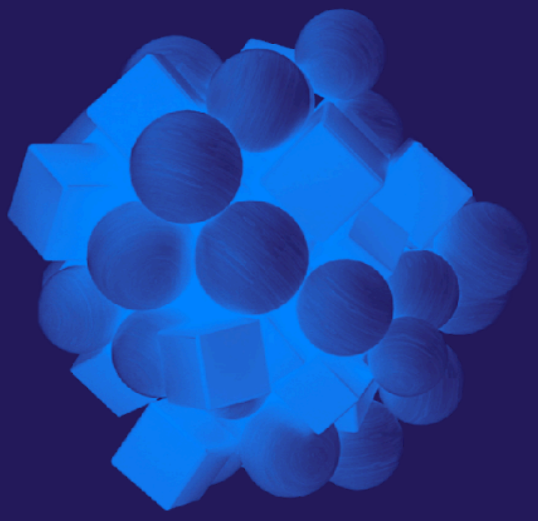
Person Detection with Overhead 3D LiDAR Technology demo report



Inspired by Konecranes

Elinkeino-, liikenne- ja ympäristökeskus
Työ- ja elinkeinoministeriö
Arbets- och näringsministeriet
Tampereen ammattikorkeakoulu
Aalto-yliopisto
Aalto-universitetet
Aalto University
Metropolia
HAMK
Hämeen ammatti-korkeakoulu
OAMK
OULUN AMMATTIKORKEAKOULU
Tampereen yliopisto

Person Detection with Overhead 3D LIDAR on Crane

Aalto University
24.11.2025
TECHBOOST Project Report

TECHBOOST AI! Aalto-yliopisto www.techboost.fi

Contents

1. Introduction

People often work close to overhead cranes where they are engaged in tasks such as moving loads, guiding equipment, or simply passing through these spaces. This project aims to create a solution for human detection using an overhead 3D Lidar in for the crane and therefore addresses the problem of robustly detecting the locations of people in the crane working area.

Additionally, the sensor mode allows for a secure way to handle processing human detection information by using 3D Lidar point cloud data rather than a camera feed. Specifically, LiDAR measures distance only and therefore does not capture clear images of people's faces or clothing. That makes it easier to protect privacy while still understanding where people are in the crane area. Upon receiving the point cloud data, the system is able to detect multiple people in the crane area by utilizing machine learning based methods.

The outputs of the system can subsequently be used in various applications such as safe autonomous operation of the crane and the navigation of smart robotics that may interact with people. Essentially, the crane controls and robots can slow down, stop, or reroute based on where people are. Additionally, operators and workers get a clearer awareness of human presence and position in the area.

Thus, the project provides a solution to the crucial problem of safe and private person detection in a work area. Furthermore, the system will work even in challenging indoor environments such as low light, glare, or dust, which are cases where cameras often fail. Moreover, since the outputs are standard 3D boxes, it can be plugged seamlessly into existing software such as Open3D and ROS2+Rviz for visualization. The approach of the project can also be extended to more sensors as needs grow. For instance, multiple LiDARs can be combined to remove blind spots and obtain wider coverage in larger areas.

2. Project Overview

2.1 Background

Modern worksites need reliable ways to know where people are so that machines can operate safely around them. Various methods exist for detecting people with different types of sensors. Cameras are common since they allow for high detail vision, and many object detection algorithms have been developed for camera images. However, they do not directly measure depth information, which is crucial for 3D localization, and moreover, they collect identifiable images of people. For safety-critical use cases such an overhead crane moving loads, these limits matter.

An alternative to camera images is to use a 3D lidar sensor, which is a sensor that can provide depth information about the image. The sensor provides depth information as a cluster of measured 3D points, or a point cloud, which can later be used to detect objects in a scene. In addition to the more detailed geometry in the image, the point cloud preserves the privacy of people in the line of sight of the sensor by not utilizing sensitive image data.

However, the detection of people in a 3D point cloud is a challenging balance of performance and accuracy. The sensor produces large volumes of data, and the detection model must run fast enough to be deployable in real time. The additional challenge here is the overhead viewpoint (top-down), for which there is a lack of public dataset. Henceforth, our approach addresses: (1) a compact and efficient model fine-tuned for the overhead angle, and (2) an edge deployment that runs close to the sensor. Together, the system can output human-detections, with sufficient accuracy, that can be used for safety critical autonomous operation of the crane.

2.2 Objectives & Scope

The goal of the project is to deliver a human detection system for the overhead crane at Aalto that identifies the instances of people within the crane's working area to support safe operations and future autonomy.

The system utilizes a Jetson microcomputer (with ROS 2 for publishing detections on ROS topics for easy integration), and an overhead 3D Lidar which gives point clouds in real-time as input to the prediction model that subsequently provides the detections of people working inside the crane area. Using this setup, a small dataset was collected and manually annotated with a 3D bounding box labelling tool with the aim of fine-tuning a suitable model on site-specific data. A pretrained model is then used as a base and finetuned with the collected dataset for the overhead view. Visualization tools used for monitoring and testing purposes include RViz and Open3D.

The project explicitly excludes person identification and also doesn't cover downstream tasks such as intent estimation or activity recognition. Also, the overhead LiDAR position is fixed and the performance is evaluated on data collected in typical lighting and dust levels (extreme ambient conditions are out of scope). Further, the project solely provides the detection layer which can be used in the future for tasks such as close-loop crane control.

2.3 Key Technologies & Innovations

Core technologies:

- Sensor and compute: RoboSense overhead LiDAR and NVIDIA Jetson Orin NX (JetPack 6.2.1).
- Middleware: ROS 2 Humble for messaging, integration, and lifecycle management.
- Detection model: PointPillars (open-source), fine-tuned on the collected overhead LiDAR dataset using transfer learning.
- Tooling: labelCloud for 3D annotations, RViz and Open3D for visualization, and Python-based code creation (evaluation and export utilities).

Methodology highlights:

- Overhead-view adaption: we labeled a small viewpoint-specific dataset to reduce domain gap with standard (front/side) datasets.
- Edge optimized: pipeline is built to be able to run on-device, reducing network dependency and giving the ability to keep raw data local.
- Dual evaluation schemes: we report metrics with ground truths (for accuracy) and without labels using TTA (for robustness).

What stands out:

- Improved efficiency/cost savings as the system runs on a compact, low power Jetson module and requires no external server or cloud. The setup of simple ROS topic interface accelerates integration with existing stacks.
- A novel advancement is made evident as the project tailors a proven detector (primarily on autonomous driving datasets) to a rare overhead LiDAR viewpoint, thus filling a gap in available solutions and datasets. Additionally, there is privacy by design (as a result of no images), which eases stakeholder concerns and reduces compliance overhead.
- Scalability and adaptability are ensured as the fine-tuning pipeline plus clear data and labeling recipe means that the system can be replicated to new sites conveniently by collecting a small local dataset and fine-tuning the model using the software provided. The modular ROS architecture allows drop-in use with additional extensions such as tracking or fusion.
- Future sustainability and environmental benefits can be realized through enhanced situational awareness that can potentially reduce unnecessary crane movements and idle time which consequently contributes to energy savings over time.

User value:

- Safety teams can get reliable privacy respecting information of people working or moving in the crane area.
- Engineers and integrators get a module that they can connect to alarms, fences, or path-planning activities.
- Project owners and funders will get a solution that is immediately deployable and adaptable to other sites, as well as built on open, maintainable components.

3. Implementation Strategy (Technical)

3.1 Design & Development

As stated in the previous section, the on-board computer (Jetson Orin NX) was flashed with Jetpack 6.2.1 and is ROS2 and CUDA enabled. The hardware acceleration of the internal GPU allows for fast inference of the machine learning models. The lidar publishes a topic of the point clouds which are fed into the finetuned PointPillar model through the ROS interface and the bounding box detections are published in a separate topic. The visualization is using the RViz tool which shows the point cloud and the human bounding boxes (Figure 1 presents an example).

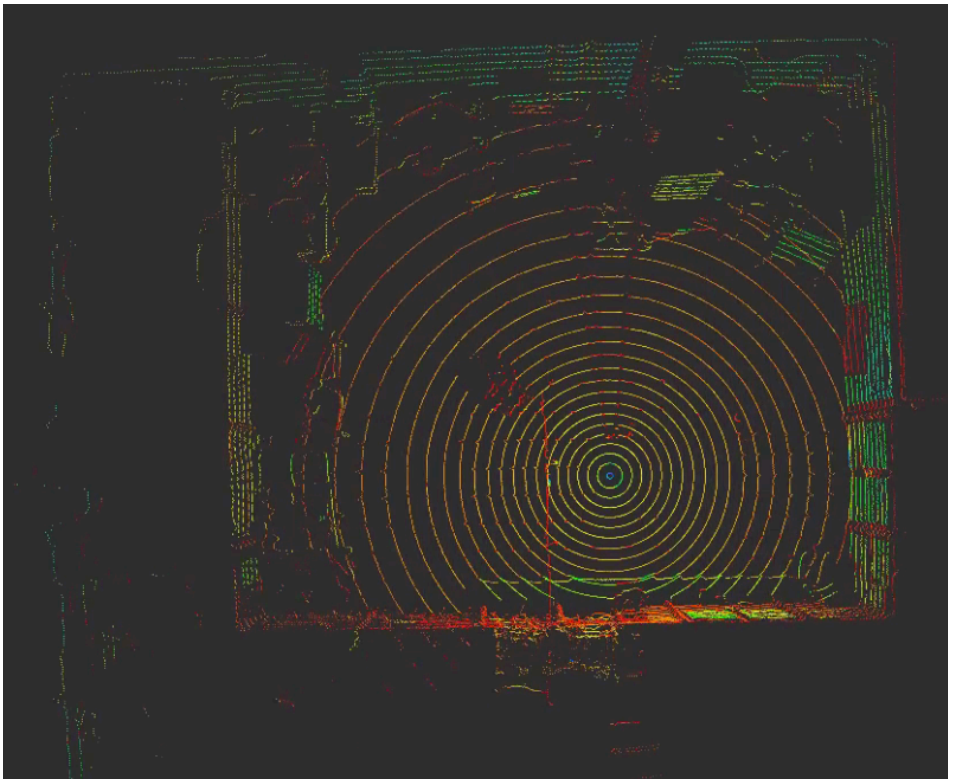


Figure 1: Live Rviz visualization of the working area, represented as a point cloud from the LiDAR.

Figure 2 presents the connection architecture. In compiling the training data, point clouds were labeled with labelCloud software, and the OpenPCDet pretrained PointPillar model was finetuned with 30 annotated point clouds.

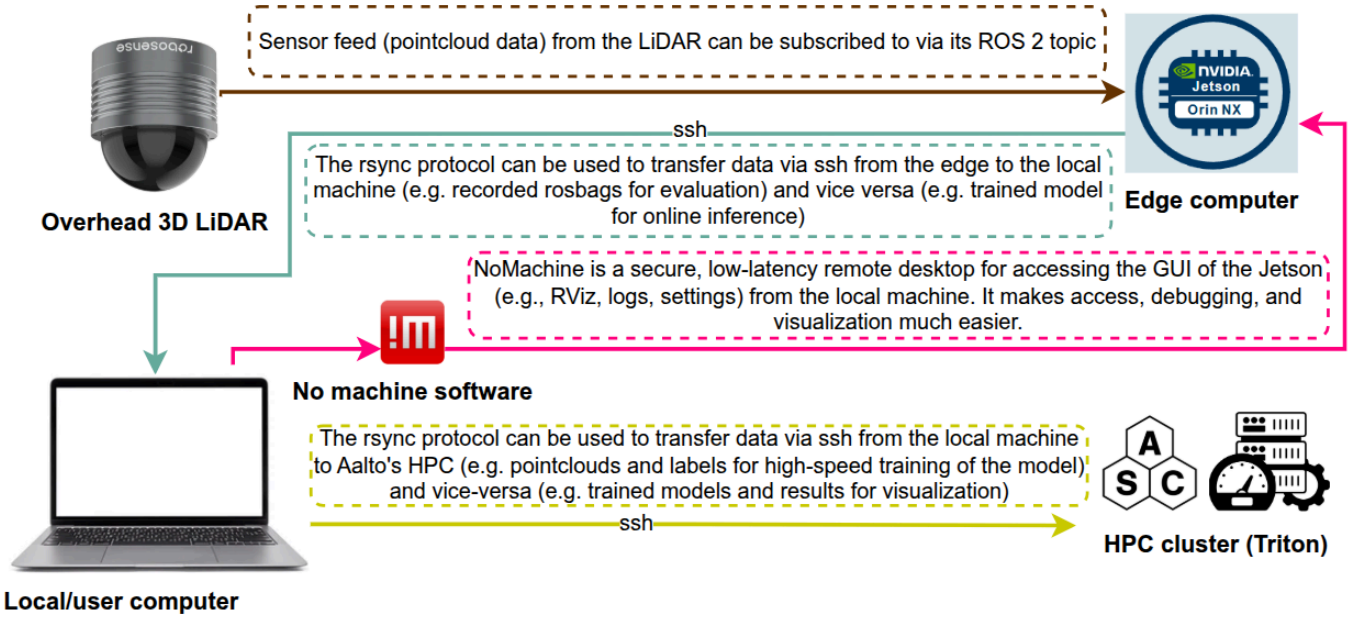


Figure 2: Main elements and connections of the implemented system.

Figure 3 below shows the final training plot. The horizontal axis is training time (epochs) and the vertical axis is "Loss, which is a single number that tells how incorrect the model's predictions are and hence, the value which we intend to reduce as much as possible. It is clear from the figure that as training progresses, both lines drop quickly and then level off into a low, stable range close to a loss value of 0. The key point is that both the training and validation lines move together and end up close to each other, which means that the model isn't memorizing the

training data, but doing well on unseen data (good generalization). Small wiggles are expected and simply reflect the randomness in the data and optimization. Overall, this shape of a steady decline followed by a plateau with training and validation closely aligned, indicates a successfully trained model.

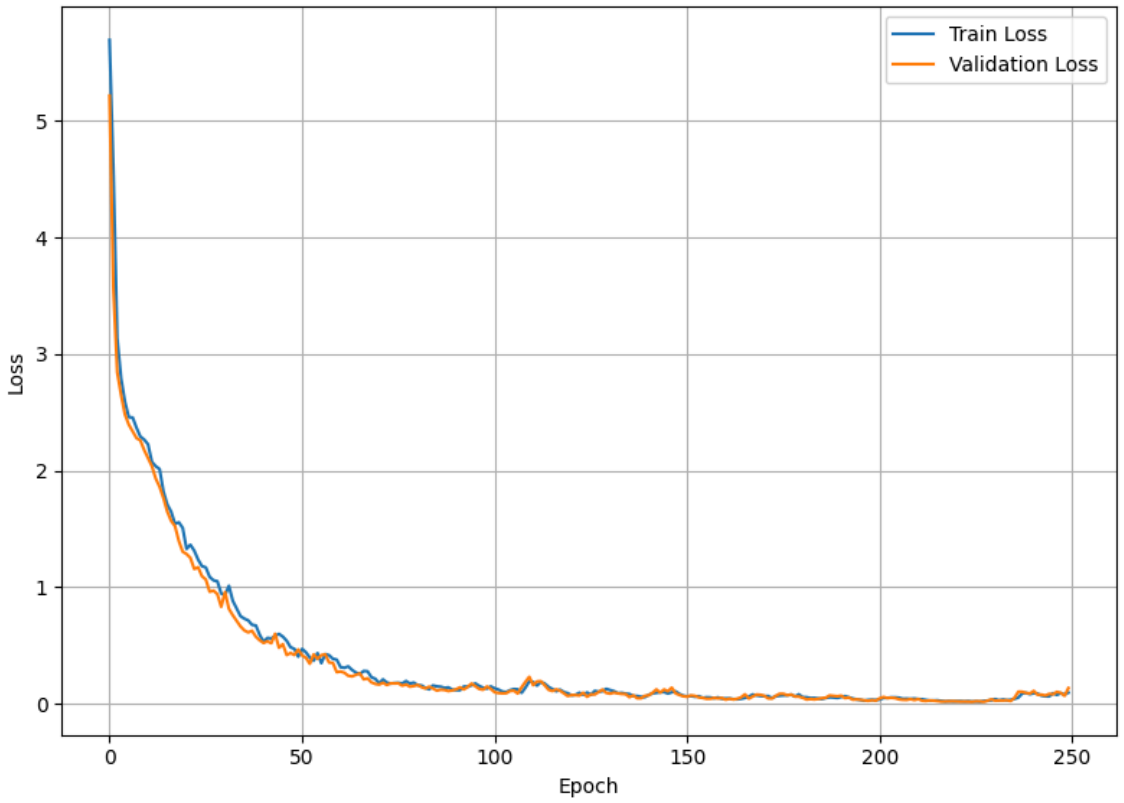


Figure 3: Loss curves generated during the training of the LiDAR person detection PointPillars model.

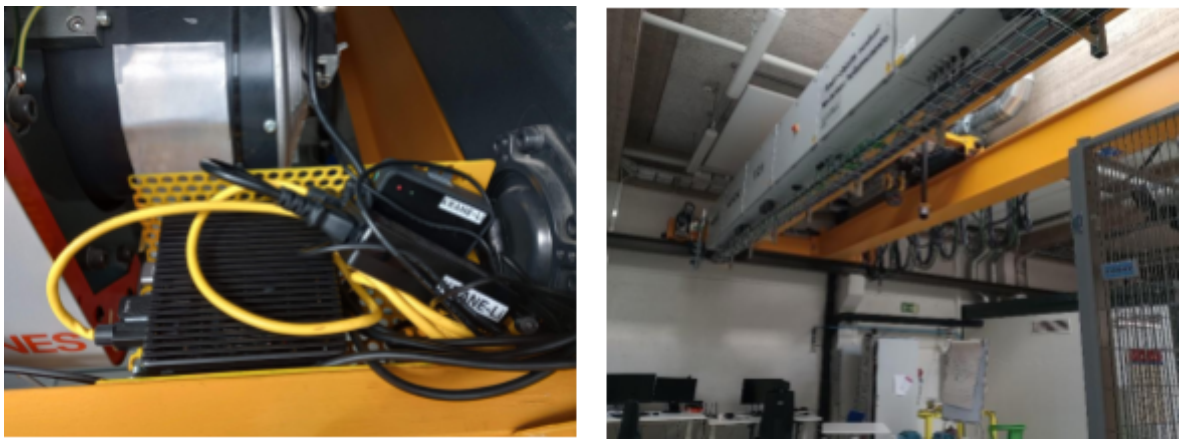


Figure 4: Overhead LiDAR installation in the Aalto crane lab.

3.2 Testing and validation strategy

As mentioned previously, data was collected inside the Aalto-Crane laboratory using the overhead LiDAR sensor (Figures 4 and 5). Participants moved naturally—walking, standing, and occasionally sitting—beneath the sensor. For instance, some participants were also conducting a separate demo under the LiDAR during the data collection, and it was recorded. For evaluation, each recording was saved as a ROS bag and later used to extract point cloud samples for model inference. Chapter 4 elaborates further on the evaluation strategy.

Figure 5: Close up view of the overhead LiDAR and the crane.

4. Impact & Performance Evaluation (Analysis)

4.1 Key Success Metrics

To measure the impact of the project work, we use key success metrics to evaluate the performance of the person detection model. We specifically report two views of prediction performance: (1) a human-annotated (labelled) test set to measure correctness and accuracy, and (2) an unlabeled set using test-time augmentation (TTA) to measure model stability (how consistent detections remain under small input changes). Note that the TTA–based evaluation is used to judge how consistent predictions are under perturbations (such as tiny rotations, noise, etc.).

Outlined below are the leveraged metrics under both these evaluation schemes

With ground-truth labels (measure of accuracy):

i. **Precision**: share of predicted people that are correct. High precision means few false alarms.
ii. **Recall**: share of actual people that are successfully detected. High recall means few misses.
iii. **F1 score**: a single number which balances both precision and recall
iv. **Average precision (AP)**: area under the precision-recall curve across score thresholds. This captures performance across operating points.
v. **Mean intersection-over-union (mIoU)**: average birds eye view (BEV) overlap between the predicted and ground-truth boxes. This metric highlights the accuracy of the localization.
vi. **Distance-based slicing**: all the above metrics are reported within cumulative radii from the center of the LiDAR to reveal how performance changes with the range.

Without ground-truth labels (measure of stability):

i. **Presence rate (mean)**: fraction of TTA runs in which a detection reappears. Higher the value, the more consistent detections are.
ii. **IoU to reference (mean)**: overlap between boxes from perturbed runs and the no-augmentation baseline. Higher the value, the steadier the box placements are.
iii. **Center L2 std (m)**: standard deviation of box centers. Lower the value, the less positional jitter there is.
iv. **Yaw std (°)**: orientation stability. Lower the value, the steadier the heading.
v. **Confidence std**: variation of detection scores. Lower the value, the steadier the confidence.

4.2 Results

Figures 6 and 7 below present samples of successful person detections visualized in Open3D and RViz respectively.

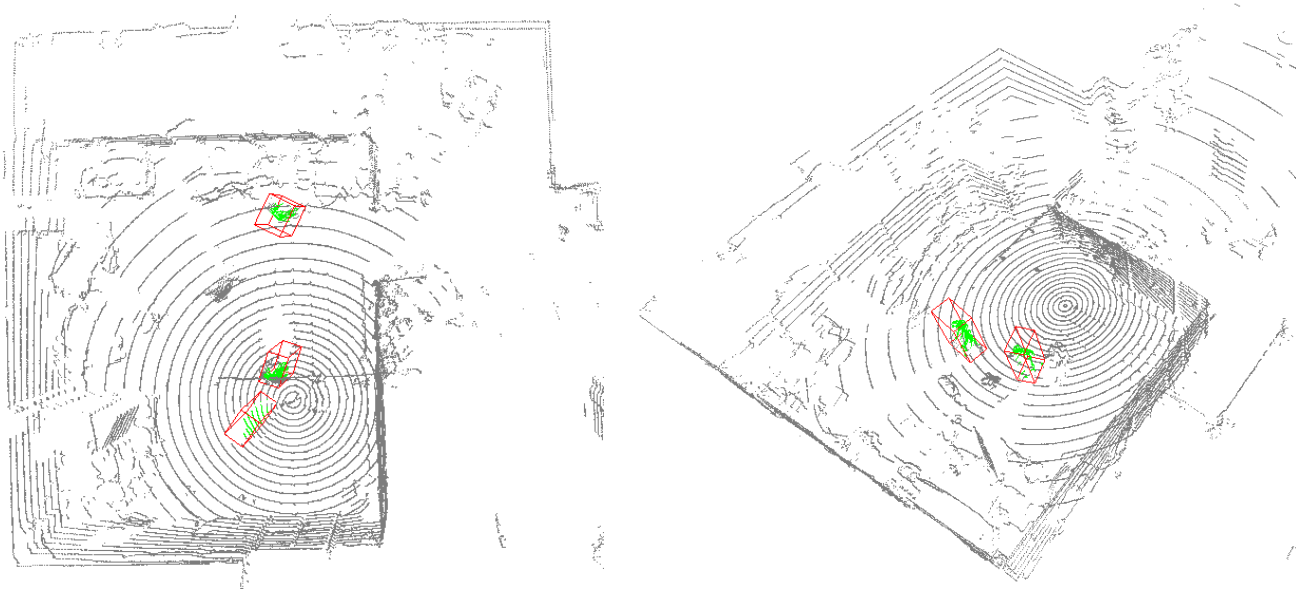


Figure 6: Two successful person detectionl results, each from a distinct viewpoint, visualized in Open3D.

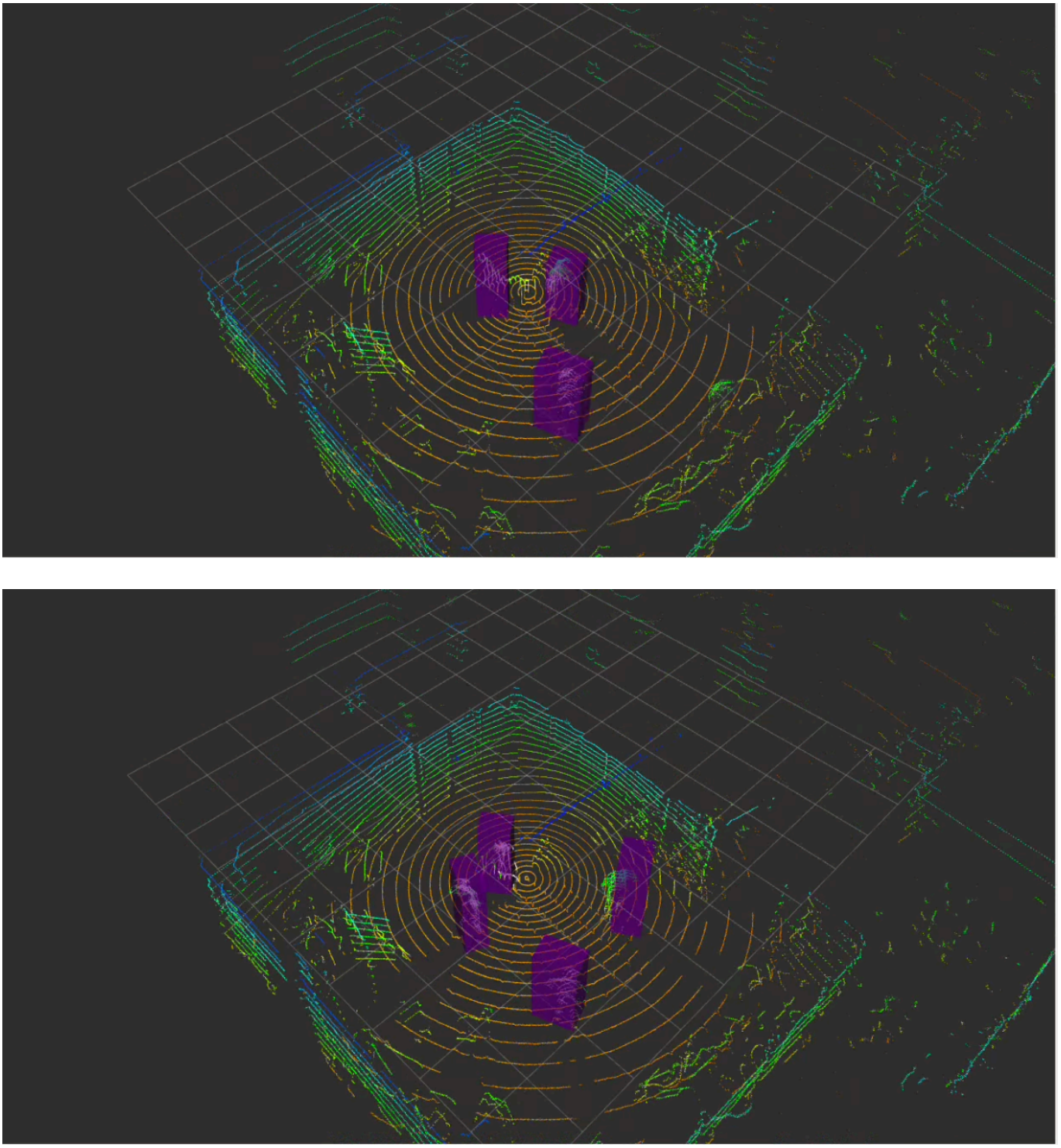


Figure 7: Two successful person detection results visualized in RViz.

4.2.1 Evaluation with ground truth labels

The trained detection model was evaluated on 76 LiDAR frames with human-annotated labels. Thus, we used the human-annotated ground-truth labels to compare the predictions and estimate how accurate the model's predictions are.

Conventional evaluation metrics for detection were computed. As shown in Table 1 below, we evaluated the metrics for various radii. For instance, in the row of values for the Distance of 0.5 m, we have filtered out all points

which are more than 0.5 m away from the center of the LiDAR. Figure 8 is a graphical representation of these values that depict the trends of how each metric changes as you move further from the LiDAR.

Table 1: Quantitative metrics for evaluation of model accuracy; with ground-truth labels

Distance	Metrics				
	Precision	Recall	AP	F1	mIoU
0.50 m	0.889	0.800	0.800	0.842	0.464
1.00 m	0.955	0.724	0.713	0.824	0.472
1.50 m	0.895	0.756	0.723	0.819	0.426
2.00 m	0.929	0.600	0.566	0.729	0.426
2.50 m	0.952	0.641	0.624	0.766	0.444
3.00 m	0.904	0.573	0.545	0.701	0.451

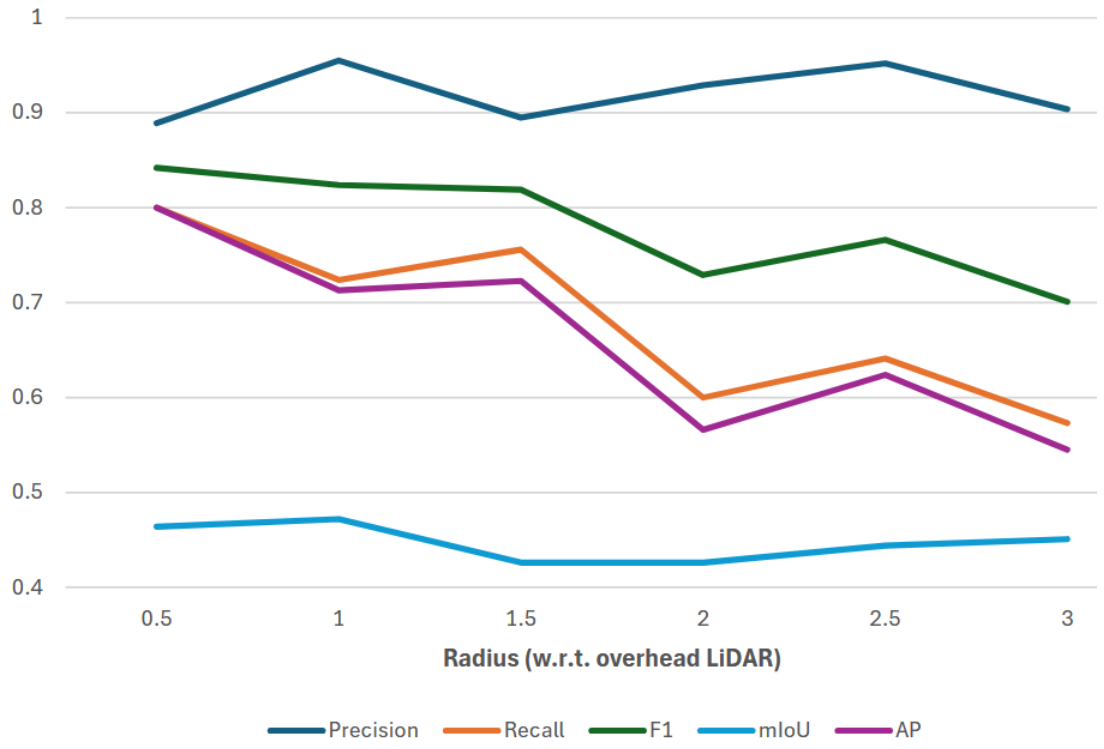
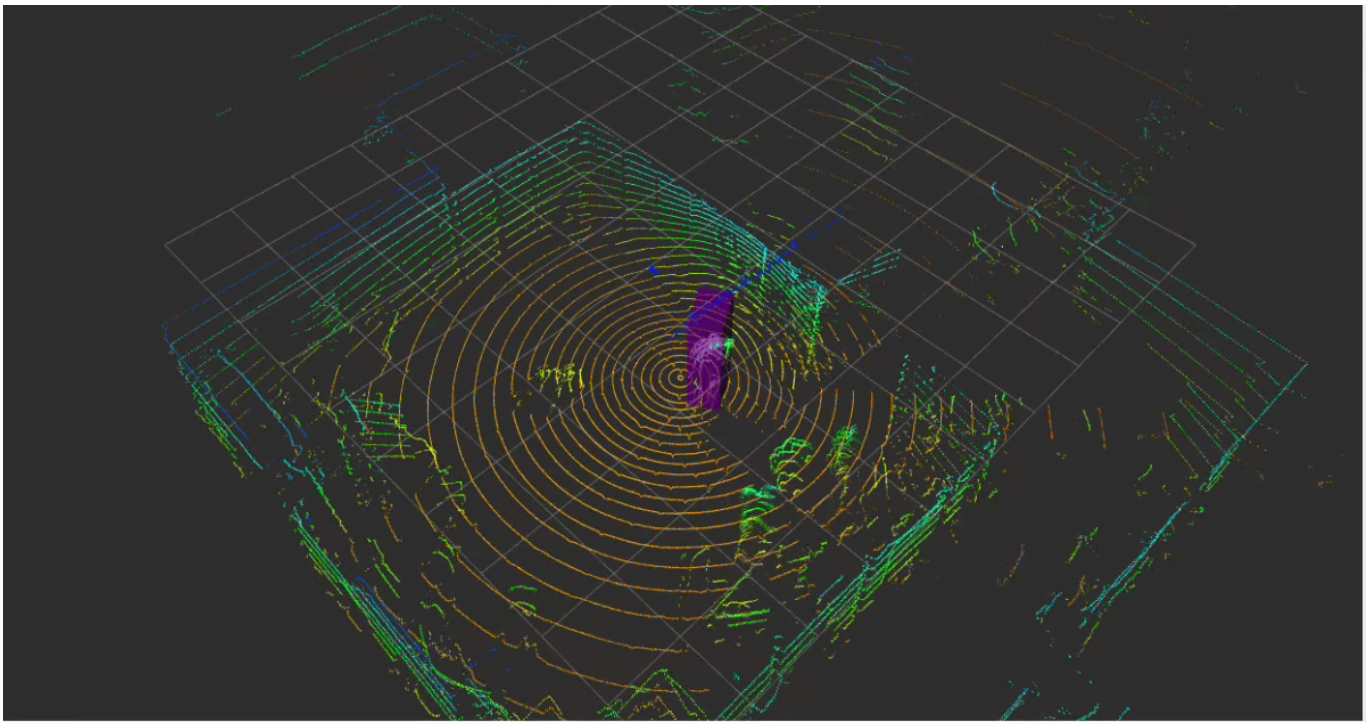


Figure 8: Metric distribution based on distance from the LiDAR

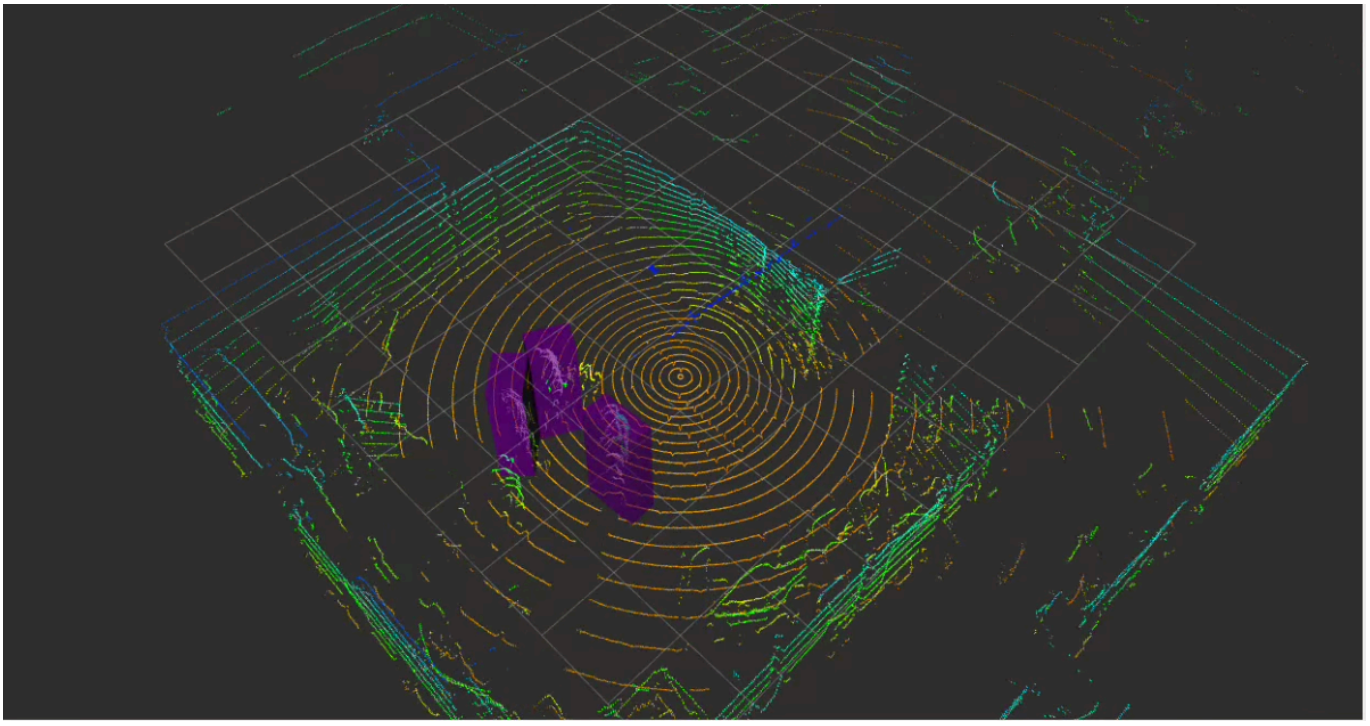
Below are some key interpretations based on the above table of results and corresponding plot:

- High precision across all distances indicates that the detector rarely flags non-person objects as persons (hence less false positives), which is good for avoiding false alarms around the crane.
- Recall decreases with distance, especially beyond the 1.5 m mark. This proves that there are more misses farther from the sensor, which is expected due to sparser points. In practice, this means people close to the LiDAR are detected more reliably than those near the edge of the workspace.
- F1 and AP metrics follow the same trend, high at close range and steadily decreasing toward 3.0 m. This suggests a practical operating envelope where automation or detection alerts can be most dependable.
- mIoU remains relatively steady, implying that when a person is detected, the box localization quality is reasonably consistent across distances.

- The overall takeaway is that in close range to the LiDAR (up to 1-1.5m range), the model provides reliable, low-false-alarm detections. The examples in Figure 9 provides instances of undetected persons (circled in red).



(a) Only one person has been successfully detected due to the remaining three being too far from the LiDAR



(b) One of the four people has not been detected due to sparse points as the range increases

Figure 9: Examples of limitations with the current approach

4.2.2 Evaluation with test time augmentation (TTA)

This analysis evaluates the stability and consistency of the trained PointPillars-based LiDAR person detection model on a total of 94 LiDAR frames without leveraging ground truth labels. As an alternative to human-annotated ground-truth labels, a test-time augmentation (TTA)–based stability analysis was employed to estimate how consistent the model's predictions are under small input perturbations. Its approach is as follows:

1. The same trained model was run multiple times on the same test frames.
2. Each run applied small random perturbations (small changes/disturbances) to the input point clouds (outlined in Table 2).
3. The resulting bounding boxes were compared across runs to quantify the model's stability.

A total of 20 runs (repeated inferences) with test-time augmentation (TTA) were performed, with the first run without any perturbation.

This evaluation approach provides an indirect but meaningful measure of how invariant the model is to small environmental or sensor variations—something particularly relevant to real-world deployment.

Table 2: Augmentation parameters leveraged during TTA

Parameter	Description	Value
Rotation	Randomly rotates the entire point cloud slightly to mimic small orientation changes in the sensor or subject movement.	±10°
Horizontal translation	Applies small shifts along the ground plane to simulate minor position offsets in the LiDAR or person's location.	0.05 m (std)
Vertical translation	Adds small up/down shifts to represent minor height variations or calibration noise.	0.03 m (std)
Scale variation	Uniformly scales the entire point cloud to test sensitivity to small scaling differences.	0.0075 m (std)
Per-point jitter	Add random (gaussian) noise to simulate sensor measurement uncertainty.	0.005 m (std)

Table 3 below presents a summary of the selected metrics, their recorded values, and corresponding interpretations.

Table 3: Quantitative metrics for evaluation of model stability; without ground-truth labels

Metric	What it measures	Value	Interpretation
Presence rate (mean)	Fraction of runs in which a detection consistently appears	0.834	83% of detections persist across perturbations: high consistency
IoU mean to reference mean	Average overlap between perturbed and baseline boxes	0.481	Moderate spatial alignment under noise
Center L2 std	Average positional variability of box centers	0.157m	Small (~15 cm) variation: stable localization
Confidence std	Variation in detection confidence across runs	0.0006	Extremely low variation: stable classification confidence
Yaw std	Variation in predicted orientation	4.6°	Minimal change in orientation estimates: robust directional consistency

Overall, these results indicate that the detector is robust to small, realistic perturbations, and detections largely reappear and stay close to their baseline location and orientation. This highlights that the model's outputs are stable enough to be used for downstream applications (such as tracking or fusion, and safety logic) with minimal temporal flicker (detections persist and stay nearly in place).

4.3 Challenges & Solutions

Finding an algorithm for detecting the positions of people from an overhead point cloud view was challenging. Objects, such as humans, inside pointclouds can be difficult to interpret visually and therefore machine learning is required. In addition to the difficult detection problem the position of the crane lidar proposes another issue with the data. Most available datasets for human detection from point clouds are from the "line of sight" perspective (robot in the same plane as a human). However, with the overhead position of the lidar there is a clear lack of available datasets that can be used to train the machine learning model. This problem was addressed by labelling a dataset tailored to the current environment.

Another issue with training a model with data is the risk of overfitting. If the model parameters are entirely fitted to a small dataset, it can make it difficult for the model to generalize to changes in the environment. Although it is infeasible to create a large dataset for this project, we utilized existing pretrained models that have general knowledge about object geometry and finetuned them to this purpose. During the finetuning process, we started with the pretrained weights from the OpenPCDet repository and improved them with our own dataset with a small learning rate for a few epochs. This created a model that can detect people from the overhead position, while reducing the risk of overfitting and undesirable behaviour.

5. Conclusion & Future Work

5.1 Key Takeaways

We delivered a working pipeline that detects people from an overhead 3D LiDAR mounted on a crane and provides a visualization of the human bounding boxes. It runs on Jetson Orin NX, publishes ROS 2 topics, and avoids camera imagery, which helps mitigate privacy concerns. Furthermore, the fine-tuned PointPillars model shows the expected training and validation loss behavior, indicating good generalization rather than overfitting. We also implemented a ground-truth evaluation and a test-time robustness check (TTA without labels).

Results were also analyzed by distance from the LiDAR so that stakeholders can see how performance changes across the working radius with this single overhead LiDAR configuration. A key operation-based learning is that although overhead views reduce occlusions relative to side-looking sensors, it still faces challenges such as sparse returns at long range which results in undetected instances. Yet, the findings of the project demonstrated feasibility of person detection with an overhead LiDAR.

5.2 Scalability & Potential for Expansion

The approach can be generalized across similar sites and industries (such as warehouses, gantry cranes, and factory cells) that can mount a 3D lidar in the overhead position. The finetuned model can be used directly, however, it is recommended to create a dataset tailored specifically to a new environment for better accuracy. The integration will be modular as ROS 2 topics enable drop-in connections to elements such as safety monitors and fleet dashboards. Moreover, our choice of both the hardware and software stacks creates a strong foundation for the inference system to be deployed for real-time predictions.

Multiple paths for potential commercialization were also identified for the future. First, it would be possible to put together and offer a complete edge kit which would include the Jetson computer, the LiDAR sensor, and our software, such that the kit can be delivered and commissioned on site to work out of the box. Then, for sites which may already have a suitable LiDAR installed, it could be possible to provide the software-only as a license or cloud-managed service, keeping costs lower and deployment faster. Finally, the system can also be packaged as a

safety add-on that can connect directly to crane controllers to trigger presence or zone alerts with respect to detected people instances, thereby making it a simple upgrade to existing control systems

5.3 Recommendations for Further Investment

We recommend investing in a comprehensive data program. Specifically, the labelled dataset can be expanded to include additional clutter in the working area, other objects which are human-sized, corner cases across distance bins, frequent occlusions from crane loads and structure, and adverse conditions such as dust. Alongside this, a simple improvement loop can be set up to regularly collect short recordings from the crane, pick and label such tricky scenarios, retrain the detector with those examples, and run the same quality/performance evaluation for the model. Versions which pass this check can be then installed.

Multi-sensor integration can also be considered to increase coverage and reduce blind spots. For instance, a second overhead LiDAR can be added to see around occlusions and into corners. With such a second viewpoint, the system is less likely to miss a person who might, for instance, be briefly hidden by a lifted load or a structure. Together, better data and better coverage will increase robustness, especially reducing missed detections and even false alarms.