From Point Cloud Perception Toward People Detection

Assia Belbachir[1,2][a], Antonio M. Ortiz[1][b], Atle Aalerud[1][c] and Ahmed Nabil Belbachir[1][d]

[1] *NORCE Norwegian Research Centre, Norway*

[2] *Sorbonne Université, LIP6-UMR 7606 CNRS, France*

Keywords: Point Cloud, LiDAR, People Detection.

Abstract: Point clouds have become significant data inputs for 3D representation, enabling accurate analysis of 3D scenes and objects. People detection from point clouds is a challenging task due to data sparsity, irregularity, occlusion, and real-time detection constraints. Existing methods based on handcrafted features or deep learning have limitations in handling occlusions, pose variations, and fast detection. This paper introduces a Random Forest classifier for people detection in point clouds, aiming to achieve both accuracy and fast performance. The point cloud data are acquired using a multi-point LiDAR system. First experiments demonstrate the effectiveness of the approach and its efficient detection compared to Multiple Layer Perceptron (MLP) in our collected Dataset.

1 INTRODUCTION

Point cloud perception has emerged as a crucial research area in computer vision, enabling accurate analysis and understanding of 3D scenes and objects. With the increasing availability of depth sensors such as LiDAR and RGB-D cameras, point clouds have become a popular representation for capturing the geometric structure of the environment. A key application of point cloud perception is people detection, which plays an important role in various domains including autonomous driving, surveillance, and human robot interaction, among others.

People detection from point clouds is a challenging task. The inherent sparsity, irregularity, and occlusion present in point clouds make accurate detection of humans a complex problem. Existing approaches often rely on handcrafted features or deep learning-based methods that primarily focus on local geometric patterns or global contextual information. However, these approaches have limitations in dealing with occlusion, pose variations, and cluttered scenes, leading to sub optimal performance.

Two main methods are used for people detection: (1) handcrafted feature-based methods and (2) deep learning-based methods. Handcrafted feature-based methods extract geometric features, such as local surface normals, curvatures, or multi-scale shape descriptors, to represent the local characteristics of point neighborhoods. These methods can use clustering or classification algorithms to differentiate humans from other objects or background. While these approaches have shown promising results in some scenarios, they struggle to handle complex occlusion patterns, pose variations, and real-time detection.

Recently, deep learning-based methods have gained significant attention, leveraging the ability of deep neural networks to learn features directly from point clouds. These methods often employ architectures such as PointNet (Qi et al., 2017b), PointNet++ (Qi et al., 2017d), or Graph Convolutional Networks (GCNs) (Li et al., 2018a) to capture local and global contextual information. Deep learning-based approaches have shown superior performance in handling complex scenes, occlusion, and pose variations. However, they may suffer from high computational costs and require large amounts of labeled training data.

The Random Forest classifier (RFC) is a well-established machine learning algorithm with a strong track record of high performance across various domains. While the field of machine learning continually evolves, the Random Forest classifier remains a robust and reliable choice for classification tasks in term of scalability and reactivity (Cutler et al., 2007).

[a] https://orcid.org/0000-0002-1294-8478
[b] https://orcid.org/0000-0002-7145-8241
[c] https://orcid.org/0000-0001-6462-235X
[d] https://orcid.org/0000-0001-9233-3723

Belbachir, A., Ortiz, A., Aalerud, A. and Belbachir, A.
From Point Cloud Perception Toward People Detection.
DOI: 10.5220/0012258800003543
In *Proceedings of the 20th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2023) - Volume 1*, pages 520-526
ISBN: 978-989-758-670-5; ISSN: 2184-2809

Before training our model for RFC, we removed the ground information. We conducted a comparison between the Random Forest Classifier (RFC) and the Multi-Layer Perceptron (MLP) through various evaluation metrics using the datasets.

The objective of this paper is twofold: first, it applies Random Forest classification mechanisms for people detection from a point cloud perception, and going one step beyond, the use of a trained model enables real-time detection of people. The point cloud perception used as input for the Random Forest classification is acquired through a multi-point LiDAR (Aalerud et al., 2020) that uses a combination of multiple mirrors to increase the number of detected points, thus increasing the resolution and data quality.

The remainder of this paper is organised as follows: Section 2 reviews the related State of the Art and presents the main drawbacks of previous work; Section 3 details the fundamentals of the proposed approach and the main novelties of this work; Section 4 presents the experiments and results highlighting the obtained performance; and finally, Section 5 presents some conclusions and future work, paving the way to the application of the proposed approach in diverse domains.

2 STATE OF THE ART

The literature presents multiple Point Cloud methods that can be classified into seven categories: (i) Point Cloud Segmentation (such as PointNet (Qi et al., 2017c), PointNet++ (Qi et al., 2017e), and PointCNN (Li et al., 2018b)), (ii) Point Cloud Registration (Huang et al., 2021), (iii) Point Cloud Reconstruction (Lin et al., 2018), (iv) Point Cloud Classification (Uy et al., 2019), (v) Point Cloud Denoising (Javaheri et al., 2017), (vi) Point Cloud Generation (Yang et al., 2019a), and (vii) Point Cloud Compression (Cao et al., 2019).

This paper focuses on people detection, which involves point cloud segmentation followed by classification.

The work in (Dai et al., 2021) presents MV3D, a multi-view 3D object detection network that utilizes point cloud data for autonomous driving applications. It combines bird's-eye view and front view representations with 3D voxel-based feature learning. MV3D achieves state-of-the-art performance in object detection tasks, including people detection. However, MV3D's computational complexity and reliance on multi-view inputs may limit its real-time applicability in resource-constrained scenarios.

Frustum PointNets (Qi et al., 2017a) introduces a frustum-based approach for 3D object detection, including people detection, using RGB-D data. The method extracts point features within frustum regions and employs PointNet to classify and regress object properties. Frustum PointNets demonstrates superior performance on challenging datasets. While effective, the frustum-based approach may suffer from incomplete object detection when objects extend outside the frustum, limiting its performance in scenarios with large-scale occlusion.

The authors of PointRCNN (Shi et al., 2019) propose a two-stage framework for 3D object detection that generates proposals and performs accurate detection using point cloud data. The method leverages region proposal network (RPN) and PointNet++ for feature extraction, achieving state-of-the-art results in various object detection benchmarks, including people detection. However, the effectiveness of this method relies on region proposal, which may limit its efficiency for real-time applications. Additionally, it may struggle with densely packed scenes due to limited coverage of the proposed solution.

STD (Yang et al., 2019b) is a sparse-to-dense 3D object detection framework that reconstructs dense and complete point clouds from sparse input. The method leverages depth completion and 3D feature learning to improve object detection accuracy. STD achieves competitive results in object detection, including human detection tasks. However, STD's reliance on depth completion may introduce errors in scenes with challenging depth estimation or limited sensor capabilities.

Complex-YOLO (Simon et al., 2018) proposes a real-time 3D object detection framework for point clouds. The method employs complex convolutional layers and anchor-based predictions to achieve efficient and accurate object detection. Complex-YOLO demonstrates competitive performance on various datasets. However, it may struggle with highly occluded objects due to limitations in the anchor-based prediction scheme.

Another technique introduces Histograms of Oriented Gradients (HOG) features combined with a Random Forest classifier for people detection (Dalal et al., 2006). This method captures shape and appearance characteristics of people, leading to accurate detection. However, the HOG-based approach may struggle with complex occlusion patterns and require careful parameter tuning for optimal performance. Additionally, using HOG for real-time detection is not appropriate as it may lack of the necessary performance despite some efforts for reducing features calculation time (Pedersoli et al., 2008).

Other solutions are available such as the "Segment

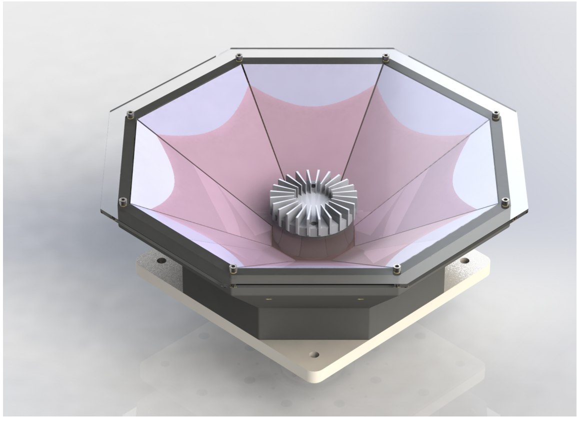

Figure 1: LiDAR used to capture the dataset input data (Aalerud et al., 2020).

Anything" project by Meta[1], which focuses on image segmentation.

The work presented in this paper utilizes a newly developed LiDAR to accurately locate people using a Random Forest classifier, outperforming state of the art solutions reducing installation and data processing costs, with a potential application in diverse fields such as industrial environments, autonomous driving, human-robot interaction, safety and security, surveillance, etc.

3 PROPOSED APPROACH

This section provides an overview of the dataset used in the present study, including details on its collection process and characteristics. Then, the methodology used for people detection is explained, outlining the steps involved in recognising people entities within the dataset.

3.1 Sensor Description

The dataset is reliant on a LiDAR system that has its origins in the research described in (Aalerud et al., 2020). This LiDAR system prototype, as shown in Figure 1, stands out due to its innovative design, featuring eight individual mirror segments inclined at an angle of 34°. More specifically, the LiDAR model used in this system is the OS1-128, produced by Ouster [2]. The OS1-128 boasts an impressive array of 128 laser emitters evenly distributed over a scanning span of 45°. Each of these laser emitters can generate a remarkable 4096 individual data points in a single frame.

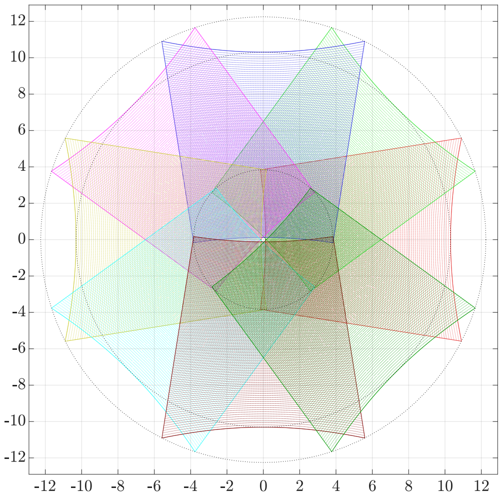

Figure 2: Illustration LiDAR's coverage in meter from a distance of 10 meter.

For the dataset generation process, a configuration was chosen that employs 2048 data points per frame and operates at a consistent frame rate of 10 frames per second (10 fps). This configuration was selected to facilitate the capture of detailed and dynamic environmental data, making it well-suited for various applications where precision and real-time sensing are critical.

A visual representation of the LiDAR system's capabilities can be seen in Figure 2 from a 10-meter distance. This Figure provides an informative illustration of the system's field of view (FOV), demonstrating its ability to detect objects and surfaces within a meter range when situated 10 meters away from the subject. Additionally, this FOV visualization highlights the LiDAR system's capacity to capture densely packed spatial data for a specific area of interest.

3.2 Dataset

The dataset was collected in a crowded environment (conference), where several humans were walking, sitting, and standing with tables around, stairs, walls, etc. An example of the collected point cloud is illustrated in Figure 3. The LiDAR provides relative positional information, represented by the coordinates (x, y, and z). In our case of study, the LiDAR is fixed in a specific height in order to collect the positional information.

[1] Segment Anything link available at https://segment-anything.com/, consulted on October 2023

[2] Additional details can be found at https://ouster.com/blog/introducing-the-os-1-128-lidar-sensor/, accessed in October 2023

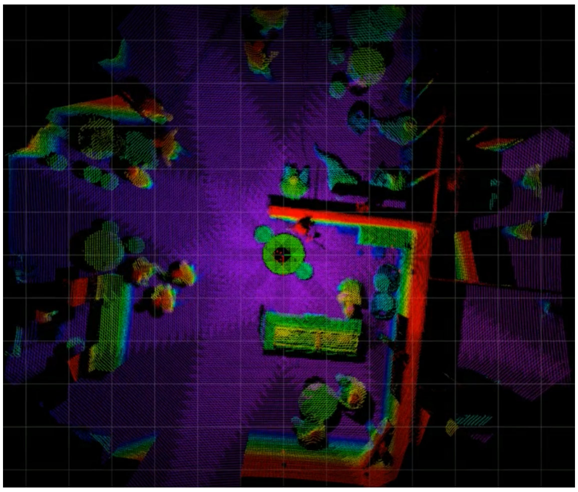

Figure 3: Visualisation of the obtained point cloud using the developed LiDAR.

3.3 Developed Architecture

Figure 4 represents the developed architecture in order to detect people from the perceived point cloud. First, raw data are collected from the LiDAR. Second, information related to the ground is removed. In the context of ground removal in point cloud data, a region growing algorithm can be applied to identify and separate the ground points from other objects. The algorithm typically starts with a seed point known to be on the ground surface. These points are more or less known due to the knowledge of the approximate sensor position. It then examines the neighboring points and checks if they satisfy certain criteria, such as having similar elevation values within a predefined threshold. If a neighboring point meets the criteria, it is added to the region. This process continues in an iterative way until no more points can be added to the region. By utilizing a region growing algorithm, the ground points are segmented and then removed for the case of study. Third, the main dataset is used to train our Random Forest Classifier. The generalized developed algorithm is shown in Algorithm 1.

Random Forest Classifier. The Random Forest classifier is an ensemble learning algorithm that combines the outputs of multiple decision tree classifiers to perform classification tasks (Belgiu and Drăguţ, 2016). It creates an ensemble, or a collection, of decision trees where each tree is trained independently on a randomly sampled subset of the training data. During the training process, each decision tree in the ensemble learns to classify the data based on different features and thresholds.

```
Data: Collected Database
Result: People detection in the Point Cloud
Generation of training (TrDB) and testing
  database (TsDB);
while Point Cloud Perception $\in$ TrDB do
    Read current perception;
    Ground removal;
    Features extraction;
    Training using Random Forest Classifier;
end
while Point Cloud Perception $\in$ TsDB do
    Read current perception;
    Ground removal;
    Features extraction;
    if Features corresponds to people then
        people detection;
    end
end
```

Algorithm 1: The developed algorithm to detect people.

When making predictions using the Random Forest classifier, each decision tree in the ensemble independently assigns a class label to the input data. The final prediction is determined by combining the predictions of all the decision trees, typically through majority voting. The class label that receives the majority of votes from the decision trees is selected as the final predicted class label.

Training and Testing. For training, the labeled dataset is used to train the Random Forest classifier, enabling it to learn the distinguishing features of people entities. The classifier leverages the collective knowledge of the decision trees to make accurate predictions. During the testing phase, the trained Random Forest classifier is deployed on new, unseen point cloud data to detect the presence of people. The classifier assigns class labels to the input data based on the majority votes from the decision trees in the ensemble. This majority voting mechanism ensures robust and reliable people detection results.

By training and testing the Random Forest classifier on the labeled dataset, the proposed approach achieves fast people detection in point cloud data. Additionally, we removed the ground before training our data which contributes to accurate and reliable detection of people in real-world scenarios.

4 EXPERIMENTAL RESULTS

To evaluate the performance of the proposed approach, a series of experiments were conducted on

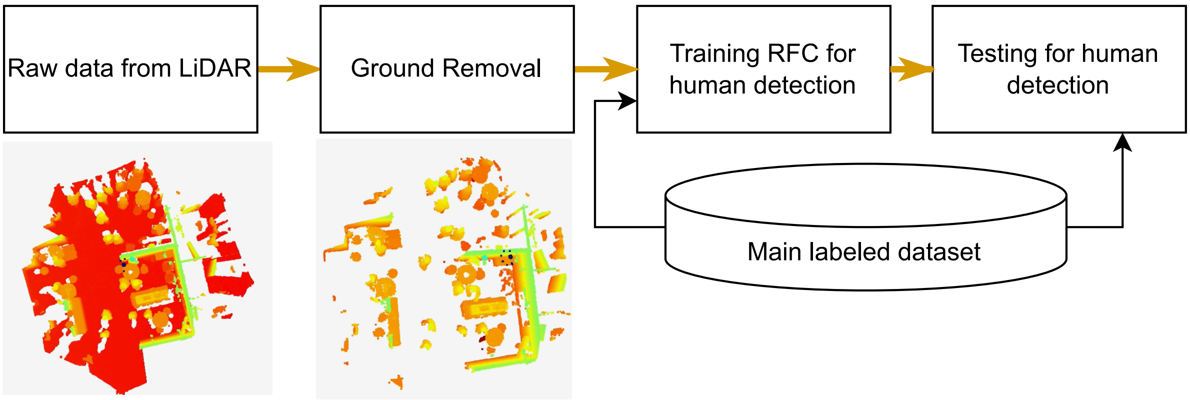

Figure 4: Illustration of the developed methods for human detection.

the dataset comprising various real-world scenarios. The dataset consisted of point cloud data captured from a conference. We divided our dataset into two: Dataset1 and Dataset2. Each of them contains 19 humans each, which are walking, standing, and sitting.

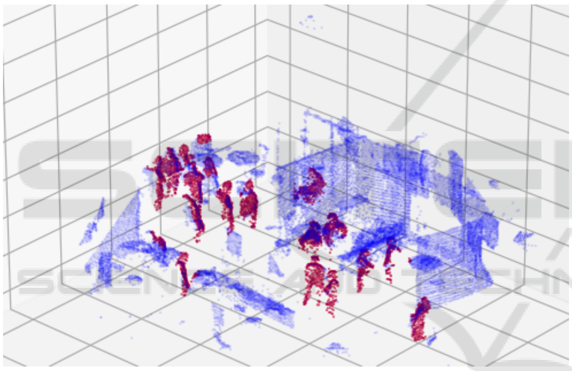

Figure 5: Illustration of a human detection example. Blue color represents the perceived point cloud from the LiDAR while red color represents the detected humans in the scene using the Random Forest classifier.

It is worth mentioning that all experiments presented herein have been performed using a MacBook Pro with Apple M1 Pro processor and 16GB RAM, and no GPU, cluster, or advanced equipment has been used, proving that the proposed solution can be easily replicated and effectively implemented in real-world scenarios.

In the conducted experiments, Dataset1 and Dataset2, were utilized. These datasets include people activity and behavior during conference sessions and were collected using the LiDAR that was recording from an elevation of 10 meters.

Dataset1, despite being smaller, still encompasses a diverse range of people poses and activities observed during the recorded scene. It serves as a representative sample to evaluate the proposed approach and assess its performance on a relatively limited dataset from a conference. The smaller size of Dataset1 allows for a focused analysis of the approach effectiveness in detecting people with different configurations within the conference environment.

Table 1: Example of annotated data.

```
"psr": {
    "position": {
        "x": 1.8860341036140744,
        "y": 2.0482052752197473,
        "z": -1.9140917764041696
                },
    "scale": {
        "x": 0.4826413579974756,
        "y": 0.8237351830145311,
        "z": 1.801308247033618
            },
    "rotation": {
        "x": 0.013942779107655656,
        "y": -0.007629356641117054,
        "z": 1.119411576214597
                }
```

On the other hand, Dataset2 is a larger and more comprehensive dataset, capturing a broader range of people behaviors. With its increased size and variety of instances, Dataset2 provides a more extensive and robust evaluation of the proposed approach. The larger amount of data in Dataset2 enables a more comprehensive assessment of the approach's performance across various scenarios and configurations encountered during the conference.

By using both Dataset1 and Dataset2, the experiments provide insights into the scalability of the proposed approach. The evaluation results obtained from these datasets contribute on understanding the approach's performance in both limited and extensive data settings derived from conference environments.

Each dataset is divided into a learning (80%) and a testing dataset (20%). We annotated each person with

Table 2: Illustration of the obtained results for human detection using RFC (Random Forest Classifier) and MLP (Multi-Layer Perceptron).

Method	Dataset	Precision	Computation time (s)	Learning time (s)
RFC	Data1	0.95	0.013	0.08
MLP	Data1	0.95	0.013	0.079
RFC	Data2	0.95	**0.014**	**0.069**
MLP	Data2	0.74	0.016	0.079

three types of information: positional information relative to the liDAR (x, y and z), scale (x, y and z), and rotation (x, y and z) parameters, thus providing more information to feature selection and learning. An example of an annotated data is represented in Table 1.

The availability of annotated data serves as a solid foundation for learning and evaluating the implemented approach. The evaluation of the proposed approach encompasses a comprehensive analysis based on multiple evaluation metrics, including Precision, Recall, F1-score, Support, Computation time, and Learning time.

This comprehensive assessment provides a thorough understanding of the approach's performance across various dimensions, ensuring a robust evaluation of its effectiveness.

Precision measures the accuracy of the positive predictions for a given class. It is calculated as the ratio of true positives to the sum of true positives and false positives. A higher precision indicates fewer false positives. Computation time represents the time that is spent in order to detect a human from a point cloud perception. Learning time refers to the duration taken by the method to acquire knowledge from the provided training data.

Table 2 shows the obtained results using Random Forest classifier (RCF) and Multi-Layer Perceptron (MLP) (Goodfellow et al., 2016). From Dataset1, we can see that both approaches are similar in all the criteria. Even from the computational and learning time both approaches are similar.

However, for Dataset2, Random Forest classifier gives a better result in term computational time (13.5% faster). Additionally, the learning time of RFC is faster (12,7%) than MLP.

This result confirms that, using our LiDAR which perceives 10fps, the proposed approach can be considered as a real-time solution. Due to its fast learning, we can also think of embedding our approach as continuous learning technique if new objects need to be detected. Therefore, RCF can be extended to human tracking.

5 CONCLUSIONS

This paper presented an approach for fast people detection in point cloud perception using a Random Forest classifier. The main challenge addressed is related to the fast detection of people using real collected data.

The utilization of a multi-point LiDAR system contributed to enhanced data quality and resolution, enabling robust and fast people detection. The experimental results demonstrated the effectiveness of the approach in diverse environments, showcasing its potential for practical applications in various domains.

This paper contributes to the advancement of a precise and fast people detection in point cloud perception by providing a reliable and efficient solution with a new developed LiDAR. The approach was compared to Multi-Layer Perceptron (MLP). For a bigger dataset scenario, the proposed approach has obtained better results.

The scalability and versatility of the Random Forest classifier make it an appealing choice for extending the approach to other object recognition tasks in point cloud analysis. The proposed approach has the potential to significantly impact the fields of autonomous driving, surveillance, human-robot interaction, industrial environments, safety and security, etc., enabling safer and more efficient interactions between humans and intelligent systems, with a high market potential.

Overall, the proposed approach opens new avenues for advancing in real-time human tracking in point cloud perception, contributing to the broader goal of improving situational awareness and enabling intelligent systems to effectively perceive and interact with their dynamic environments.

Future work includes larger scale testing and further validation activities in different environments (e.g. industrial areas, roads, etc.), as well as the comparison with other existing approaches for real-time people detection.

ACKNOWLEDGEMENTS

This work is co-funded by the Research Council of Norway under the project entitled “Next Generation 3D Machine Vision with Embedded Visual Computing” with grant number 325748.

REFERENCES

Aalerud, A., Dybedal, J., and Subedi, D. (2020). Reshaping field of view and resolution with segmented reflectors: Bridging the gap between rotating and solid-state lidars. *Sensors*, 20(12).

Belgiu, M. and Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS journal of photogrammetry and remote sensing*, 114:24–31.

Cao, C., Preda, M., and Zaharia, T. (2019). 3d point cloud compression: A survey. In *The 24th International Conference on 3D Web Technology*, pages 1–9.

Cutler, D. R., Edwards Jr., T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., and Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, 88(11):2783–2792.

Dai, D., Chen, Z., Bao, P., and Wang, J. (2021). A review of 3d object detection for autonomous driving of electric vehicles. *World Electric Vehicle Journal*, 12(3).

Dalal, N., Triggs, B., and Schmid, C. (2006). Human detection using oriented histograms of flow and appearance. volume 3952, pages 428–441.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). Deep learning. *Nature*, 521(7553):436–444.

Huang, X., Mei, G., Zhang, J., and Abbas, R. (2021). A comprehensive survey on point cloud registration. *arXiv preprint arXiv:2103.02690*.

Javaheri, A., Brites, C., Pereira, F., and Ascenso, J. (2017). Subjective and objective quality evaluation of 3d point cloud denoising algorithms. In *2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pages 1–6. IEEE.

Li, Y., Bu, R., Sun, M., Wu, W., Di, X., and Chen, B. (2018a). Pointcnn: Convolution on x-transformed points. In Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc.

Li, Y., Bu, R., Sun, M., Wu, W., Di, X., and Chen, B. (2018b). Pointcnn: Convolution on x-transformed points. *Advances in neural information processing systems*, 31.

Lin, C.-H., Kong, C., and Lucey, S. (2018). Learning efficient point cloud generation for dense 3d object reconstruction. In *proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.

Pedersoli, M., Gonzalez, J., Chakraborty, B., and Villanueva, J. J. (2008). Enhancing real-time human detection based on histograms of oriented gradients. In *Computer Recognition Systems 2*, pages 739–746. Springer.

Qi, C. R., Liu, W., Wu, C., Su, H., and Guibas, L. J. (2017a). Frustum pointnets for 3d object detection from RGB-D data. *CoRR*, abs/1711.08488.

Qi, C. R., Su, H., Mo, K., and Guibas, L. J. (2017b). Pointnet: Deep learning on point sets for 3d classification and segmentation.

Qi, C. R., Su, H., Mo, K., and Guibas, L. J. (2017c). Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 652–660.

Qi, C. R., Yi, L., Su, H., and Guibas, L. J. (2017d). Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Qi, C. R., Yi, L., Su, H., and Guibas, L. J. (2017e). Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*, 30.

Shi, S., Wang, X., and Li, H. (2019). Pointrcnn: 3d object proposal generation and detection from point cloud. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, pages 770–779. Computer Vision Foundation / IEEE.

Simon, M., Milz, S., Amende, K., and Gross, H. (2018). Complex-yolo: Real-time 3d object detection on point clouds. *CoRR*, abs/1803.06199.

Uy, M. A., Pham, Q.-H., Hua, B.-S., Nguyen, T., and Yeung, S.-K. (2019). Revisiting point cloud classification: A new benchmark dataset and classification model on real-world data. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1588–1597.

Yang, G., Huang, X., Hao, Z., Liu, M.-Y., Belongie, S., and Hariharan, B. (2019a). Pointflow: 3d point cloud generation with continuous normalizing flows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4541–4550.

Yang, Z., Sun, Y., Liu, S., Shen, X., and Jia, J. (2019b). STD: sparse-to-dense 3d object detector for point cloud. *CoRR*, abs/1907.10471.