Exercise Pose Detection with LIDAR: Mapping Human Movement

Ray Huda, Evan Webster, Mitchell Zinck, Marzieh Amini
School of Information Technology,
Carleton University, Ottawa, Canada
rayhuda@cmail.carleton.ca, evanwebster@cmail.carleton.ca, mitchellzinck@cmail.carleton.ca, marzieh.amini@carleton.ca

***Abstract*—This research presents a comprehensive approach to detecting and classifying human exercise poses using 3D LIDAR technology coupled with deep learning methodologies. Our study builds upon the foundational work of point cloud processing, introducing a series of deep learning models to recognize and classify human exercises accurately. The project progressed through initial concept, design, implementation of a human activity recognition model, and culminated in the development of a convolutional neural network (CNN) model that processes 3D LIDAR data for exercise pose detection. This paper outlines our methodology, including the novel use of the Point Pillars network and pseudo-image transformation for effective 3D to 2D data conversion, enhancing the applicability of conventional 2D CNNs for 3D spatial data analysis. By leveraging this network and developing bespoke CNNs, we mapped human movements to detect initially three, then six key exercises, overcoming traditional limitations of 2D vision-based systems. The subsequent application of the "smush" technique, demonstrated a significant improvement in the detection accuracy of the additional dynamic exercises, underscoring the potential of our approach in enhancing fitness monitoring technologies. We further discuss our iterative development process, the challenges faced, and propose directions for future research. This research not only demonstrates the effective use of LIDAR in capturing complex activities, but also lays the groundwork for future enhancements in mapping human movement using LIDAR.**

***Index Terms*—LIDAR, Point Cloud, Exercise Pose Detection, Point Pillars, Convolutional Neural Networks, CNN, Human Activity Recognition, HAR**

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 [4]. Our project was motivated by the need for a more sophisticated analysis of human movements, leveraging the depth perception, privacy, and low-light performance advantages of 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 outlines the project methodology and framework development. Sections IV reports on the results of these developed frameworks. Discussion of results and areas of interest for future work are presented in Section V, and Section VI concludes the paper.

II. Related Work

Our methodology involved iterative development stages, starting with a comprehensive literature review, followed by the design and implementation of multiple deep learning models tailored for LIDAR data.

A. Literature Review

We commenced our literature review with a focus on Point Pillars and PointNet models for their effectiveness in object detection and classification within LIDAR point clouds. This foundational work informed our approach to transforming 3D point clouds into a format suitable for CNN analysis, crucial for recognizing human exercises in three-dimensional space [1]. As shown in Figure 1, the Point Pillars network transforms raw 3D LIDAR data into a structured format, enabling effective analysis by CNNs.

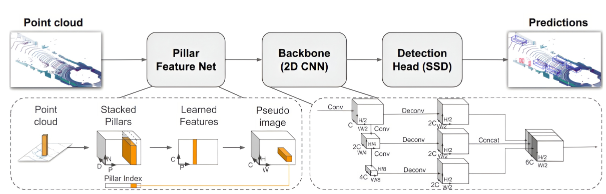

Figure 1: Network overview of Point Pillars.

The Point Pillars network process involves several key steps, transforming raw 3D LIDAR data into a structured format that can be effectively analyzed by convolutional neural networks (CNNs). Below is a detailed overview of the network process, illustrated by the included Point Pillars network diagram.

a) 3D Object Detection: Point Pillars starts with the detection and categorization of 3D objects within the LIDAR point cloud data. This is critical for identifying spatial movement and is the foundation for further analysis [1].

b) Point Cloud Partitioning: The process begins by partitioning the 3D space into a grid of pillars, simplifying the complex point cloud into a manageable structure. Each pillar represents a vertical column of LIDAR points, facilitating a simplified analysis of the space [1].

c) Feature Learning from Pillars: Within each pillar, a feature is learned using a mini PointNet architecture. This step captures the unique attributes of all points within the pillar, including their shape and movement, which is crucial for identifying exercises and human poses [1].

d) Pseudo-Image Generation: The learned features from each pillar are then organized into a 2D pseudo-image based on their spatial locations. This transformation allows the application of conventional 2D CNNs to analyze the data, leveraging the power of established image recognition technologies [1].

e) Deep Learning for Object Detection: The pseudo-image undergoes a series of convolutional layers, where patterns and relationships within the data are extracted. This culminates in the final layer, where specific activities are recognized through learned object detection techniques, enabling the identification of various exercises [1].

f) Activity Recognition Precision: A key feature of Point Pillars is its ability to process LIDAR data with high precision, accurately identifying and categorizing human movements. This precision is essential for the effective recognition of motion activities [1].

g) Real-Time Data Processing: Designed for efficiency, Point Pillars facilitates the real-time processing of LIDAR data. This capability is vital for applications requiring immediate analysis and feedback, such as in fitness monitoring and rehabilitation contexts [2].

h) Adaptability and Scalability: The adaptable architecture of Point Pillars allows it to handle a diverse range of environments and activities. Its scalability ensures that the framework can be extended beyond the initial project scope to encompass a wider array of applications [2].

The incorporation and development of Point Pillars into our project represents a significant innovation in the field of exercise detection, providing a robust framework for analyzing complex human movements through LIDAR data.

III. Methodology

A. Detector Development

a) Human Detector: 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.

b) Three-Pose and Six-Pose Detectors: 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.

c) Implementation of the "Smush" Technique: To enhance the detection accuracy of dynamic activities, we introduced the "smush" technique, a multi-frame analysis approach. This method significantly improved our model's ability to recognize and classify continuous movements by incorporating temporal information across sequential frames.

B. Design and Structure of Detectors

The design process involved a thorough hardware and literature review; critical selection between Python and MATLAB frameworks based on the available libraries and ease of use; and the decision to utilize the Point Pillars network for data processing. Implementation in MATLAB was selected for its integration with our given Velodyne, sixteen channel, LIDAR hardware; robust LIDAR data handling capabilities; inbuilt tool kits, libraries; and LIDARLabeler algorithm and application for data labelling during preprocessing [3]. These assets allowed the project to focus on the development and optimization of a baseline exercise recognition platform.

a) Human Detector: The human detector, our initial model, was designed to accurately identify human figures within LIDAR data. Using a 50% confidence threshold, it achieved an Average Object Size (AOS) of 0.94469 and an Average Precision (AP) of 0.99773, demonstrating its effectiveness in human detection within complex environments.

b) Three-Pose Detector: Expanding upon the human detector, the three-pose detector was capable of recognizing walking, standing, and crouching poses. Modifications to the network allowed it to handle an increased range of actions, setting the stage for more sophisticated exercise detection.

c) Six-Pose Detector: The development of the six-pose detector marked a significant advancement in our project. This model extended our capability to identify six exercises: walking, standing, squatting, push-ups, sit-ups, and jumping jacks. The complexity of training increased exponentially with the addition of dynamic movements, necessitating the implementation of multi-frame analysis to enhance detection accuracy.

d) Multiframe Analysis ("Smush"): To address the limitations observed in detecting dynamic activities such as walking, we introduced the "smush" technique. This multi-frame analysis method improved the model's performance by considering sequential frames, thereby capturing the temporal aspects of human movement more accurately.

C. Data Collection and Preprocessing

Data collection involved capturing LIDAR point clouds from 13 individuals performing six key exercises in varied environments. These point clouds were then converted into pseudo-images using the Point Pillars framework, enabling the application of CNNs for activity recognition.

As shown in Table 1, 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.

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

D. 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.

IV. Results

As depicted in Table 2, 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.

As depicted in Table 3, the introduction 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.

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

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.

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

	No Smush		Five-Frame Smush	
Exercise Pose	AOS	AP	AOS	AP
Standing	0.0061645	0.0062156	0.23521	0.23529
Walking	0.021534	0.021534	0.14534	0.14534
Squatting	0.046228	0.046228	0.19341	0.19341
Push-ups	0.0058065	0.0058065	0.28867	0.28867
Sit-ups	0.023306	0.023342	0.28723	0.28726
Jumping Jacks	0.024567	0.02461	0.11619	0.11619
AVERAGE	0.021267	0.021289	0.21101	0.21103

V. Discussion

A. Results Analysis

The results presented in Table 3 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. 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 example, 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 contect, 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. Our approach not only advances the field in terms of technology but also in terms of methodology, as it underscores the importance of incorporating temporal dynamics into the analysis. The utilization of a multi-frame approach could be considered a best practice for future studies aiming to achieve high accuracy in human activity recognition. In conclusion, the analysis of our results affirms the potential of the ”smush” technique in enhancing exercise pose detection. By offering a solution to one of the primary challenges in LIDAR-based recognition systems—the accurate classification of dynamic movements—our methodology sets the stage for the development of more sophisticated and reliable human activity monitoring systems.

C. Future Work

Our research opens several avenues for future exploration and improvement:

a) Expansion of Exercise Library: By incorporating a wider variety of exercises, especially those with subtle distinctions, we can refine our model's sensitivity to nuanced movements, enhancing its utility across diverse fitness regimes.

b) Integration of Advanced Neural Networks: Incorporating advanced neural networks such as RNNs and LSTMs could offer a more sophisticated framework for capturing and enhancing temporal dependencies in exercise sequences. This approach is expected to further improve the model's ability to process dynamic movements and transitions between exercises and therefore increase the model's accuracy.

c) High-Resolution LIDAR Data: Exploring the use of 128-channel LIDAR data promises higher resolution and detail in captured movements. This could lead to more accurate models by providing richer datasets for training and validation.

d) Diverse Environmental Conditions: Testing our model in varied environmental settings and lighting conditions will ensure its robustness and applicability in real-world scenarios. This is crucial for developing a system that remains effective regardless of external factors.

e) Real-Time Feedback Systems: Developing a real-time feedback mechanism for users performing exercises could significantly impact fitness training and rehabilitation by providing immediate corrective guidance, enhancing the effectiveness of workouts and reducing the risk of injury.

VI. Conclusion

This study represents a significant step forward in the field of exercise pose detection using LIDAR and deep learning. Our development of the human detector and subsequent models for exercise detection have demonstrated the potential of LIDAR technology in accurately mapping human movement. The successful implementation of the ”smush” technique underscores the importance of temporal analysis in understanding dynamic activities. Looking ahead, the integration of advanced neural network architectures and the exploration of high-resolution LIDAR data hold promise for further advancements in this area. By addressing the challenges encountered and capitalizing on the insights gained, this paper contributes to the ongoing innovation in health monitoring and fitness technology. Our research not only showcases the potential of LIDAR in exercise detection but also paves the way for future developments that could revolutionize how we approach physical training, rehabilitation, and personal fitness monitoring.

References

[1] “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 (accessed Nov. 14, 2023).

[2] “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 (accessed Nov. 14, 2023).

[3] “Puck LIDAR Sensor, High-Value Surround LIDAR,” Velodyne LIDAR. https://velodyneLIDAR.com/products/puck/ (accessed Nov. 14,2023).

[4] J. Roche, V. De-Silva, J. Hook, M. Moencks and A. Kondoz, ”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, doi: 10.1109/TCYB.2021.3085489. https://ieeexplore.ieee.org/document/9464313

[5] 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). https://link.springer.com/article/10.1007/s10462-021-10116-xciteas