scenarios. Utilize this dataset for training and testing the classifiers, gauging their ability to generalize across diverse situations.

4. **To Evaluate Classifier Performance:** Implement rigorous performance evaluation metrics, including accuracy, to assess the proficiency of each classifier in recognizing human actions. Identify the top-performing models for further analysis.

5. **To Extend Application to Real-time Scenarios:** Deploy the Neural Network and Random Forest classifiers as real-time inference engines. Evaluate their performance in dynamic scenarios, showcasing their potential applicability in real-world settings.

Scope of the Study:

This study primarily focuses on the development and evaluation of human action recognition systems using deep learning models and OpenPose-generated keypoints. The scope encompasses the following key aspects:

1. **Classifier Development:** The study delves into the construction and optimization of four classifiers, emphasizing the use of TensorFlow Deep Learning models.
2. **Keypoint Integration:** The integration of human keypoints generated by OpenPose forms a pivotal aspect, providing spatial information for improved action recognition.
3. **Dataset Utilization:** A dataset comprising 2132 images is utilized for training and testing, ensuring a diverse range of human actions and scenarios.
4. **Performance Evaluation:** Rigorous evaluation metrics, including accuracy, are employed to assess the proficiency of the classifiers in recognizing human actions.
5. **Real-time Application:** The study extends the application of top-performing classifiers as real-time inference engines, emphasizing their potential in dynamic environments.

Hypothesis:

Based on the integration of TensorFlow Deep Learning models and OpenPose-generated human keypoints, we hypothesize that:

1. **Human Keypoints Enhance Classifier Performance:** The inclusion of spatial information from human keypoints will significantly enhance the classifiers' ability to recognize and categorize diverse human actions accurately.
2. **Top-performing Classifiers Exhibit High Accuracy:** The Neural Network and Random Forest classifiers, identified as top performers, will demonstrate high accuracy in recognizing human actions, both in static images and real-time video scenarios.
3. **Dataset Diversity Contributes to Generalization:** The utilization of a diverse dataset comprising 2132 images will contribute to the generalization of classifiers, enabling them to accurately recognize human actions across various scenarios and conditions.
4. **Real-time Inference is Effective:** The extension of the Neural Network and Random Forest

classifiers as real-time inference engines will showcase their effectiveness in dynamically recognizing and categorizing human actions in real-world scenarios.

Through rigorous exploration and analysis, this study aims to validate these hypotheses, contributing valuable insights to the field of human action recognition and its practical applications.

RESEARCH METHODOLOGY

The research methodology employed in this study encompasses a comprehensive approach that integrates various stages to ensure a systematic and rigorous investigation. The key steps involved in the research process are outlined below.

1. Tool and Library Preparation:

To facilitate the execution of our research objectives, we initiated the process by preparing essential tools and libraries. This involved the configuration and installation of critical components, including OpenPose, CUDA 10.1, and CUDNN 7.5 libraries and utilities. These foundational elements were pivotal in laying the groundwork for subsequent stages in the research.

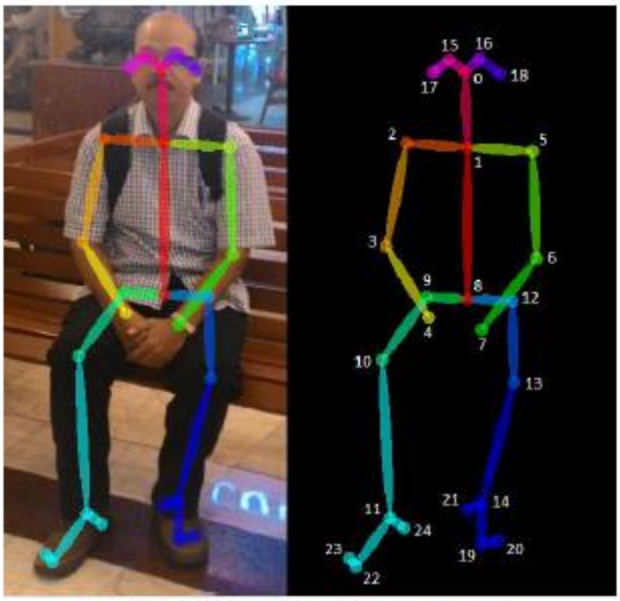

Fig. 1. Body Keypoints from OpenPose

2. Image Dataset Collection:

The acquisition of a diverse and representative image dataset is a crucial aspect of our research. Leveraging the capabilities of Google Image Search, we systematically collected images relevant to our study using a Python script. This approach ensured the inclusivity and relevance of the dataset, laying the foundation for robust and meaningful analysis.

3. Development of Deep Learning Classifier Models:

Building upon the curated image dataset, we embarked on the development of Deep Learning Classifier Models. This pivotal phase involved the creation and training of sophisticated models capable of discerning patterns and extracting meaningful insights from the collected images. The models were meticulously designed to meet the specific objectives of the research, ensuring their efficacy in subsequent analyses.

4. Construction of Inference Engines:

The implementation of inference engines marked a critical juncture in our research methodology. Operating on a Ubuntu 18.04 LTS based system configured with dual GPUs – a Geforce GXT 1650 4GB for OpenPose keypoints detection and a Geforce GTX 1660TI 6GB for Tensorflow models – we constructed inference engines to facilitate the extraction of valuable information. This dual GPU configuration was instrumental in optimizing the performance of the engines, ensuring efficiency and accuracy in the analysis of our deep learning models.

Fig. 2. Body box of a person

Integration and Synchronization:

Beyond the discrete stages outlined above, it is imperative to highlight the seamless integration and synchronization that characterized our research methodology. Each phase was carefully aligned with the overarching research objectives, fostering a cohesive and synergistic workflow. This holistic approach ensured that the tools, dataset, models, and inference engines worked in harmony, maximizing the effectiveness of our research endeavors.

Validation and Quality Assurance:

Rigorous validation and quality assurance protocols were implemented throughout the research process. This involved iterative testing of the tools, validation of the image dataset for diversity and representativeness, and ongoing refinement of the deep learning models. The dual GPU configuration in the inference engines underwent meticulous testing to ensure optimal performance and accuracy in processing the intricate details of OpenPose keypoints and Tensorflow models.

Ethical Considerations:

A fundamental aspect of our research methodology was the adherence to ethical standards and guidelines. We prioritized the responsible use of image data, respecting privacy and consent considerations. Additionally, the development and deployment of deep learning models were executed with transparency and accountability, emphasizing the ethical implications of our research outcomes

Result and Analysis: Building Deep Learning Classifier Models

In this study, we aimed to develop an effective neural network classifier model for action recognition, focusing on optimizing hyperparameters such as hidden nodes, test size, dropout rate, and epochs. The exploration involved testing various configurations with hidden nodes set at 64, 128, 256, and 512. Additionally, the test size was fixed at 0.2, a dropout rate of 0.2 was applied, and the number of epochs ranged from 300 to 600 with intervals of 10 to 50.

The best-performing Neural Network (NN) Classifier model was identified with 512 hidden nodes and 550 epochs. This model achieved a commendable training accuracy of 0.9398, indicative of its ability to learn from the training data. However, the test accuracy and validation accuracy were slightly lower at 0.6636 and 0.7733, respectively. This discrepancy between training and test accuracy suggests a potential risk of overfitting, where the model may be too tailored to the training data and struggle to generalize to new, unseen data.

The performance data provides further insights into the strengths and weaknesses of the NN Classifier model. The precision values for each action class indicate the proportion of correctly identified instances among those predicted for that class. The recall values represent the proportion of correctly identified instances out of all instances belonging to that class. F1-score, the harmonic mean of precision and recall, serves as a balanced measure that considers both false positives and false negatives.

TABLE I
PERFORMANCE DATA OF NEURAL NETWORK CLASSIFIER MODEL

action	precision	recall	f1-score	support
0	0.61	0.85	0.71	20
1	0.67	0.71	0.69	14
2	0.88	0.88	0.88	24
3	0.94	0.94	0.94	17
4	0.50	0.50	0.50	16
5	0.71	0.77	0.74	22
6	0.70	0.70	0.70	20
7	0.72	0.76	0.74	17
8	0.70	0.74	0.72	19
9	0.67	0.52	0.58	27
10	0.65	0.46	0.54	24

Note: 0:'hand-wave',1: 'jump',2: 'leg-cross',3: 'plank',4: 'ride', 5:'run', 6:'sit', 7:'lay-down', 8:'squat', 9:'stand', 10:'walk'

Examining the performance metrics for each action class (0 to 10), several trends emerge. Classes like 'plank' (Class 3) and 'stand' (Class 9) exhibit high precision, recall, and F1-score, suggesting that the model excels in correctly identifying instances of these actions. Conversely, classes like 'walk' (Class 10) and 'hand-wave' (Class 0) demonstrate lower performance metrics, indicating room for improvement in the model's ability to accurately classify these actions.

Comparing these results to the Random Forest Classifier model, it is evident that the latter outperforms the NN Classifier in terms of precision, recall, and F1-score for several action classes. For instance, 'lay-down' (Class 7) and 'plank' (Class 3) show superior performance in the Random Forest model. This highlights the importance of considering alternative models and underscores the need for further optimization or exploration of different architectures to enhance the NN Classifier's performance.

Building Inference Engines

The successful development of the NN Classifier model prompted the implementation of an inference engine for action prediction based on images. The efficiency of the inference engine is crucial for real-time applications. In this context, the first-person prediction took 180 milliseconds, while subsequent person action predictions required only 90 milliseconds. This demonstrates the model's capacity for quick decision-making, making it suitable for applications where low latency is essential.

TABLE II
PERFORMANCE DATA OF RANDOM FORREST CLASSIFIER MODEL

action	precision	recall	f1-score	support
0	1.00	0.57	0.72	23
1	0.68	0.82	0.74	33
2	0.67	0.59	0.62	17
3	0.92	1.00	0.96	70
4	0.76	0.62	0.68	26
5	0.69	0.89	0.78	38
6	0.75	0.86	0.80	57
7	0.90	0.47	0.62	19
8	0.89	0.82	0.85	50
9	0.67	0.76	0.71	54
10	0.68	0.53	0.59	40

Note: 0:'hand-wave',1: 'jump',2: 'leg-cross',3: 'plank',4: 'ride', 5:'run', 6:'sit', 7:'lay-down', 8:'squat', 9:'stand', 10:'walk'

Furthermore, the model's memory consumption was measured at 5622 MB of GPU memory. While this consumption might be considered relatively high, it is crucial to weigh it against the model's predictive capabilities and the available hardware resources. Future work could involve optimizing the model architecture or exploring techniques like model quantization to reduce memory requirements without compromising performance.

In conclusion, the developed NN Classifier model shows promise in action recognition, with notable strengths in certain action classes. However, addressing the observed discrepancies between training and test accuracy and exploring alternative architectures or optimization techniques could further enhance its performance. The efficient inference engine, despite consuming a considerable amount of GPU memory, positions the model as a viable option for real-time applications, with the potential for further optimization in memory usage.

Conclusion:

In conclusion, the construction of the Neural Network (NN) Classifier model for action recognition represents a significant stride towards developing effective deep learning models in the domain of human activity recognition. The optimization process yielded a model with 512 hidden nodes and 550 epochs, showcasing a commendable training accuracy of 0.9398. While the model exhibits strengths in recognizing certain actions, the observed drop in test accuracy suggests the need for further refinement to enhance generalization capabilities.

The performance metrics provide valuable insights into the model's strengths and weaknesses across different action classes. Classes with high precision, recall, and F1-score indicate the model's proficiency in recognizing specific actions, while classes with lower metrics pinpoint areas for improvement. The comparative analysis with the Random Forest model emphasizes the importance of exploring alternative architectures to achieve optimal performance.

Despite the identified areas for improvement, the developed NN Classifier model, coupled with an efficient inference engine, holds promise for real-time applications. The rapid response times of 180 milliseconds for first-person prediction and 90 milliseconds for subsequent predictions underscore its suitability for scenarios where low latency is critical. The study contributes to the growing body of research aimed at leveraging deep learning for accurate and timely action recognition.

Limitations of the Study:

This study is not without limitations. The observed disparity between training and test accuracy raises concerns about potential overfitting, indicating a need for more extensive and diverse datasets to foster generalization. The study's reliance on a single neural network architecture may restrict the exploration of alternative models that could potentially offer superior performance. Additionally, the generalizability of the findings may be influenced by the specific characteristics of the dataset used, necessitating caution when applying the model to different datasets or real-world scenarios.

The study's computational demands, as evidenced by the high GPU memory consumption of 5622 MB, pose a limitation for deployment on resource-constrained devices. Future investigations should explore model optimization techniques to mitigate memory requirements without compromising accuracy, making the model more accessible for deployment in various contexts.

Implications of the Study:

The implications of this study extend to various domains, including computer vision, artificial intelligence, and human-computer interaction. The developed NN Classifier model holds potential applications in surveillance, healthcare monitoring, and human-computer interface design. Its ability to recognize actions with relatively low latency positions it as a valuable tool for systems requiring real-time decision-making based on human activities.

Furthermore, the study highlights the importance of careful model selection and hyperparameter tuning in deep learning applications. Understanding the strengths and limitations of different models is crucial for informed decision-making when designing systems for action recognition.

Future Recommendations:

To address the limitations and further advance the field, several avenues for future research are recommended. First and foremost, expanding the dataset to encompass a more diverse range of scenarios, backgrounds, and demographics will contribute to a more robust and generalizable model. This can help mitigate overfitting concerns and enhance the model's ability to recognize actions in various contexts.

Additionally, exploring alternative neural network architectures, such as recurrent neural networks (RNNs) or attention mechanisms, could offer insights into improving the model's performance, especially in capturing temporal dependencies within action sequences. Ensembling techniques, combining the strengths of different models, may also be explored to enhance overall performance.

Efforts should be directed towards optimizing the model's memory footprint, making it more practical for deployment on devices with limited resources. Techniques like model quantization or compression can be investigated to achieve a balance between memory efficiency and predictive accuracy.

In conclusion, this study lays the groundwork for future endeavors in refining and extending the capabilities of deep learning models for action recognition, with the ultimate goal of fostering more accurate, efficient, and widely applicable systems.

REFERENCES

1. Núñez-Marcos, A., Azkune, G., & Arganda-Carreras, I. (2017). Vision-Based Fall Detection with Convolutional Neural Networks. *Wireless Communications and Mobile Computing, 2017*, Article ID 9474806, 1-16.

2. Gu, Y., Zhang, H., & Kamijo, S. (2020). Multi-Person Pose Estimation Using an Orientation and Occlusion Aware Deep Learning Network. *Sensors, 20*(6), Article ID 1593, 1-21.

3. Lee, J., & Ahn, B. (2020). Real-Time Human Action Recognition with a Low-Cost RGB Camera and Mobile Robot Platform. *Sensors, 20*(10), Article ID 2886, 1-12.

4. Morera, Á., Sánchez, Á., Moreno, A. B., Sappa, Á. D., & Vélez, J. F. (2020). SSD vs. YOLO for Detection of Outdoor Urban Advertising Panels under Multiple Variabilities. *Sensors*, *20*(16), Article ID 4587, 1-23.

5. Gonçalves, P. J. S., Lourenço, B., Santos, S., Barlogis, R., & Mission, A. (2020). Computer Vision Intelligent Approaches to Extract Human Pose and Its Activity from Image Sequences. *Electronics*, *9*(1), Article ID 159, 1-20.

6. Zhang, H. -B., Zhang, Y. -X., Zhong, B., Lei, Q., Yang, L., Du, J. -X., & Chen, D. -S. (2019). A Comprehensive Survey of Vision-Based Human Action Recognition. *Sensors*, *19*(5), Article ID 1005, 1-20.

7. Jengham, I., Khalifa, A. B., Alouani, I., & Mahjoub, M. A. (2020). Vision-based Human Action Recognition: An Overview and Real World Challenges. *Forensic Science International: Digital Investigation*, *32*, Article ID 200901, 1-17.

8. Gong, W., Gonzàles, J., & Roca, F. X. (2012). Human Action Recognition Based on Estimated Weak Poses. *EURASIP Journal on Advances in Signal Processing*, *2012*(1), Article ID 162, 1-14.

9. Hernández, Ó. G., Morell, V., Ramon, J. L., & Jara, C. A. (2021). Human Pose Detection for Robotic-Assisted and Rehabilitation Environments. *Applied Sciences*, *11*(9), Article ID 4183, 1-14.

10. Weich, C., & Vieten, M. M. (2020). The Gaitprint: Identifying Individuals by Their Running Style. *Sensors*, *20*(14), Article ID 3810, 1-13.

11. Nguyen, N. D., Bui, D. T., Truong, P. H., & Jeong, G. M. (2018). Position-Based Feature Selection for Body Sensors regarding Daily Living Activity Recognition. *Journal of Sensors*, *2018*, Article ID 9762098, 1-13.

12. Tufek, N., Yalcin, M., Altintas, M., Kalaoglu, F., Li, Y., & Bahadir, S. K. (2019). Human Action Recognition Using Deep Learning Methods on Limited Sensory Data. *IEEE Sensors Journal*, *20*(6), 3101-3112.

13. Jin, Y., Chen, G., Lao, K., Li, S., Lu, Y., Gan, Y., ... Luo, J. (2020). Identifying Human Body States by Using a Flexible Integrated Sensor. *npj Flexible Electronics*, *4*(28), 1-8.

14. Tan, Y., Yan, W., Huang, S., Du, D., & Xia, L. (2020). A Motion Deviation Image-based Phase

Feature for Recognition of Thermal Infrared Human Activities. *Engineering Letters*, *28*(1), 48-55.

15. Akula, A., Shah, A. K., & Gosh, R. (2018). Deep Learning Approach for Human Action Recognition in Infrared Images. *Cognitive System Research*, *50*, 146-154.
16. Mendes, P. A. S., Mendes, M., Coimbra, A. P., & Crisostomo, M. M. (2019). Movement Detection and Moving Object Distinction Based on Optical Flow. *Proceedings of the World Congress on Engineering (WCE) 2019*, 3-5.
17. Chen, C., Wang, Y., Yi, K., Wang, T., & Xiang, H. (2020). Semantic Analysis of Action with Spatio-Temporal Features Based on Object Detection. *Engineering Letters*, *28*(2), 616-623.
18. Chao, Z., Hidalgo, G., Simon, T., Wei, S. -E., & Sheik, Y. (2017). Real-time Multi-Person 2D Pose Estimation using Part Affinity Fields. *Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 7291-7299.
19. Ijjina, E. P., & Chalavadi, K. M. (2017). Human Action Recognition in RGB-D Videos using Motion Sequence Information and Deep Learning. *Pattern Recognition*, *72*, 504-516.
20. Ma, M., Marturi, N., Li, Y., Leonardis, A., & Stolkin, R. (2018). Region-sequence Based Six-Stream CNN Features for General and Fine-grained Human Action Recognition in Videos. *Pattern Recognition*, *76*, 506-521.
21. Antonik, P., Marsal, N., Brunner, D., & Rontani, D. (2019). Human Action Recognition with a Large Scale Brain-Inspired Photonic Computer. *Nature Machine Intelligence*, *1*(11), 530-537.
22. Xiaopeng, J., Jun, C., Wei, F., & Dapeng, T. (2018). Skeleton Embedded Motion Body Partition for Human Action Recognition using Depth Sequences. *Signal Processing*, *143*, 56-68.
23. Pham, H. -H., Khoudour, L., Crouzil, A., Zegers, P., & Velastin, S. A. (2018). Exploiting Deep Residual Networks for Human Action Recognition from Skeletal Data. *Computer Vision and Image Understanding*, *170*, 51-66.
24. Charles, J., Pfister, T., Everingham, M., & Zisserman, A. (2014). Automatic and Efficient Human Pose Estimation for Sign Language Videos. *International Journal of Computer Vision*, *110*(1), 70-90.
25. Banjongkan, A., Pongsena, W., Kerdprasop, N., & Kerdprasop, K. (2020). A Comparative Study of

Learning Techniques with Convolutional Neural Network Based on HPC-Workload Dataset. *International Journal of Machine Learning and Computing*, *10*(1), 10-17.

26. Chen, K. (2019). Sitting Posture Recognition Based on OpenPose. *Proceedings of 2019 4th International Conference on Insulating Materials, Material Application and Electrical Engineering (IMMAEE)*, 1341-1346.

27. Groos, D., Ramampiaro, H., & Ihlen, E. A. F. (2021). EfficientPose Scalable single-person pose estimation. *Applied Intelligence*, *51*, 2518-2533.