

LiDAR Point Cloud Feature Learning with One-Class SVM for Efficient Human Recognition in Indoor Environment

Hadil Eltaif

Département d'informatique
Université de Sherbrooke
Sherbrooke, QC, Canada.
hadil.eltaif@usherbrooke.ca

Charles Gouin-Vallerand

Département des systèmes d'information
et méthodes quantitatives de gestion
Université de Sherbrooke
Sherbrooke, QC, Canada.
charles.vallerand@usherbrooke.ca

Yacine Yaddaden

Département de mathématiques,
informatique et génie
Université du Québec à Rimouski
Rimouski, Québec, Canada.
yacine_yaddaden@uqar.ca

Abstract—Accurate person detection is a key enabler for intelligent monitoring in smart home environments, supporting applications such as assisted living, activity recognition, and adaptive automation. 3D LiDAR sensors are particularly well suited for this task, as they provide precise geometric information while preserving privacy. However, robust indoor detection remains challenging due to cluttered interiors, occlusions, diverse human activities, and the limited availability of annotated datasets. We propose a novel LiDAR-based person detection approach that leverages cross-dataset transfer between deep feature learning and One-Class classification. Deep geometric features are first learned from the L-CAS 3D People Dataset and transferred to the SZTAKI LiDAR Gait-and-Activity dataset, where they are exploited by a lightweight One-Class SVM trained only on the “person” class. This enables the detection of non-person anomalies without requiring extensive negative samples. We report multiclass classification results on the L-CAS dataset, achieving over 98% accuracy. On the more challenging SZTAKI-LGA dataset, the proposed method reaches up to 96% accuracy and 96% F1-score. Finally, the full pipeline is validated on a small in-house dataset collected in our laboratory, highlighting its practical feasibility in a realistic indoor setting and providing a foundation for future extensions toward multi-person tracking and advanced smart-environment applications.

Index Terms—3D LiDAR, Person Detection, Deep Learning, One-Class classification, Smart Indoor Environments

I. INTRODUCTION

With the rapid development of **smart indoor environments**, accurate and privacy-preserving **person detection** has become essential for monitoring, activity recognition, and context-aware services [1]. While RGB/RGB-D systems offer rich semantic cues, they remain sensitive to lighting conditions and raise privacy concerns. Similarly, 2D LiDAR enables real-time sensing but lacks volumetric information, limiting its reliability in cluttered or crowded indoor scenes [2].

In contrast, **3D LiDAR sensors** provide dense geometry-only point clouds, offering robustness to illumination changes and inherently preserving user privacy [3]. These properties make them attractive for indoor monitoring.

However, most LiDAR-based person detection research has focused on outdoor autonomous driving using large supervised

datasets such as KITTI [4] or Waymo [5]. Indoor environments introduce different challenges: close-range scanning often yields partial human shapes; furniture and walls create strong occlusions; and varying layouts produce a significant *domain shift*. Furthermore, annotated indoor LiDAR datasets remain scarce, typically containing few subjects and offering limited negative examples.

These constraints motivate the need for models that can **learn from limited supervision** and still generalize across heterogeneous indoor scenes. To address this, we explore a hybrid approach combining **deep geometric feature learning** with **One-Class classification**. Deep models are first trained on the L-CAS dataset [6] to learn transferable geometric representations. The resulting embeddings are then used for person-versus-non-person detection through a lightweight **OC-SVM** [7] trained only on person samples. The multiclass evaluation on L-CAS is therefore used as an intermediate stage to assess and compare the quality of the learned representations, rather than as the final task addressed in this work.

We evaluate this pipeline through cross-environment experiments on both the **SZTAKI-LGA** dataset [8] and a new **laboratory dataset** collected with a different LiDAR sensor. Unlike prior works relying on large supervised datasets or multisensor fusion, this study focuses on the underexplored yet practical setting of **positive-only indoor LiDAR detection**.

This paper makes the following contributions:

- A **hybrid detection framework** combining deep LiDAR feature learning with One-Class classification for privacy-aware indoor monitoring.
- A lightweight **OC-SVM** trained solely on person embeddings, reducing annotation effort while enabling cross-environment generalization.
- **Cross-dataset validation** on L-CAS, SZTAKI-LGA, and a new laboratory dataset, demonstrating robustness across different indoor environments and sensor configurations.
- A **preliminary in-lab dataset** collected in a smart-home environment, enabling end-to-end experiments to validate the proposed approach.

The remainder of the paper is organized as follows. In Section II, we present the fundamentals of LiDAR-based person monitoring and review related works. Section III provides a detailed description of the proposed detection approach. Section IV details the used datasets. In Section V, we define the experimental protocol and introduce the benchmark datasets used in this study. Section VI presents and discusses the obtained results. Finally, Section VII concludes the paper and outlines future research directions.

II. BACKGROUND

A. Fundamentals

In indoor environments, LiDAR-based person monitoring involves several complementary tasks that collectively enable understanding human presence and movement within a scene. The three core tasks are [9]:

- 1) **Detection of Regions of Interest (ROI)** — identifying candidate areas and their 3D locations within the LiDAR point cloud. The system outputs point clusters or bounding boxes corresponding to potential individuals.
- 2) **Classification** — determining whether each detected cluster truly corresponds to a person or another object in the environment (e.g., *chair*, *table*). While detection answers *where* a person is, classification answers *what* the detected cluster represents.
- 3) **Tracking** — maintaining consistent identities for detected persons across consecutive LiDAR frames. Tracking assigns unique IDs and estimates trajectories, enabling continuous monitoring of individuals over time.

Together, these tasks form the foundation of LiDAR-based person monitoring. Detection provides spatial localization, classification adds semantic understanding, and tracking ensures temporal consistency, enabling higher-level applications for instance: occupancy analysis, safety monitoring, and human–robot collaboration. Moreover, these steps are essential prerequisites for reliable Human Activity Recognition (**HAR**).

B. Related Work

Person detection in indoor environments has been approached using various sensing modalities, including RGB-D cameras, 2D LiDAR, 3D LiDAR, and more recently mmWave radar. RGB-D systems provide rich information [10], [11], but their limited range, field of view, lighting sensitivity, and privacy constraints reduce their suitability for smart-home applications.

2D LiDAR sensors are popular due to their low cost and real-time performance [12], [13]. Most methods cluster 2D scans and classify them using handcrafted geometric descriptors, but the lack of height information limits their ability to separate people from nearby objects in cluttered indoor scenes.

To address these limitations, **3D LiDAR sensors** offer denser point clouds with richer geometric structure. Early approaches relied on handcrafted descriptors, such as the geometry-based pipeline of Navarro-Serment et al. [14] or the height–density features of Kidono et al. [15]. Later works

introduced adaptive strategies: Yan et al. [16]–[18] proposed online learning and multisensor transfer on the **L-CAS** dataset.

Deep architectures such as PointNet and PointNet++ [19], [20] further advanced 3D feature learning. However, these models are **data-intensive** and exhibit **domain shift** when applied to unseen indoor spaces or different sensors, which limits their reliability for practical indoor detection.

mmWave radar has also gained interest for indoor sensing due to its robustness to occlusions and low cost. Yet, mmWave point clouds are **sparse and noisy**, with lower spatial resolution than LiDAR [21], while 3D LiDAR offers significantly **denser geometry and higher angular resolution** [22]—properties crucial for shape-based discrimination and transferable embeddings.

Nevertheless, **indoor LiDAR-based person detection remains underexplored**, with most existing work targeting autonomous driving or general 3D scene understanding. Few studies explicitly address cross-environment generalization under limited supervision, and only a handful of indoor datasets such as **L-CAS** [6] and **SZTAKI-LGA** [8] support this line of research.

III. PROPOSED SYSTEM

The proposed LiDAR-based person detection framework follows a generic two-stage design (*offline* and *online*), as illustrated in Fig. 1. This structure is independent of any specific dataset and can be instantiated with different training sources. In the **offline phase**, a deep feature extractor is trained to learn discriminative geometric representations from annotated LiDAR clusters. These learned embeddings are then used to train a lightweight **One-Class SVM (OC-SVM)** that models the distribution of person samples without requiring labeled negatives. In the **online phase**, incoming LiDAR clusters are processed in real time: the feature extractor generates embeddings, which are then classified by the OC-SVM as *person* or *non-person*.

In our implementation, the feature extractor is trained on the L-CAS dataset, while the OC-SVM is trained and evaluated on SZTAKI-LGA and our laboratory data.

A. Pre-processing

To ensure accurate feature learning and robust generalization, both the **L-CAS** [6] and **SZTAKI-LGA** [8] datasets undergo a standardized pre-processing pipeline, which includes the following stages:

1) **Point-Cloud Normalization**: Each input cluster P is translated to its centroid and scaled to fit inside a unit sphere. This step ensures **translation** and **scale invariance**, aligning clusters from different environments to a common reference frame.

2) **Fixed-Size Resampling**: LiDAR clusters naturally vary in size because of occlusions, sensor noise, and distance from the sensor. However, neural networks require inputs of uniform size. To this end, each cluster is resampled to a fixed number of points, denoted by N . Following standard practice in point cloud learning such as in [19], [20], we evaluate three

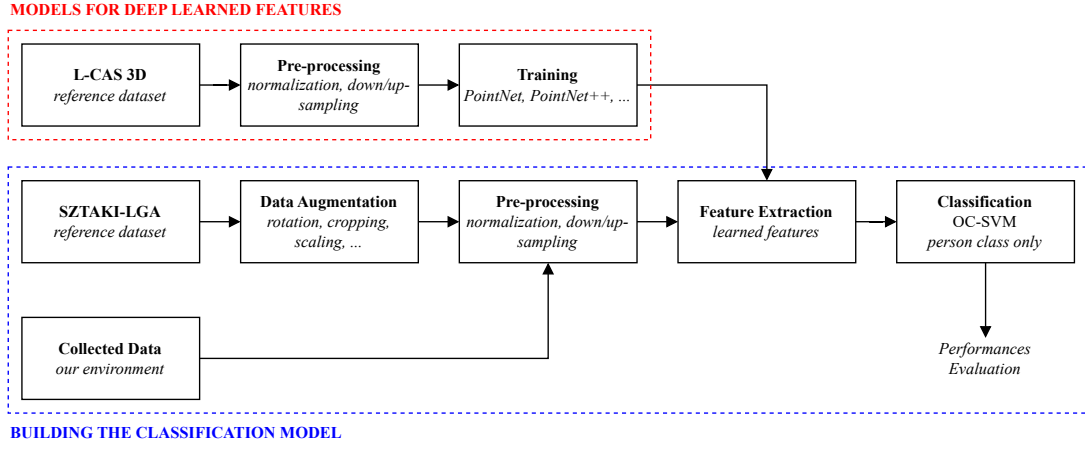


Fig. 1. Proposed LiDAR-based person detection framework, composed of an **offline phase** for feature learning and model training, and an **online phase** for classification on our laboratory LiDAR streams.

commonly used sizes $N \in \{256, 512, 1024\}$. Here, N is not a threshold but the target number of points after resampling. For clusters with more than N points, we apply **Farthest Point Sampling (FPS)** to select a representative subset of keypoints that preserves both local and global geometry. FPS iteratively picks points that are maximally distant from each other, ensuring uniform spatial coverage of the cluster shape [20].

For clusters with fewer than N points, we apply a simple **progressive upsampling** strategy, where keypoints are duplicated in a balanced manner and slightly jittered. This avoids exact point repetition while maintaining the structural consistency of the cluster, without relying on any learned upsampling model.

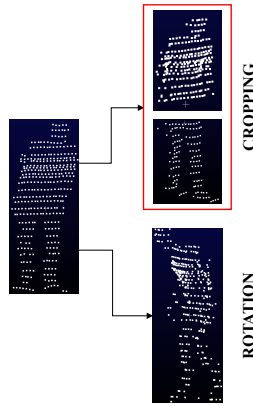


Fig. 2. Illustration of SZTAKI-LGA data augmentation. **(Top)** Body segmentation into upper and lower parts. **(Bottom)** Discrete yaw rotations to simulate viewpoint variations.

3) **Data Augmentation (SZTAKI-LGA only)**: The **SZTAKI-LGA dataset** contains few labeled samples (3,416), so we apply simple augmentations to increase variability. As shown in Fig. 2, we split each cluster into upper and lower body parts to simulate occlusions, and we apply small yaw rotations to

vary the viewpoint. Although a real LiDAR would capture slightly different points at different orientations, small rigid rotations are a standard augmentation in point-cloud learning [19], [20] and help the model learn more robust geometric features. Here, rotations are used only to improve invariance, not to replicate exact LiDAR measurements.

B. Feature Extraction

After pre-processing, we extract discriminative geometric features from 3D LiDAR clusters using deep learning architectures trained on the annotated **L-CAS dataset**. The task is formulated as a **three-class classification problem** with labels: *pedestrian*, *group*, and *non-person*.

1) **Deep Learning Architectures**: We evaluate three representative architectures designed for point cloud data:

- 1) **PointNet** [19]: A baseline model that applies shared Multi-Layer Perceptrons (MLPs) on individual points and aggregates features using symmetric max-pooling.
- 2) **PointNet++** [20]: Extends PointNet with hierarchical feature learning, capturing both local and global structures. Two variants are tested: (i) Single-Scale Grouping (**SSG**) for compact local features, (ii) Multi-Scale Grouping (**MSG**) for richer context aggregation.
- 3) **PointNet with Self-Attention** [23]: Integrates a lightweight attention module to adaptively emphasize the most informative regions within the point cloud.

PointNet [19] and PointNet++ [20] are widely used architectures in point-cloud learning and are used here as reference feature extractors.

2) **Training Objective**: All models are trained in a supervised classification setup over three classes. Clusters are uniformly resampled to fixed sizes $N \in \{256, 512, 1024\}$ (see Sec. III-A). After convergence, the classification layers are discarded, and the **feature encoders** are kept frozen.

3) **Feature Encoder Extraction**: The frozen encoders map each input cluster into a fixed-length **embedding vector** $\mathbf{x} \in \mathbb{R}^{1024}$, capturing both local and global geometric patterns,

where the dimensionality of the vector is set to 1024. These embeddings serve as inputs for the downstream OC-SVM classifier.

C. OC-SVM Classification

The One-Class Support Vector Machine (**OC-SVM**) [7] is an unsupervised anomaly detection method that learns a boundary enclosing *normal* data points (**inliers**) while rejecting unseen or abnormal ones (**outliers**). In our case, this boundary models the distribution of person clusters, while all other objects are treated as outliers. Unlike multi-class SVMs, the OC-SVM focuses solely on the target class, making it suitable for indoor settings where non-person objects are highly diverse. Its effectiveness in positive-only scenarios has also been demonstrated in biometric modalities such as iris and palmprint authentication [24].

As shown in Fig. 1, deep embeddings extracted from the frozen feature encoder are used as inputs to the OC-SVM. During training, the model estimates a compact region—often interpreted as a hypersphere in the feature space—that encloses the person embeddings. At test time, incoming LiDAR clusters are classified as *person* if they fall inside this region, and *non-person* otherwise.

Formally, let $X = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ denote the set of person embeddings. The OC-SVM decision function is defined as:

$$f(\mathbf{x}) = \text{sign}(\langle \mathbf{w}, \phi(\mathbf{x}) \rangle + b), \quad (1)$$

where $\phi(\cdot)$ is a non-linear mapping into a high-dimensional kernel space, $\mathbf{w}$ defines the separating hyperplane, and b is the bias term. The parameters are obtained by solving:

$$\min_{\mathbf{w}, \xi, \rho} \frac{1}{2} \|\mathbf{w}\|^2 + \frac{1}{\nu N} \sum_{i=1}^N \xi_i - \rho, \quad (2)$$

where ξ_i are slack variables allowing soft-margin violations, ρ is the decision threshold, and $\nu \in (0, 1]$ controls the maximum fraction of outliers permitted during training.

D. Online Inference

In the **Online phase** (e.g. with our laboratory smart home), the trained framework is deployed on online LiDAR streams from the Velodyne VLP-32C sensor. To detect dynamic objects, we perform background subtraction by selecting one frame as a reference and comparing each new frame using a fast KD-Tree search. Points missing from the reference are considered dynamic, which are then grouped into clusters using Euclidean clustering. In the visualization, the static background is shown in gray, while the dynamic clusters are highlighted in distinct colors, as illustrated in Fig. 3.

For each dynamic cluster, the pipeline performs:

- 1) **Pre-processing**: normalization and fixed-size resampling.
- 2) **Feature Extraction**: obtaining a 1024-D embedding using the frozen deep encoder.

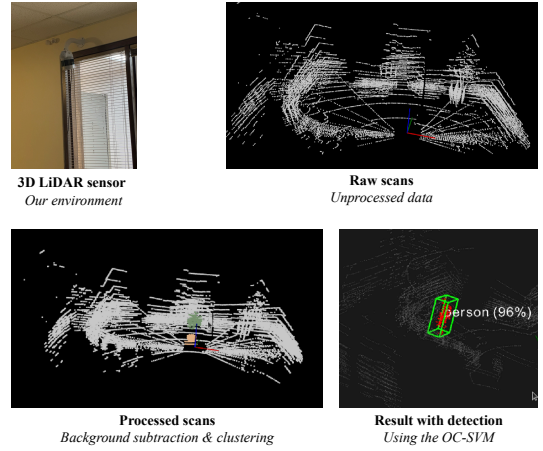


Fig. 3. Overview of the LiDAR-based detection pipeline in the experimental environment: (top left) sensor setup, (top right) raw scans, (bottom left) background subtraction and clustering, and (bottom right) final detection using the OC-SVM classifier.

- 3) **Classification**: testing the embedding on the trained **OC-SVM** to decide whether the object is a *person* or *non-person*.

This streamlined design allows for efficient and accurate person detection in our laboratory environment, while keeping computational overhead low.

IV. DATA DESCRIPTION

We base our experiments on two complementary benchmark datasets: the **L-CAS Dataset** [6] and the **SZTAKI-LGA Dataset** [8]. The **L-CAS** dataset is employed for supervised multi-class classification and for training deep models to learn transferable feature representations. The **SZTAKI-LGA** dataset is then used for cross-dataset evaluation and for training the **OC-SVM** classifier.

A. L-CAS 3D Point Cloud People Dataset

The **L-CAS 3D Point Cloud People Dataset** was collected at the University of Lincoln (UK) using a Velodyne VLP-16 LiDAR. It contains **28,002 indoor frames**, of which **5,492** are manually annotated, resulting in **9,122 labeled instances** (6,068 single-person and 3,054 group instances).

Annotations are provided as per-frame files describing **3D point clusters** and bounding boxes. Key statistics are summarized in Table I.

TABLE I
L-CAS 3D POINT CLOUD PEOPLE DATASET: KEY STATISTICS

PROPERTY	VALUE
Acquisition Site	University of Lincoln (UK)
Sensor	Velodyne VLP-16
Total Frames	28,002
Annotated Frames	5,492
Labeled Instances	9,122 (6068 single-person, 3,054 group of people)
Classes	Pedestrian, Group of people

B. SZTAKI-LGA Activity Recognition Dataset

The **SZTAKI LiDAR Gait-and-Activity (LGA) Dataset** [8] was collected using a **Velodyne HDL-32E** 3D LiDAR sensor and contains high-quality point clouds of people performing various actions. The activity subset used in this work includes **3,416 original PCD samples** covering five categories: bend to ground, check watch, phone call, wave, and two-handed wave. After data augmentation, the dataset contains **4,343 samples** in total. Key statistics are summarized in Table II.

TABLE II
SZTAKI-LGA ACTIVITY RECOGNITION DATASET: KEY STATISTICS

ACTIVITY	NUMBER OF PCD SAMPLES
Bend to Ground	1,018
Check Watch	552
Phone Call	684
Wave	731
Two-Handed Wave	431
Total (original)	3,416
Total (after augmentation)	4,343
Increase (%)	$\approx 30\%$

C. Laboratory Dataset

To complement the evaluations on public benchmarks, we also conducted preliminary tests using a small dataset collected in the smart apartment of the DOMUS Laboratory at the Université de Sherbrooke (see Table III). This dataset was acquired with a **Velodyne VLP-32C** sensor and includes **1,790 point-cloud samples** recorded from four participants in realistic indoor conditions. The scenes feature typical smart-home layouts with furniture-induced occlusions, allowing us to assess the behavior of the proposed approach in more complex environments. Its primary purpose is to validate the practical feasibility of our pipeline, while a more comprehensive and systematically structured in-lab dataset is currently being collected to support large-scale evaluation.

TABLE III
LABORATORY DATASET: KEY STATISTICS

PROPERTY	VALUE
Acquisition Site	Indoor laboratory environment (Domus Lab)
Sensor	Velodyne VLP-32C
Total Samples	1,790
Participants	4 different persons
Classes	Person

V. EVALUATION PROTOCOL

We evaluate our approach through three complementary protocols: **(i)** multi-class classification on L-CAS to assess feature learning and compare with existing methods, **(ii)** one-class evaluation on SZTAKI-LGA, which forms the core of our LiDAR-based person detection task, and **(iii)** validation on laboratory data acquired with a VLP-32C to assess deployment robustness. Across all protocols, we report **Accuracy, Precision, Recall, F1-score**, and **confusion matrices**. Protocol 1 does not correspond to the final target setting of the paper; it

is used only to assess the quality of the learned embeddings on L-CAS. Protocols 2 and 3 address the actual task of person-versus-non-person detection using a one-class classifier trained on person data only.

A. Protocol 1: Multi-class Evaluation on L-CAS

This protocol is used as an intermediate feature-learning evaluation stage rather than as the final recognition task addressed by the paper. The **L-CAS** dataset is split into **70%** training, **15%** validation, and **15%** testing, with three classes: *person*, *group*, and *non-person objects*. To enrich the non-person class, we add **2,000 samples** (*chair*, *table*, *sofa* and *background elements*) from the **S3DIS** [25], also reused later as outliers in the one-class evaluation.

B. Protocol 2: One-class Evaluation on SZTAKI-LGA

The **SZTAKI-LGA** dataset provides balanced *person* and *non-person* clusters. In this protocol, an OC-SVM is trained on **70% of the person clusters**, while the remaining 30% (mixed with non-person samples) are split into a **validation set** (15%) and a **test set** (15%). This setup evaluates the model's ability to accept genuine person clusters while rejecting impostors.

C. Protocol 3: Validation on Laboratory Data

To assess deployment robustness, we replicate the one-class protocol on our **in-house laboratory dataset** acquired with a Velodyne VLP-32C. An OC-SVM is retrained on **60%** of the *person* embeddings; the remaining **40%** are mixed with the standardized *impostor* set (as in Protocol 2) and split equally into **validation** (20%) and **test** (20%). This setup evaluates recognition on newly acquired LiDAR data without relying on pre-recorded benchmarks.

The OC-SVM is trained only on positive samples; the validation split is used strictly to tune ν and the decision threshold. This is necessary to cope with indoor domain shift and ensures an unbiased one-class evaluation.

VI. RESULTS & DISCUSSION

We present results on the L-CAS, SZTAKI-LGA, and laboratory datasets, covering multi-class classification, one-class evaluation, and real-world validation. The analysis examines the impact of network architecture and point resolution, provides class-wise performance, and compares our approach with existing methods.

A. Evaluation on the L-CAS Dataset

Table IV summarizes results for PointNet and PointNet++ variants at three point resolutions (256, 512, 1024). Metrics include Accuracy, macro-Precision, macro-Recall, and macro-F1.

Since L-CAS is unbalanced, macro-F1 is the most reliable indicator. The best performance is achieved by **PointNet++ MSG at 256 points**, exceeding 98% on all metrics. Its multi-scale grouping captures both local and global features, while using 256 points reduces computation, enabling efficient deployment.

TABLE IV
L-CAS—DATASET PERSON DETECTION PERFORMANCE, BY ARCHITECTURE AND POINT RESOLUTION.

POINT RESOLUTION	ARCHITECTURE	ACCURACY	PRECISION	RECALL	F1-SCORE
processed_256	PointNet	98.20%	98.39%	97.87%	98.12%
	PointNet++ SSG	96.58%	96.98%	96.61%	96.79%
	PointNet++ MSG	98.8%	99%	98.76%	98.88%
	PointNet-Attention	98.14%	98.18%	97.98%	98.07%
processed_512	PointNet	98.32%	98.57%	97.91%	98.23%
	PointNet++ SSG	96.88%	97.28%	97.01%	97.14%
	PointNet++ MSG	98.38%	98.40%	98.61%	98.51%
	PointNet-Attention	97.90%	97.99%	97.54%	97.76%
processed_1024	PointNet	98.56%	98.39%	98.46%	98.42%
	PointNet++ SSG	97.84%	98.27%	97.78%	98.02%
	PointNet++ MSG	98.32%	98.27%	98.61%	98.44%
	PointNet-Attention	98.08%	98.21%	97.83%	98.02%

TABLE V
SZTAKI-LGA DATASET — PERSON DETECTION PERFORMANCE, BY ARCHITECTURE AND POINT RESOLUTION.

POINT RESOLUTION	ARCHITECTURE	ACCURACY	PRECISION (PERSON)	RECALL (PERSON)	F1 (PERSON)
processed_256	PointNet	96%	98%	94%	96%
	PointNet++ SSG	83%	78%	92%	84%
	PointNet++ MSG	86%	87%	84%	85%
	PointNet-Attention	95%	96%	94%	95%
processed_512	PointNet	94%	95%	93%	94%
	PointNet++ SSG	94%	94%	93%	94%
	PointNet++ MSG	92%	91%	93%	92%
	PointNet-Attention	95%	96%	94%	95%
processed_1024	PointNet	95%	94%	97%	95%
	PointNet++ SSG	96%	100%	93%	96%
	PointNet++ MSG	96%	100%	93%	96%
	PointNet-Attention	96%	97%	94%	96%

TABLE VI
OUR LABORATORY DATASET — PERSON DETECTION PERFORMANCE, BY ARCHITECTURE AND POINT RESOLUTION.

POINT RESOLUTION	ARCHITECTURE	ACCURACY	PRECISION (PERSON)	RECALL (PERSON)	F1 (PERSON)
processed_256	PointNet	95%	99%	91%	95%
	PointNet++ SSG	91%	92%	89%	91%
	PointNet++ MSG	97%	99%	94%	97%
	PointNet-Attention	94%	93%	95%	94%
processed_512	PointNet	96%	94%	98%	96%
	PointNet++ SSG	94%	94%	93%	94%
	PointNet++ MSG	98%	97%	98%	98%
	PointNet-Attention	94%	92%	98%	95%
processed_1024	PointNet	95%	94%	97%	95%
	PointNet++ SSG	96%	95%	97%	96%
	PointNet++ MSG	98%	99%	97%	98%
	PointNet-Attention	96%	97%	95%	96%

TABLE VII
COMPARISON OF MULTICLASS CLASSIFICATION RESULTS WITH *state-of-the-art* METHODS FOR 3D LiDAR-BASED PERSON DETECTION ON THE L-CAS DATASET.

METHOD	DESCRIPTION / CLASSIFIER	PRECISION
Yan et al. [16] (2017)	Online learning with P/N experts (SVM classifier)	$\approx 46.2\%$
Yan et al. [17] (2018)	Multisensor transfer learning (RGB-D + 2D LiDAR, SVM classifier)	$\approx 59\%$
Yan et al. [18] (2020)	Adaptive clustering + 10D slice distance feature (SVM classifier)	$\approx 60\%$
Ours	PointNet++ MSG with 256 points (end-to-end deep classification)	99%

The confusion matrix (Table VIII) shows strong class-wise robustness: the *Person* class is recognized with 99.1% accuracy (903/911), the *Group* class with 96.5%, and the *non-person* class with 100%.

TABLE VIII
CONFUSION MATRIX FOR THE MULTI-CLASS CLASSIFICATION TASK ON THE L-CAS DATASET (**POINTNET++ MSG, 256 POINTS**).

Actual		Predicted		
		PERSON	GROUP	NON-PERSON
	PERSON	903	7	1
	GROUP	14	442	2
	NON-PERSON	0	0	300

Table VII reports the few published methods evaluated on the L-CAS dataset. Only a small number of studies have used this benchmark, and all rely on handcrafted geometric descriptors rather than deep point-cloud models. To our knowledge, no recent LiDAR-based indoor detection or modern deep learning approaches report results on L-CAS. This limited set of baselines is mainly due to the low adoption of the dataset, which has been used almost exclusively by its creators, unlike widely used outdoor benchmarks such as KITTI or Waymo. To the best of our knowledge, this work provides one of the few deep-learning evaluations on the L-CAS dataset, with PointNet++ MSG (256 points) reaching **99%** precision.

For the SZTAKI-LGA dataset, no direct state-of-the-art comparison is available for this specific task setting. The dataset was originally designed for *activity recognition*, and we found no prior work addressing a “person vs. non-person” detection task or using SZTAKI-LGA in a one-class setting. This reflects a broader gap in the literature: indoor LiDAR person detection remains underexplored, and existing datasets do not provide standardized baselines. Despite this, our results on SZTAKI-LGA remain consistently high, demonstrating the potential of transferable geometric embeddings for indoor detection.

B. Evaluation on the SZTAKI-LGA Dataset

For SZTAKI-LGA, pretrained networks serve as feature extractors and OC-SVM is used for anomaly detection. Table V reports Accuracy, Precision, Recall, and F1-score for the *Person* class.

The best performance is obtained with **PointNet++ SSG** and **MSG** at 1024 points, both reaching **96% accuracy and F1-score**. These results confirm that higher point density improves feature discrimination.

The confusion matrices (Tables IX and X) highlight a trade-off: - **MSG** achieves 100% specificity, rejecting all impostors but missing 47 persons. - **SSG** recovers more persons but misclassifies two impostors.

Thus, MSG favors anomaly rejection, while SSG offers slightly higher sensitivity—both validating the approach’s robustness.

C. Evaluation on our laboratory Dataset

We next assess generalization on our **laboratory dataset** collected with a Velodyne VLP-32C.

TABLE IX
CONFUSION MATRIX WITH THE PROPOSED SYSTEM EVALUATED ON THE SZTAKI-LGA DATASET (**POINTNET++ MSG, 1024 POINTS**).

Actual		Predicted	
		PERSON	NON-PERSON
	PERSON	604	47
	NON-PERSON	0	652

TABLE X
CONFUSION MATRIX WITH THE PROPOSED SYSTEM EVALUATED ON THE SZTAKI-LGA DATASET (**POINTNET++ SSG, 1024 POINTS**).

Actual		Predicted	
		PERSON	NON-PERSON
	PERSON	605	46
	NON-PERSON	2	650

Applying OC-SVM models trained on SZTAKI-LGA embeddings directly yields lower performance (Fig. 4): **PointNet (1024)** achieves 76.0%, followed by **PointNet (512)** at 71.3%, and **PointNet-Attention (512)** at 60.3%. This gap reflects a strong **domain shift** between SZTAKI-LGA (Velodyne HDL-32E) and our lab setup (Velodyne VLP-32C).

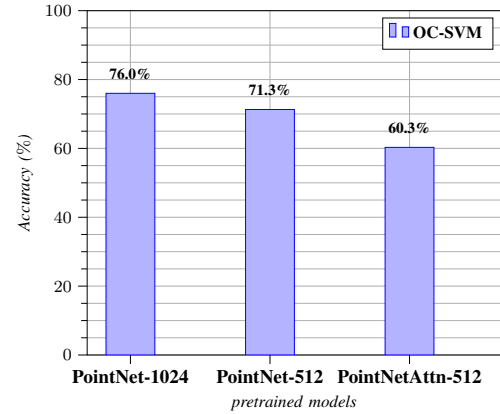


Fig. 4. Performance of the proposed approach when trained on the **SZTAKI-LGA dataset** and subsequently evaluated on our **laboratory dataset**.

Retraining the OC-SVM on embeddings from our own dataset significantly improves results. Table VI shows that **PointNet++ MSG 1024** reaches **98% accuracy**, with 99% precision, 97% recall, and a corresponding **98% F1 score**. As in previous experiments, the dataset was split into **60% training, 20% validation, and 20% testing**. This ensures that the reported performance is obtained on **unseen test data**, and not on samples used during OC-SVM training, confirming the reliability of the evaluation.

The confusion matrix in Table XI confirms this strong performance, showing very few errors across classes.

Overall, these results show that despite limited **experimental LiDAR data**, our two-stage method (deep feature learning + one-class classification) provides robust and reliable detection under practical indoor conditions. On our workstation (Intel Core i7, 64 GB RAM, RTX 4070), preprocessing takes tens of milliseconds, feature extraction a few hundred milliseconds,

TABLE XI
CONFUSION MATRIX OF THE PROPOSED SYSTEM EVALUATED ON THE
LABORATORY DATASET (POINTNET++ MSG, 1024 POINTS).

Actual	Predicted	
	PERSON	NON-PERSON
	PERSON	NON-PERSON
PERSON	346	12
NON-PERSON	3	355

and OC-SVM prediction is nearly instantaneous (0.064 ms). These measurements reflect processing time only and do not include LiDAR acquisition delays, which may slightly increase end-to-end latency. Overall, the latency remains acceptable for indoor monitoring, and further analysis in near real-time conditions is planned.

VII. CONCLUDING REMARKS

This paper presented a LiDAR-based approach for indoor person detection based on deep feature learning and a lightweight OC-SVM classifier. Trained on L-CAS and evaluated across SZTAKI-LGA and in-lab data, the proposed approach showed promising transferability of learned geometric features for person detection across indoor settings.

Experiments on data collected in the **DOMUS Laboratory** further confirmed the practical feasibility of the pipeline under realistic deployment conditions, while also revealing the challenges of direct transfer across different sensors and environments. These results highlight the importance of lightweight adaptation and of larger, more diverse indoor datasets to improve generalization.

Overall, the hybrid combination of deep geometric embeddings and OC-SVM proved effective for indoor person detection with limited supervision. Future work will focus on expanding the in-lab dataset and extending the approach toward multi-person tracking and activity recognition.

ACKNOWLEDGMENT

The authors acknowledge the use of generative artificial intelligence tools to assist in improving the clarity, grammar, and overall readability of the manuscript. The scientific content, methodology, results, and conclusions presented in this work were developed entirely by the authors, who take full responsibility for the integrity and accuracy of the content.

REFERENCES

- [1] A. Chakraborty, M. Islam, F. Shahriyar, S. Islam, H. U. Zaman, and M. Hasan, "Smart home system: a comprehensive review," *Journal of Electrical and Computer Engineering*, vol. 2023, no. 1, p. 7616683, 2023.
- [2] 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, 2023.
- [3] S. Chen, B. Liu, C. Feng, C. Vallespi-Gonzalez, and C. Wellington, "3d point cloud processing and learning for autonomous driving: Impacting map creation, localization, and perception," *IEEE Signal Processing Magazine*, vol. 38, no. 1, pp. 68–86, 2020.
- [4] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset," *The international journal of robotics research*, vol. 32, no. 11, pp. 1231–1237, 2013.
- [5] K. Chen, R. Ge, H. Qiu, R. Ai-Rfou, C. R. Qi, X. Zhou, Z. Yang, S. Ettinger, P. Sun, Z. Leng, M. Mustafa, I. Bogun, W. Wang, M. Tan, and D. Anguelov, "Womd-lidar: Raw sensor dataset benchmark for motion forecasting," 2023.
- [6] Z. Yan, T. Duckett, and N. Bellotto, "Online learning for human classification in 3d lidar-based tracking," in *Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Vancouver, Canada, September 2017.
- [7] B. Schölkopf, R. C. Williamson, A. Smola, J. Shawe-Taylor, and J. Platt, "Support vector method for novelty detection," *Advances in neural information processing systems*, vol. 12, 1999.
- [8] C. Benedek, B. Gálai, B. Nagy, and Z. Jankó, "Lidar-based gait analysis and activity recognition in a 4d surveillance system," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 1, pp. 101–113, 2016.
- [9] J. Gómez, O. Aycard, and J. Baber, "Efficient detection and tracking of human using 3d lidar sensor," *Sensors*, vol. 23, no. 10, p. 4720, 2023.
- [10] K. Zhou, A. Paiement, and M. Mirmehdi, "Detecting humans in rgb-d data with cnns," in *2017 Fifteenth IAPR international conference on machine vision applications (MVA)*. IEEE, 2017, pp. 306–309.
- [11] H. Li, J. Liu, G. Zhang, Y. Gao, and Y. Wu, "Multi-glimpse lstm with color-depth feature fusion for human detection," in *2017 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2017, pp. 905–909.
- [12] K. O. Arras, O. M. Mozos, and W. Burgard, "Using boosted features for the detection of people in 2d range data," in *Proceedings 2007 IEEE international conference on robotics and automation*. IEEE, 2007, pp. 3402–3407.
- [13] L. Beyer, A. Hermans, T. Linder, K. O. Arras, and B. Leibe, "Deep person detection in 2d range data," *arXiv preprint arXiv:1804.02463*, 2018.
- [14] L. E. Navarro-Serment, C. Mertz, and M. Hebert, "Pedestrian detection and tracking using three-dimensional lidar data," *The International Journal of Robotics Research*, vol. 29, no. 12, pp. 1516–1528, 2010.
- [15] K. Kidono, T. Miyasaka, A. Watanabe, T. Naito, and J. Miura, "Pedestrian recognition using high-definition lidar," in *2011 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2011, pp. 405–410.
- [16] Z. Yan, T. Duckett, and N. Bellotto, "Online learning for human classification in 3d lidar-based tracking," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 864–871.
- [17] Z. Yan, L. Sun, T. Duckett, and N. Bellotto, "Multisensor online transfer learning for 3d lidar-based human detection with a mobile robot," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 7635–7640.
- [18] Z. Yan, T. Duckett, and N. Bellotto, "Online learning for 3d lidar-based human detection: experimental analysis of point cloud clustering and classification methods," *Autonomous Robots*, vol. 44, no. 2, pp. 147–164, 2020.
- [19] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 652–660.
- [20] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "Pointnet++: Deep hierarchical feature learning on point sets in a metric space," *Advances in neural information processing systems*, vol. 30, 2017.
- [21] W. et al., "Millimeter-wave radar detection and localization of static humans in indoor environments," *Remote Sensing*, 2024.
- [22] K. et al., "Mmwave radar and vision fusion for object detection in autonomous driving," *Sensors*, 2022.
- [23] A. Paigwar, O. Ercent, C. Wolf, and C. Laugier, "Attentional pointnet for 3d-object detection in point clouds," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 0–0.
- [24] H. Eltaif, Y. Yaddaden, R. Cherif, and Y. Benahmed, "Enhancing biometric authentication efficiency: A hybrid approach exploiting iris modality and leveraging one-class svm," in *2023 IEEE Third International Conference on Signal, Control and Communication (SCC)*. IEEE, 2023, pp. 01–06.
- [25] I. Armeni, O. Sener, A. R. Zamir, H. Jiang, I. Brilakis, M. Fischer, and S. Savarese, "3d semantic parsing of large-scale indoor spaces," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1534–1543.