

Exercise Pose Detection Using 3D LiDAR: Mapping Human Movement

Mitchell Zinck, Ray Huda, Evan Webster, Marzieh Amini*

The School of Information Technology

*Department of Systems and Computer Engineering

Carleton University, Ottawa, Canada

{mitchellzinck, rayhuda, EvanTWebster}@email.carleton.ca, *marzieh.amini@carleton.ca

Abstract—Recently, light detection and ranging (LiDAR) technology has emerged as a promising approach to enhance sensing systems, particularly in human activity recognition (HAR). Advances in computer vision and LiDAR sensor development are revolutionizing HAR systems. This work proposes a novel method to detect and classify human exercise poses using 3D LiDAR technology by taking advantage of deep learning techniques. We utilize the Point Pillars network for 3D object detection and employ pseudo-image transformation for effective 3D to 2D data conversion, enhancing the applicability of 2D CNNs for 3D point cloud data analysis. To account for temporal dependencies in 3D point cloud data streams, we use the “smush” technique, which considers a few consecutive 3D LiDAR point cloud frames. This method significantly improves the detection accuracy of dynamic exercises, highlighting the potential of our approach in enhancing fitness monitoring technologies. This research not only demonstrates the effective use of 3D LiDAR in capturing complex activities but also lays the groundwork for future advancements in mapping human movement using 3D LiDAR.¹

Index Terms—LiDAR, 3D point cloud, exercise pose detection, convolutional neural networks, human activity recognition

I. INTRODUCTION

The integration of LiDAR technology with advanced deep learning models offers a promising avenue for human activity recognition (HAR), overcoming the limitations of traditional camera-based systems [1]. LiDAR or Radar sensors for daily activity monitoring provide significant advantages over other technologies, particularly in terms of privacy preservation and non-cooperative monitoring capabilities [2], [3]. Specifically, ambient sensors, especially vision-based ones, can raise concerns about data confidentiality and patient privacy, issues less prevalent with wearable sensors. However, wearable sensors require user cooperation and compliance to be worn or carried, which can be problematic and uncomfortable [4].

This work was motivated by the need for a more sophisticated analysis of human movements, leveraging the depth perception, privacy, and low-light performance advantages of 3D LiDAR. This paper details the development process of our neural network model, from initial conception to future directions, highlighting our methodology, results, and focusing on the potential impacts within health monitoring and fitness technologies. Our primary contribution of this work is the development of a novel multi-frame analysis methodology

model capable of detecting humans and classifying them into six main calisthenic exercises. The secondary contributions are the development of an accurate human detector model and the development of a three-pose detector to classify three calisthenic exercises with static movement. By achieving these objectives, we aim to address gaps in current HAR utilizing LiDAR, providing a more accurate and privacy-preserving method for exercise monitoring and analysis.

The remainder of the paper is organized as follows: Section II discusses the related work. Section III, explain the data acquisition details. Section IV outlines the project methodology and framework development. Sections V reports on the results of these developed frameworks. Discussion of results and areas of interest for future work are presented in Section VI, and Section VII concludes the paper.

II. RELATED WORK

Our approach began with a thorough review of existing work in human activity recognition (HAR) across visual, radar, and LiDAR modalities. This helped us assess the state-of-the-art techniques and understand the unique strengths of each method. In visual-based systems, Li et al. employed a Bi-LSTM network, achieving 96% accuracy in activity detection, which demonstrates the model’s strength in processing visual sequences and temporal data effectively [5]. Radar-based systems offer an advantage in user privacy and are capable of achieving over 90% accuracy without needing to capture identifiable images, making them ideal for privacy-sensitive applications [6]. On the other hand, existing LiDAR-based approaches predominantly focused on non-human detection, such as vehicle tracking, or used additional sensors to enhance detection accuracy [7].

The literature review then narrowed in on Point Pillars and PointNet models for LiDAR point cloud processing [10]. The Point Pillars network processes raw 3D data into a structured format compatible with convolutional neural networks (CNNs), supporting high-efficiency detection and classification. Recognizing the limitations of CNNs in temporal analysis, we explored recurrent models, specifically the RNN-based TimePillars approach, and examined the potential of LSTMs to add temporal context within point cloud data [8], [9]. While we chose Point Pillars and PointNet for their suitability in LiDAR-based HAR, this approach initially lacked temporal

¹This work was supported by Natural Sciences and Engineering Research Council (NSERC) of Canada.

¹<https://sites.google.com/view/lidar-h-a-r/home>

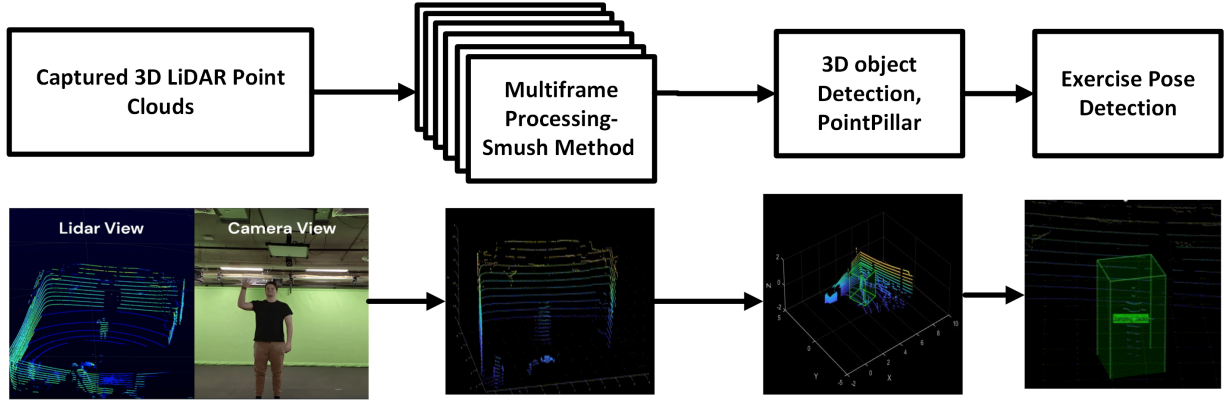


Fig. 1: Overview of the proposed method.

information. However, we address this challenge later in the paper with an innovative method to capture temporal aspects in three-dimensional point clouds, enhancing our model's accuracy for human activity recognition in dynamic scenes.

The Point Pillars network processes raw 3D LiDAR data into a structured format for analysis by CNNs. The key steps include:

3D Object Detection: Identifying and categorizing objects within the LiDAR point cloud data. **Point Cloud Partitioning:** Dividing the 3D space into a grid of pillars, simplifying the point cloud. **Feature Learning from Pillars:** Using a mini PointNet to learn features within each pillar, capturing attributes like shape and movement. **Pseudo-Image Generation:** Organizing learned features into a 2D pseudo-image for analysis by 2D CNNs. **Deep Learning for Object Detection:** Applying convolutional layers to the pseudo-image to recognize specific activities. **Activity Recognition Precision:** Achieving high precision in identifying and categorizing human movements. **Real-Time Data Processing:** Enabling real-time analysis and feedback, crucial for applications like fitness monitoring. **Adaptability and Scalability:** Handling diverse environments and activities, and extending the framework to various applications. This architecture significantly advances exercise detection and complex human movement analysis using 3DLiDAR data.

III. DATA ACQUISITION

In this section we explain the process of data acquisition. We initially setup our 3D velodyne puck 16-channel LiDAR in a relatively open room ($5m \times 10m$) and gathered data of a single person conducting general movements in the space several times. The data was then labeled using the Matlab tool LiDARLabeler. Initially, labeling was conducted solely based on the position of the person (Human Detector), and subsequently, the classified activity being performed. The same setting was then used to gather data of a single subject performing 3 actions (Three-Pose Detector). Finally, data was gathered in a large green screen room ($10m \times 12m$) to increase the range of detection and added twelve various subjects (13 in total) to increase diversity of the dataset. These subjects also performed all six actions (Six-Pose Detector) in a randomized order to ensure variance in the dataset. Figure 2 below, depicts

an example of each of the 6 activities that were performed and classified (a-f).

As shown in Table I, our full database consists of a total of 28614 frames of data. This novel database is available free and open-source to individuals and researchers with proper accreditation [14].

TABLE I: Number of Frames per Dataset

Dataset	Frames
VLP-16	11505
Alpha Prime-128	9938
WalkStandCrouch	5752
Human Detection	1086
Misc	333
Total	28614

IV. METHODOLOGY

The first stage in our development process was creating a human detector model capable of identifying human figures within LiDAR point clouds with high precision. This model laid the groundwork for subsequent activity detection by accurately locating and classifying human presence.

Building upon the human detector, we developed models to recognize three basic poses: standing, walking, and crouching, which were then expanded to detect six exercises: walking, standing, squatting, push-ups, sit-ups, and jumping jacks. This expansion required significant modifications to the network architecture to accommodate the increased complexity of movements.

To enhance detection accuracy for dynamic activities, we introduced the "smush" technique, a multi-frame analysis approach. This method improves our model's ability to recognize and classify continuous movements by incorporating temporal information across sequential 3D point cloud frames. Inspired by the "TransPillars" method [12], which employs coarse-to-fine aggregation for multi-frame 3D object detection through dwell time for richer object information, our technique takes a more simplified approach by directly concatenating the LiDAR frames to provide additional context. Fig. 1 shows the overview of the proposed method.

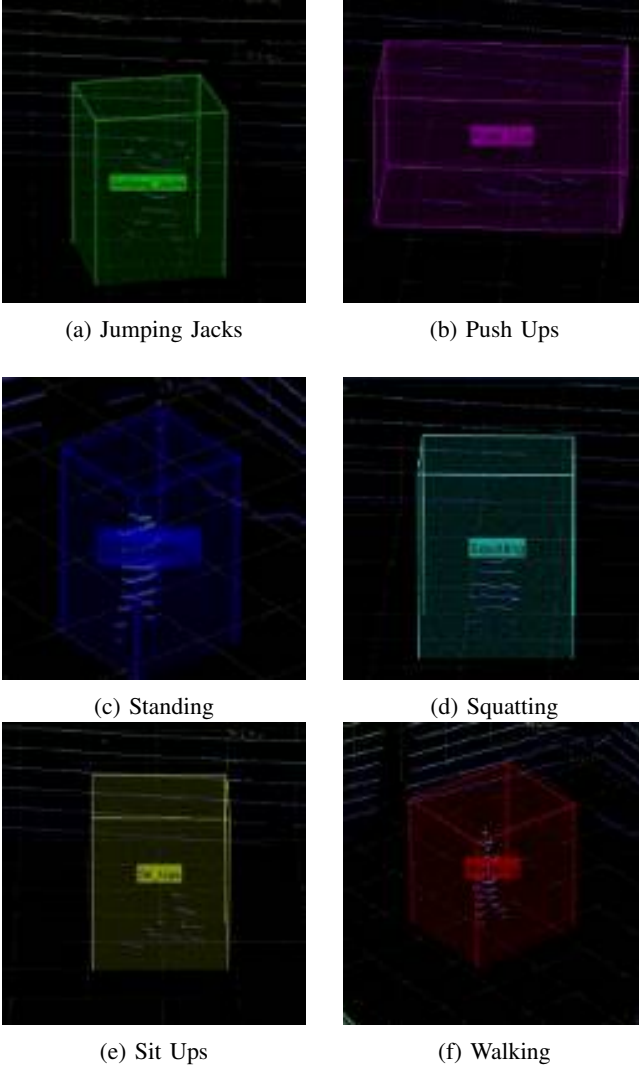


Fig. 2: (a-f) Examples of the 6 different classified activities.

V. EXPERIMENTAL RESULTS

A. Design and Structure of Detectors

Implementation of this work has been done in MATLAB due to its integration capabilities with our hardware Velodyne 16-channel LiDAR including robust LiDAR data handling capabilities; inbuilt tool kits, libraries; and LiDARLabeler algorithm and application for data labelling during preprocessing [13]. These assets allowed the project to focus on the development and optimization of a baseline exercise recognition platform.

B. Model Training and Validation

We employed a split of 70% training and 30% testing data from the collected point clouds. Training involved adjusting parameters such as confidence thresholds and epochs to optimize model performance as measured by the following metrics: AOS (Average Overlap Score), which depicts the accuracy of bounding box placement; and AP (Average Precision), which depicts average correct exercise classification.

The models underwent rigorous validation to ensure their accuracy and robustness in exercise detection.

a) Parameter Optimization: Throughout the development process, we continuously refined our models by varying parameters and incorporating feedback loops. This iterative approach allowed us to enhance model performance systematically, ensuring high precision and recall in exercise detection. Our final model employed 5430 frames of data, had a confidence threshold of 25%, and training was conducted for a total 100 epochs.

b) Implementation Challenges: One of the primary challenges we encountered was ensuring the model's generalizability across different individuals and environmental conditions. Through continuous dataset expansion, iterative testing and refinement, we aimed to develop a system robust enough to handle the inherent variability present in real-world applications. Another key challenge was regarding data formatting. The program used to record the LiDAR data records the file in a .PCAP format instead of a .PCD file as needed. This required additional data preprocessing and conversion prior to data utilization.

c) Results: As depicted in Table II, the results showcase that for highly static exercises like standing and crouching, the model showed high precision prior to applying the "smush" technique, as evidenced by AOS and AP values above 0.90.

TABLE II: Exercise Poses and resulting AOS and AP values for three-pose detection without smush

Exercise Pose	AOS	AP
Standing	0.90532	0.98498
Walking	0.78425	0.78499
Crouching	0.99882	1.00000
AVERAGE	0.89613	0.92332

As shown in Table III, by taking advantage of multi-frame analysis, known as the "smush" technique, significantly enhanced our model's performance across all exercises. Specifically, this method improved the accuracy in detecting dynamic movements, a challenge that had persisted in earlier models. The application of the "smush" technique led to noteworthy improvements in Average Overlap Score (AOS) and Average Precision (AP). In comparison of the average accuracies obtained, the five-frame smush model performed approximately ten times better than the no smush model, indicating a significantly more nuanced understanding of temporal sequences in exercises.

VI. DISCUSSION

A. Results Analysis

The results presented in Table III show the effectiveness of our "smush" technique, particularly in the context of dynamic exercises such as push-ups and sit-ups, where AOS and AP values improved markedly. The improved performance metrics attest to the method's ability to interpret temporal sequences in exercise movements more accurately. A comparative analysis with our previous system indicates a substantial leap in precision and reliability for our pose detection methodology.

TABLE III: Exercise Poses and resulting AOS and AP values for six-pose detection without smush (left) and five frame smush (right)

Exercise Pose	No Smush		Five-Frame Smush	
	AOS	AP	AOS	AP
Standing	0.00616	0.00622	0.23521	0.23529
Walking	0.02153	0.02153	0.14534	0.14534
Squatting	0.04623	0.04623	0.19341	0.19341
Push-ups	0.005806	0.005806	0.28867	0.28867
Sit-ups	0.02331	0.02334	0.28723	0.28726
Jumping Jacks	0.02457	0.0246	0.11619	0.11619
AVERAGE	0.021267	0.021289	0.21101	0.21103

One of the most compelling insights from our research is the critical role of temporal context in human pose recognition. The limitations of static frame analysis became apparent when exercises with similar initiation phases, like the transition from standing to squatting, were misclassified. However, the "smush" technique's multi-frame analysis provides a broader temporal scope, enabling the model to discern the continuity of movement, thereby significantly reducing false positives. In contrast to our previous model that utilized no temporal consideration, our new approach demonstrates positive trajectory in consistently accurate pose identification, especially in suboptimal lighting conditions and where privacy concerns preclude the use of cameras. The comparison with standard camera-based systems is especially favorable, highlighting the advantages of LiDAR in scenarios where optical devices fall short, such as privacy-sensitive areas. Additionally, our findings suggest a disparity between the detection of static and dynamic poses, with the latter posing more significant challenges in terms of temporal analysis and movement fluidity. This dichotomy emphasizes the necessity for advanced modeling techniques that can capture the intricacies of human movement without compromising on real-time analysis capabilities.

B. Challenges and Insights

During the course of our research, we encountered challenges that are inherent in the field of human pose estimation. One primary issue was the initial model's inability to distinguish between similar exercises that share overlapping movement patterns. For instance, the transition from standing to squatting occasionally caused false classifications. The application of the "smush" technique has been instrumental in overcoming this, as it augments the spatial data with broader temporal context, providing a more holistic view of the subject's movements by allowing for better differentiation based on the progression of movements rather than isolated frames. The insights gained from this study have significant implications for the development of LiDAR-based exercise detection systems.

VII. CONCLUSION

This study marks a significant advancement in exercise pose detection using LiDAR and deep learning. Our development of a human detector and subsequent exercise detection models

has demonstrated 3DLiDAR technology's potential in accurately mapping human movement. The successful implementation of the "smush" technique highlights the importance of temporal analysis in understanding dynamic activities. By considering the temporal dependencies across multiple point cloud frames, we achieved an average improvement in AP and AOS by 891% across various exercise poses, showcasing the impact of this approach. Looking forward, integrating more sophisticated neural network architectures and exploring high-resolution 3D LiDAR data offer promising avenues for further progress in exercise detection. By addressing current challenges and leveraging our findings, this research contributes valuable insights to the evolving fields of health monitoring and fitness technology. Our work not only illustrates the powerful role 3D LiDAR can play in exercise detection but also sets the stage for future innovations that could redefine physical training, rehabilitation, and personal fitness monitoring, while maintaining user privacy as a central priority.

REFERENCES

- [1] J. Roche, V. De-Silva, J. Hook, M. Moencks and A. Kondo, "A Multimodal Data Processing System for LiDAR-Based Human Activity Recognition," in *IEEE Transactions on Cybernetics*, vol. 52, no. 10, pp. 10027-10040, Oct. 2022.
- [2] H. Sadreazami, M. Amini, M. Bolic and S. Rajan, "Unsupervised Learning of Fall Incidents using Radar-based Sensing," 2025 IEEE International Symposium on Circuits and Systems (ISCAS), pp. 1-5, 2025.
- [3] H. Sadreazami, A. Khoyani, M. Amini, S. Rajan and M. Bolic, "Radar Based Fall Detection with Imbalance Data Handling and Data Augmentation," 2023 IEEE Sensors Applications Symposium (SAS), pp. 1-4, 2023.
- [4] H. Sadreazami, M. Bolic and S. Rajan, "Contactless Fall Detection Using Time-Frequency Analysis and Convolutional Neural Networks," in *IEEE Transactions on Industrial Informatics*, vol. 17, no. 10, pp. 6842-6851, 2021.
- [5] O. Rinchi, H. Ghazzai, A. Alsharoa, and Y. Massoud, "LiDAR technology for Human Activity Recognition: Outlooks and challenges," *IEEE Internet of Things Magazine*, vol. 6, no. 2, pp. 143-150, Jun. 2023.
- [6] A. D. Singh, S. S. Sandha, L. Garcia, and M. Srivastava, "Radhar," *Proceedings of the 3rd ACM Workshop on Millimeter-wave Networks and Sensing Systems*, Oct. 2019.
- [7] H. Li, A. Shrestha, H. Heidari, J. Le Kerne, and F. Fioranelli, "Bi-LSTM Network for multimodal continuous human activity recognition and fall detection," *IEEE Sensors Journal*, vol. 20, no. 3, pp. 1191-1201, Feb. 2020.
- [8] E. L. Calvo et al., "Timepillars: Temporally-recurrent 3D LiDAR object detection," 2023, <https://doi.org/10.48550/arXiv.2312.17260>.
- [9] R. Huang et al., "An LSTM approach to temporal 3D object detection in LiDAR point clouds," In *Computer Vision-ECCV*, pp. 266-282. Springer International Publishing, 2020.
- [10] "LiDAR 3-D Object Detection Using Point Pillars Deep Learning - MATLAB & Simulink," [www.mathworks.com](https://www.mathworks.com/help/LiDAR/ug/object-detection-using-pointpillars-network.html). <https://www.mathworks.com/help/LiDAR/ug/object-detection-using-pointpillars-network.html> (accessed Nov. 14, 2023).
- [11] "Point Cloud Classification Using PointNet Deep Learning - MATLAB & Simulink," [www.mathworks.com](https://www.mathworks.com/help/vision/ug/point-cloud-classification-using-pointnet-deep-learning.html). <https://www.mathworks.com/help/vision/ug/point-cloud-classification-using-pointnet-deep-learning.html> (accessed Nov. 14, 2023).
- [12] Z. Luo et al., "Transpillars: Coarse-to-fine aggregation for multi-frame 3D object detection," in 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp. 4219-4228, 2023.
- [13] "Puck LiDAR Sensor, High-Value Surround LiDAR," [Velodyne LiDAR](https://velodynelidar.com/products/puck/). <https://velodynelidar.com/products/puck/> (accessed Nov. 14, 2023).
- [14] M. Amini, R. Huda, E. Webster, and M. Zinck, 3D LiDAR Exercise Pose Detection Dataset, <https://doi.org/10.5683/SP3/UFDR7Q>.
- [15] Gupta, N., Gupta, S.K., Pathak, R.K. et al. Human activity recognition in artificial intelligence framework: a narrative review. *Artif Intell Rev* 55, 4755-4808 (2022).