Carnegie Mellon University
School of Architecture - College of Fine Arts
**Thesis Submitted in Partial Fulfillment of the requirements for the degree of Master of Science in Computational Design**

Title:

# **Recognizable Point-cloud**:
# An end-to-end point-cloud processing, visualization, and user-interaction design

*Autor*:
**Zishen Wen**

*Advisory Committee*:

| | | |
|---|---|---|
| **Daniel Cardoso Llach** | Principal Advisor | Dec 28, 2022<br>DATE |
| **Eddy Man Kim** | Reader | Dec 31, 2022<br>DATE |

# Abstract

Point clouds have been widely used in various industries, such as autonomous driving, monitoring, and surveying. However, the visualization methods developed for these fields are limited compared to the rapidly evolving techniques for point cloud processing. Most of these visualization approaches are primarily focused on technical considerations, such as memory usage, computation latency, and visualization accuracy, rather than the user experience. The aim of this thesis is to design a new framework for point cloud visualization that incorporates state-of-the-art techniques and to evaluate its effectiveness through user recognition testing. The work involves the following steps: (1) conducting a literature review and comparing different baseline models for learning-based point cloud techniques, including detection and upsampling; (2) constructing a pipeline to train models and generate multiple point cloud visualization options; (3) developing a user interface with a set of questions on object, semantic, and spatial recognition; and (4) conducting user research to assess the visualization options. This study provides a comprehensive review of point cloud upsampling and detection methods, and presents a novel framework for enhancing point cloud visualization through the incorporation of advanced techniques. It also contributes to user-centered research on point clouds by introducing various forms of recognition as key factors.

## Acknowledgments

I'd want to take this time to thank my adviser, Prof. Daniel Cardoso Llach, for his direction, support, and encouragement during my thesis research. His knowledge and perspective have been crucial in defining my study and assisting me in reaching my objectives. I'd also want to thank my reader, Prof. Eddy Man Kim for engaging with my work and contributing vital feedback and thoughts. Your support and interest are highly appreciated.

I would also like to thank my parents for their love and support. Their unwavering belief in me and my abilities has been a constant source of strength and motivation.

Lastly, I would also like to thank my friends for their friendship and for being there for me during the ups and downs of the thesis process. Their support and encouragement have meant the world to me.

Overall, I am deeply grateful to all of these individuals for their contributions to my thesis and for helping to make this journey possible. Thank you!

**1. Introduction** **6**

1.1. Design as a Way of Seeing 6

1.2. Thesis Vision 7

1.3. Thesis Summary 7

**2. Background** **9**

2.1. Information Visualization and Representation 9

2.2. Point Cloud and its Learning-based Approaches 14

(1) Point Cloud and LiDAR 14

(2) Learning-based Point Cloud Methods 16

(3) Point Cloud Classification 18

(4) Point Cloud People Detection 20

(5) Point Cloud Upsampling 24

2.3. Point Cloud Visualization 27

2.4. Recognizability 30

**3. Hypothesis and Research Questions** **35**

3.1. Hypothesis 35

3.2. Research Questions 35

(1) How to visualize point-cloud? 35

(2) How to leverage recognizability? 35

(3) How to foster new engagement? 36

3.3. Research Goal 36

**4. Methodologies** **37**

**5. Technical Implementation** **38**

5.1. Framework 38

5.2. Dataset 38

5.3. People Detection Model 40

5.4. Upsampling Model 45

5.5. Combination of Detection and Upsampling 49

5.6. Design of User Interface 50

**6. User Study 54**

6.1. Overview 54

6.2. Types of Recognizability 55

6.3. Question Design 58

**7. Results 62**

7.1. User Study Result Overview 62

7.2. Analysis 64

7.3. Summary 66

**8. Speculative Design 68**

8.1. Speculative Design 68

8.2. Design 1: Night Vision Glasses 69

8.3. Design 2: Dance with Cloud 71

8.4. Design 3: Automated construction assistant 75

**9. Conclusion 78**

9.1. Contribution 78

9.2. Limitation and Future Work 78

**bibliography 80**

# 1. Introduction

## 1.1. Design as a Way of Seeing

The book *Social Development and Symbolic Landscape* by Denis Cosgrove had a significant impact on my studies of landscape and architecture. Cosgrove argued that "landscape is a method of perceiving,[1]" and suggested using landscape as a way to understand the social and physical development of the environment. This expanded the concept of landscape beyond traditional areas such as art, gardening, and park design. As a result of reading Cosgrove's work, I began to view design not just as a process of turning ideologies into forms and implementing them, but also as a means of interpreting and understanding the world and its social and cultural development. Cosgrove's ideas helped me to see design in a more comprehensive and nuanced way.

In recent years, the rapid advancement of technology and the promotion of technological optimism have led to a focus on embracing new and efficient techniques. However, in these trends, humanism often takes a back seat as it is seen as having no contribution to "improvement." In technical development, "improvement" is often equated with higher efficiency and precision, and a humanistic perspective is viewed as a hindrance. Yet, we have seen cases where a technically "better" solution has caused social problems due to a lack of consideration for humanistic values. For instance, the widespread use of big data for predicting user behavior has raised concerns about data privacy, and the reliance on social media recommendations has had negative impacts on democracy by influencing voting decisions.

These observations have led me to reconsider the role of design, particularly computational design, in today's society. Is it possible for us to be active participants in technical development while also offering alternative perspectives to the dominant voices of scientists and engineers? Can we provide alternative ways of seeing and understanding technical questions, and reclaim the value of humanism in the process? As designers, can we contribute to technical development in a meaningful way while also bringing a humanistic perspective to the table?

---

[1] Cosgrove, Denis E. *Social formation and symbolic landscape*. University of Wisconsin Press, 1998.

## 1.2. Thesis Vision

This work is my graduate thesis for the CMU MSCD program and addresses the question: "Can we leverage the recognizability of point cloud visualization to increase engagement?.”

The research is motivated by two factors: my interest in visualization and representation as a way of perceiving the world, and my learning in the MSCD program about the importance of situating technologies within the context of human experience and social movements.

I chose to focus on point clouds as a case study of a rapidly developing technique that has not always considered the needs and perspectives of users. This work aims to demonstrate the potential for improving point-cloud visualization that considers technical advancement, user recognizability, and engagement analysis. It is not intended to create a state-of-the-art visualization tool or to improve the efficiency or quality of existing methods. Rather, it aims to bring a new approach that prioritizes the user experience to the development of visualization tools.

This work also builds on my previous research in the ReAC (Rethinking Automation in Construction) project, led by Professors Daniel Cardoso Llach, Jean Oh, and Dina El-Zanfaly, where I had the opportunity to work with point clouds and discuss my ideas with other research students.

## 1.3. Thesis Summary

This work begins by conducting a literature review on the theory of information visualization, point clouds and their learning-based methods (including detection and upsampling), point cloud visualization, and human recognition. This review provides a comprehensive background and available toolkit for the thesis, drawing on the historical context and current technical development in these areas. The literature review serves as a foundation for the research and helps to situate the work within the broader context of these fields.

The second aspect of this research involves building a demo visualization library that uses deep learning methods including detection and upsampling to enhance the visualization of point clouds. In this section, this work discusses the technical development framework, which includes the required dataset, model architectures, and training process. This work also presents the evaluation results of models developed in this approach. Additionally, this work designs a user interface to present the selected visualization results and related questions to the user. This demo visualization library serves as a proof-of-concept for our proposed approach and helps to demonstrate the potential of using deep learning methods to improve point cloud visualization.

Third, to further understand how people perceive and recognize point clouds, this research conducts user studies using augmented point cloud visualization. In this section, design questions are designed that focus on three levels of recognizability: object, semantic, and spatial. For object recognizability, I ask questions that involve counting objects. For semantic recognizability, I design questions that ask users to locate specific objects. For spatial recognizability, I present questions that ask users to estimate the size and area of a space. These questions are designed to understand how people interpret and understand point clouds, and to identify potential areas for improvement in our visualization approach.

Based on the results of the user study, this research proposes speculative applications for enhancing the use of point clouds with more engaging interfaces. In this section, I present several scenarios for using these new visualization tools in the fields of surveying, art performing and building construction. The results are presented through text descriptions and storyboard illustrations, which provide a concrete vision for how these tools could be used in real-world applications. The goal of this section is to demonstrate the potential value of the proposed visualization approach and to inspire further development and exploration of these ideas.

# 2. Background

## 2.1. Information Visualization and Representation

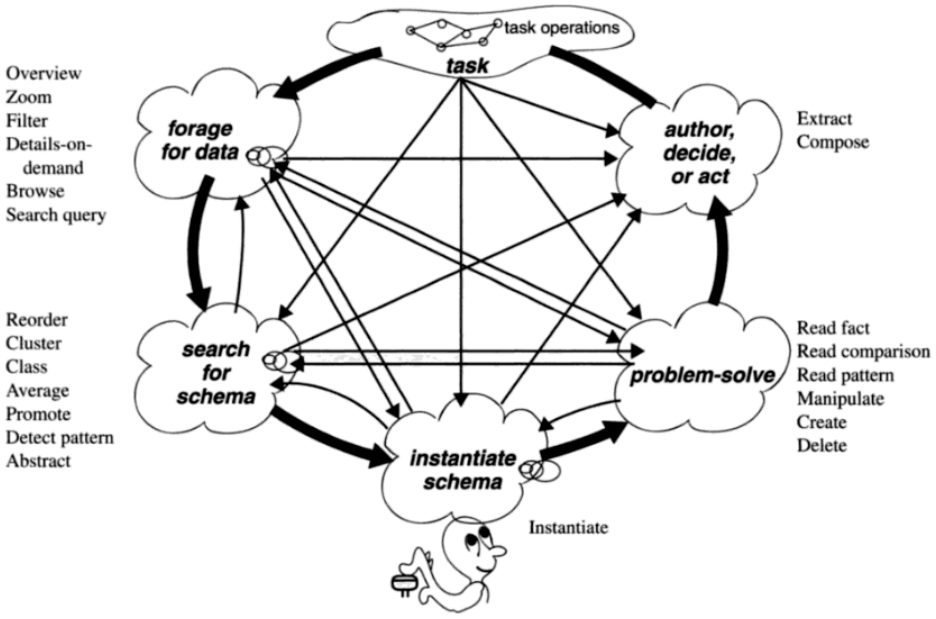



**Figure 1. The visualization process in human's perception**
Image Source: Card, Stuart K., Jock D. Mackinlay, and Ben. Shneiderman. Readings in Information Visualization : Using Vision to Think. San Francisco, Calif: Morgan Kaufmann Publishers, 1999. Print.

Visualization means displaying information with visible elements. It uses with the goal of communication. "Information visualization (InfoVis) is the communication of abstract data through the use of interactive visual interfaces"[2]. And it helps human to make decision and solve problems. It "utilizes computer graphics and interaction to assist humans in solving problems"[3].

---

[2] Keim, Daniel A., Florian Mansmann, Jörn Schneidewind, and Hartmut Ziegler. "Challenges in visual data analysis." In *Tenth International Conference on Information Visualisation (IV'06)*, pp. 9-16. IEEE, 2006.

[3] Purchase, Helen C., Natalia Andrienko, Thomas J. Jankun-Kelly, and Matthew Ward. "Theoretical foundations of information visualization." In *Information Visualization*, pp. 46-64. Springer, Berlin, Heidelberg, 2008.

The emerging of information visualization can be traced back to 18th century with two principles: **reduction** and the use of **spatial variables**[4].

**Reduction** refers to the process of filtering out noisy data in visualization. Information visualization uses primitive geometry and spatial elements such as points, lines, rectangles and circles to represent a wide range of topic such as demography, stock prices, temperature and so on. It retrieves and captures patterns from given observation and dataset. Since the world is chaotic and data are not always cleaned and structured, usually, we filter noisy to use only a small portion of our observation for visualization, with a hope that this can represent the general pattern of the thing we visualize[5]. Reductionism is not an unique idea only in visualization. It is actually a paradigm for all modern science such as physics, biology, chemistry and so on[6].

**Spatial variables** refers to the using position, size, shape or movement to represent key differences, patterns or relationships in the data[7]. In other words, by mapping information into different spatial topography, geometry shapes, covering areas, as well as color tones, we are able to present the result of information reduction to readers with hierarchies and layers. For instance, when reading a line chart, different line weight distinguishes priorities of reading. And different colors separate and regroups information for readers. The selection of different spatial variables impacts the way people perceive those information. One of the most common case is the difference between a bar chart and a line graph. The bar chart emphasizes the difference between each bar, while the line graph focuses on the trend of changes over all data.

Both reduction and spatial variables let us realize that visualization is not a one-to-one mapping between real world and visualized elements. Instead, **reduction** filtering “redundant” information and the using of **spatial variables** arranges remaining information. Therefore, the mechanism of reduction and rule of setting spatial variables will fundamentally determine information’s visualization and how it is perceived by readers.

---

[4] Manovich, Lev. "What is visualization." *paj: The Journal of the Initiative for Digital*.–2010 (2010).

[5] Same as 3.

[6] Gallagher, Richard, and Tim Appenzeller. "Beyond reductionism." *Science* 284, no. 5411 (1999): 79-79.

[7] Same as 3.

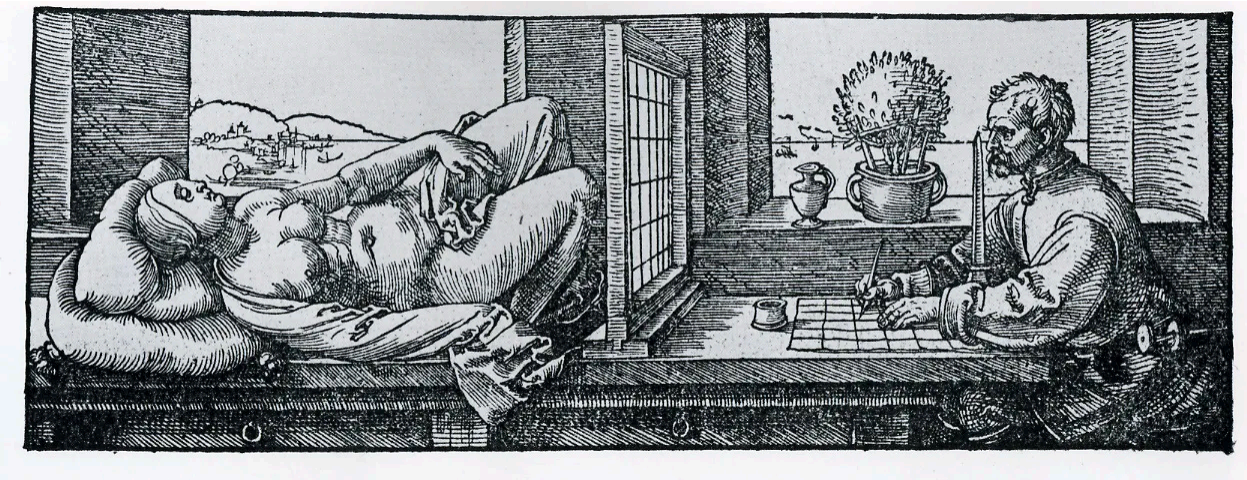


**Figure 2. Perspective Machine**
illustration published in Albrecht Dürer's The Painter's Manual, 1525.

There are a few examples can be used to represent the  history of visualization that involves **reduction** and **spatial variables**. The first one was the Drawing Machine developed by Albrecht Dürer in mid-century. In the 15th century, artists sought to more accurately depict the natural world and developed various machines to aid in their drawings. Leon Battista Alberti, an artist and mathematician, wrote the first treatise on perspective, Della Pittura, in 1435. One of the most successful drawing devices from this time was known as Alberti's Frame, which consisted of a wooden frame with horizontal and vertical threads stretched across it to form a grid. A rod, placed a foot in front of the frame and at the same height as the distance from the bottom of the frame to the middle of the grid, allowed the artist to maintain a consistent viewing position by aligning their eye with the rod and center of the grid. This helped the artist



**Figure 3. Utah Teapot**
https://en.wikipedia.org/wiki/Utah_teapot

to accurately depict the spatial relationships in their drawings. This process reduce any real object into points with relationship to a grid, and then represent them back to painted work by referring to another grid on canvas. It is in pre-digital era that people use manual process to conduct representation by seeking available rules to assist.

In digital era, people attempted to render object by mathematical modeling. One of the first example is the Utah teapot. The teapot model was created in 1975 by Martin Newell, a researcher in the field of computer graphics at the University of Utah. It was one of the first 3D models to be created using bézier curves, which are smooth, curved lines that are defined by a set of control points. The use of bézier curves allowed for more artistic freedom in the creation of the model, as it did not have to be precisely measured like previous models. The teapot model has since become a widely used benchmark and test object in the field of computer graphics.

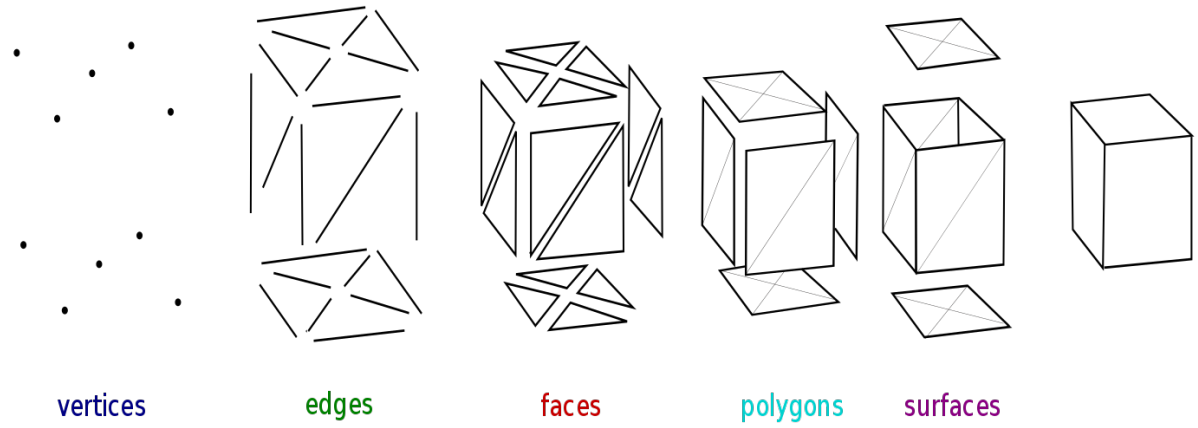



**Figure 4. Elements of Polygon Mesh**
https://en.wikipedia.org/wiki/Polygon_mesh

Later, the introduction of Polygon Mesh becomes the for computer graphics and rendering. A polygon mesh is a collection of vertices, edges, and faces that define the shape of a three-dimensional object in computer graphics and solid modeling. The faces of a polygon mesh are typically made up of triangles, quadrilaterals, or other simple convex polygons, although they can also be composed of more complex shapes such as concave polygons or polygons with holes. The study of polygon meshes is a significant field within computer graphics and geometric modeling, and there are a wide range of algorithms and operations that can be performed on them, such as Boolean logic, smoothing, simplification, ray tracing, collision detection, and rigid-body dynamics. If the edges of the mesh are rendered instead of the faces,

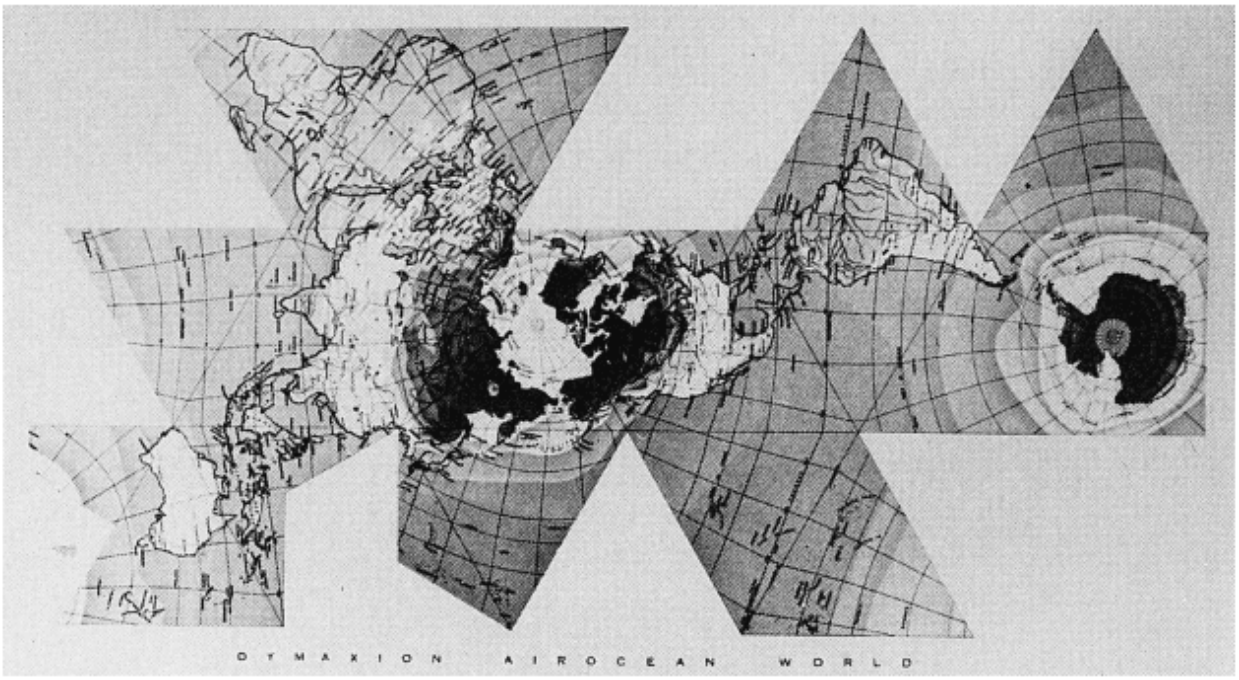



**Figure 5. R. Buckminster Fuller and Shoji Sada's world map**
Ramsey, Joseph. "A Fuller Map: Latent Meanings within Jasper Johns' Map (Based on Richard Buckminster Fuller's Dymaxion AirOcean World)." *Athanor 35* (2017): 77-84.

the model becomes a wireframe model. In contrast to polygon meshes, volumetric meshes explicitly represent both the surface and volume of a structure.

The center of visualization is a way of communication rather than showing objectivities of the world. We usually take the outcome of information visualization and representation for granted, and "forget that even basic representations are complex, challenging inscriptions and are not neutral; they are based on representational choices that impact how and what they communicate"[8]. Looking back to the architecture field, mapping is a very common method people use to visualize information, which provides opportunities for measuring and describing the world we live and imagine about. However, instead of "mirroring the world", mapping "reshaping" the world with agencies[9]. In cartography, various ways for select and visualize information and different forms of drawing will impact how it is presented to viewers. To expand this topic from architecture to the broader field of information visualization, Card et. al

[8] Cafaro, Francesco, and Jessica Roberts. "Data through Movement: Designing Embodied Human-Data Interaction for Informal Learning." *Synthesis Lectures on Visualization* 8.3 (2021): 1-127

[9] Corner, James. *The agency of mapping: speculation, critique and invention*. na, 1999.

summarize it as “to use perception to amplify cognition”[10]. It was called a “knowledge crystallization” process that humans gather information, frame representation, and then use it for communication[11].

In summary, the review of information visualization as an approaches to reveal information with perceivable elements leading to its goal as methods for thinking and communicating rather than revealing objective world. Based on this though, the evaluation and improvement of visualization techniques is also built upon user’s experience and reflection rather than the accuracy of visualization.

## 2.2. Point Cloud and its Learning-based Approaches

### (1) Point Cloud and LiDAR

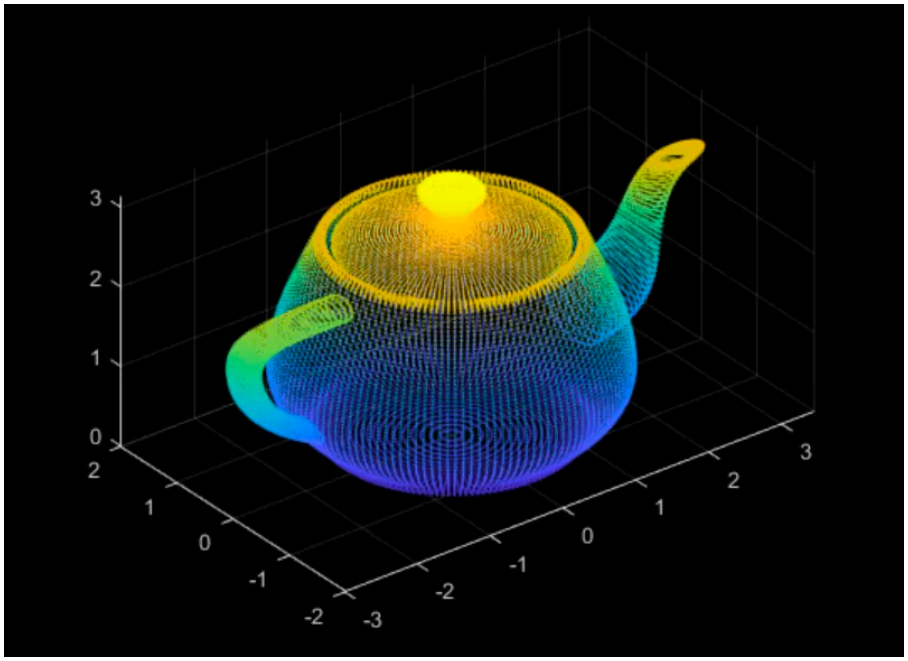



**Figure 6. Point cloud Object.**
Image Source: https://www.revopoint3d.com/point-cloud-and-3d-image/

Point cloud are gaining attention nowadays as a new spatial technique. Usually, we define a collection of points in three dimensions as a 3D point cloud. Instead of illustrating object by edges or surface, point

[10] Card, Stuart K., Jock D. Mackinlay, and Ben. Shneiderman. *Readings in Information Visualization : Using Vision to Think*. San Francisco, Calif: Morgan Kaufmann Publishers, 1999. Print.

[11] same as 2.

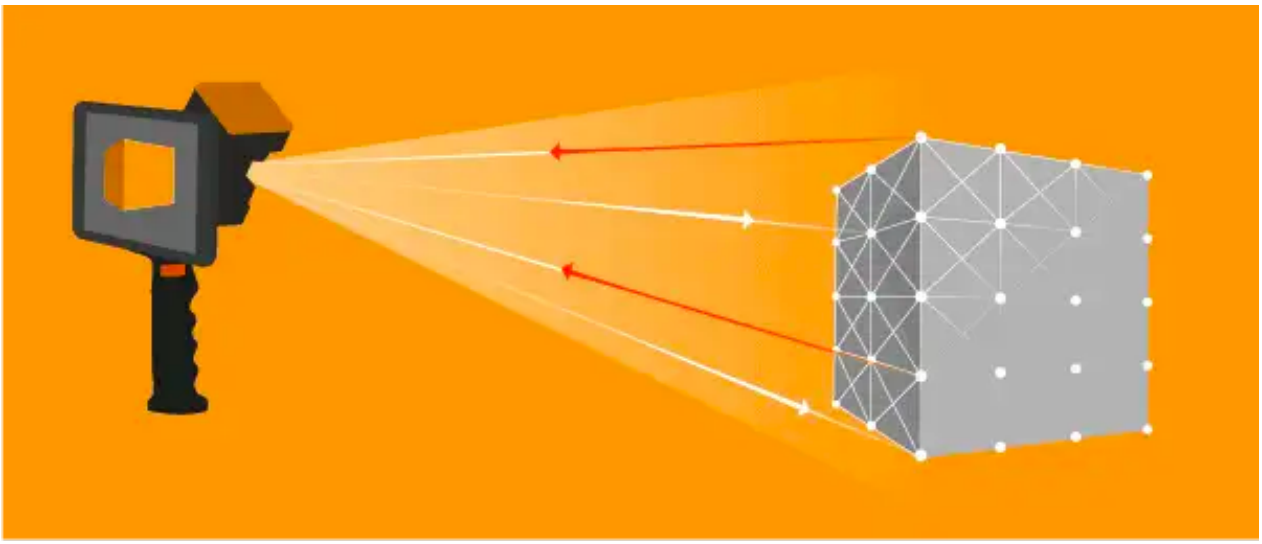


**Figure 7. Lidar Sensor.**
Image Source: https://geoslam.com/what-is-lidar/

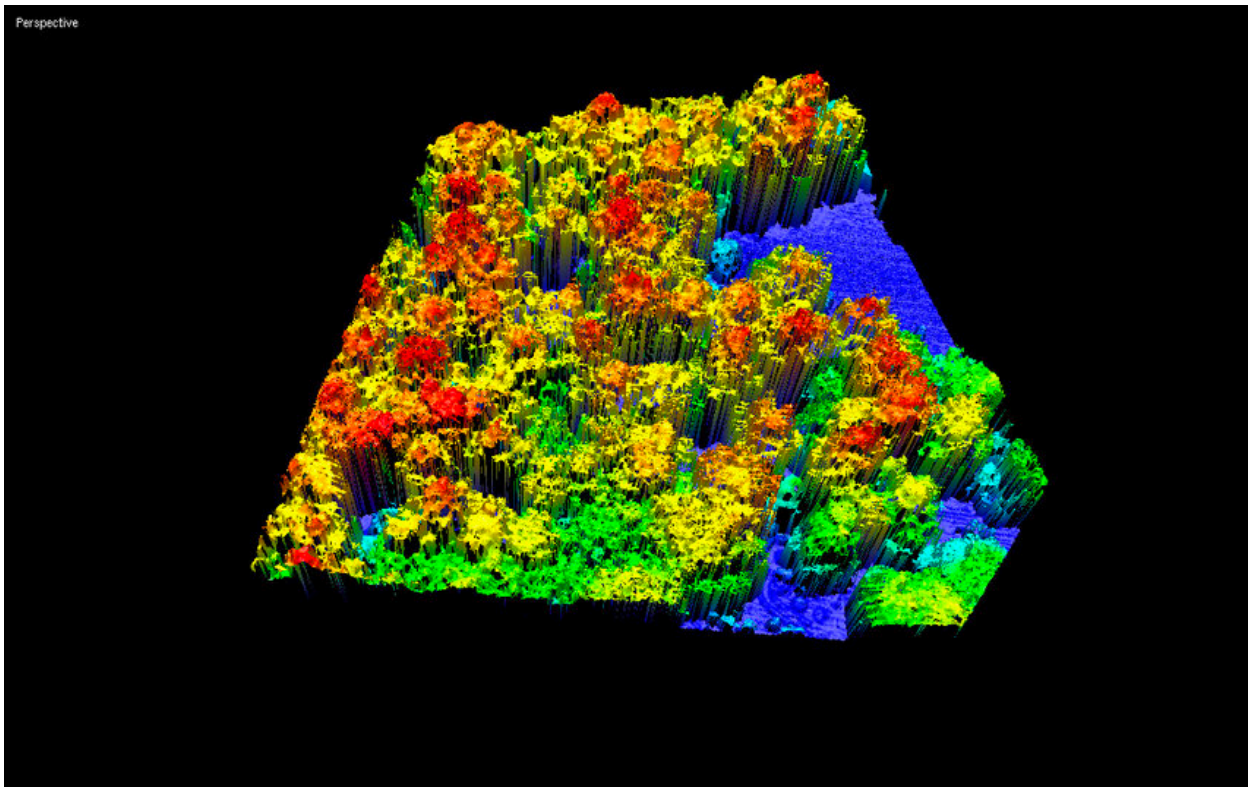



**Figure 8. 3D-LiDAR image generated from FP-mode 3D point cloud data.**
Image Source: Nakai, Y., F. Hosoi, and K. Omasa. "Estimating carbon stocks of coniferous woody canopy trees Using airborne lidar and passive optical senser." *Bretar F, Pierrot-Deseiligny M, Vosselman G* (2009): 289-292.

cloud use points to capture shapes. The surface of an object in spatial context can be represented as points. Each point in 3D point-cloud object will be described by its x, y, and z axes for spatial Cartesian coordinate system.

The most common way for point cloud generation is LiDAR sensors (Light Detection and Ranging). It contains three components for three steps in point cloud generation: emitter, rotator and photodetector. It can be understood as a process to emit laser in 180 degrees and capture reflection for location calculation.

The emerging of LiDAR technique was dated back to 1960's when laser scanner became available for spatial survey. LiDAR emits light beams (laser) towards the surface of objects and calculating its backward time to provide spatial measurement. Once the light hit an object, it will reflect back to the laser scanner, then the total time can be calculated. Then, the total distance is calculated by using the velocity of light to get the time of flight. In reality, LiDAR will emit hundreds of thousands of pulses per seconds. This laser beam will hit objects in emitting direction and then return the LiDAR sensor. Then the result of received laser can be calculated to a "point-cloud" object, containing tens of thoughts points denoting a position on an object.

LiDAR is widely used for remote sensing that will measure surface of earth, ground cover, topography and other geo-information data. People usually refer point-cloud objects generated by LiDAR as a "digital twins"[12] to illustrate the accuracy for spatial information it contains. The key advantage of LiDAR sensing is its positional or spatial accuracy, including both absolute and relative. This allows any further modeling and analysis based on spatial information can be done within or across point-cloud objects.

**(2) Learning-based Point Cloud Methods**

[12] Tao, Fei, He Zhang, Ang Liu, and A. Y. C Nee. "Digital Twin in Industry: State-of-the-Art." *IEEE transactions on industrial informatics* 15, no. 4 (2019): 2405–2415.

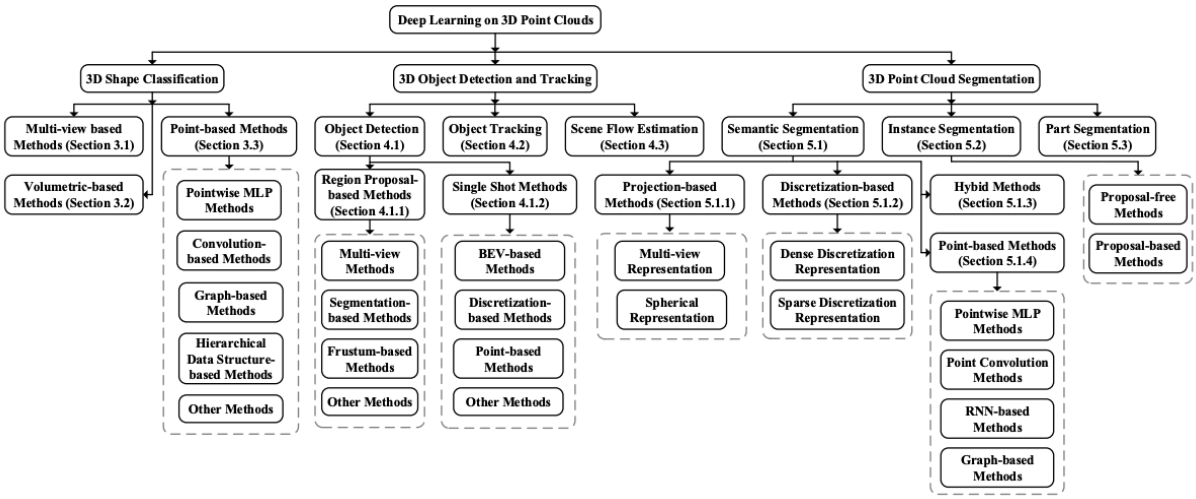



**Figure 9. A taxonomy of deep learning methods for 3D point clouds.**
Image Source: Guo, Yulan, Hanyun Wang, Qingyong Hu, Hao Liu, Li Liu, and Mohammed Bennamoun. "Deep learning for 3d point clouds: A survey." IEEE transactions on pattern analysis and machine intelligence 43, no. 12 (2020): 4338-4364.

The application of point cloud are ranging in robotics[13], construction[14], landscape[15], geography[16] and so on. The wide range of point cloud using is closely associated with adequate analysis tools and techniques developed with math and computer science. It includes conventional analysis techniques such as filtering, nearest neighbor search, model fitting, classification and segmentation, and feature extraction[17]. The conventional methods focusing on computing the spatial relationship of points and use those information to further filter, extract and cluster points for analysis.

In addition to the conventional techniques, a serial of deep learning-based point cloud techniques also emerged. “Deep-learning methods are representation-learning methods with multiple levels of

---

[13] Pomerleau, François, Francis Colas, and Roland Siegwart. "A review of point cloud registration algorithms for mobile robotics." *Foundations and Trends® in Robotics* 4.1 (2015): 1-104.

[14] Wang, Qian, and Min-Koo Kim. "Applications of 3D point cloud data in the construction industry: A fifteen-year review from 2004 to 2018." *Advanced Engineering Informatics* 39 (2019): 306-319.

[15] Urech, Philipp RW, et al. "Point cloud modeling as a bridge between landscape design and planning." *Landscape and Urban Planning* 203 (2020): 103903.

[16] Svalec, Marian, Lubos Takac, and Michal Zabovsky. "Performance enhancing of storage system for point cloud geographic data." *2015 10th System of Systems Engineering Conference (SoSE)*. IEEE, 2015.

[17] Liu, Songbin. *3D Point Cloud Analysis : Traditional, Deep Learning, and Explainable Machine Learning Methods*. Cham, Switzerland: Springer, 2021. Print.

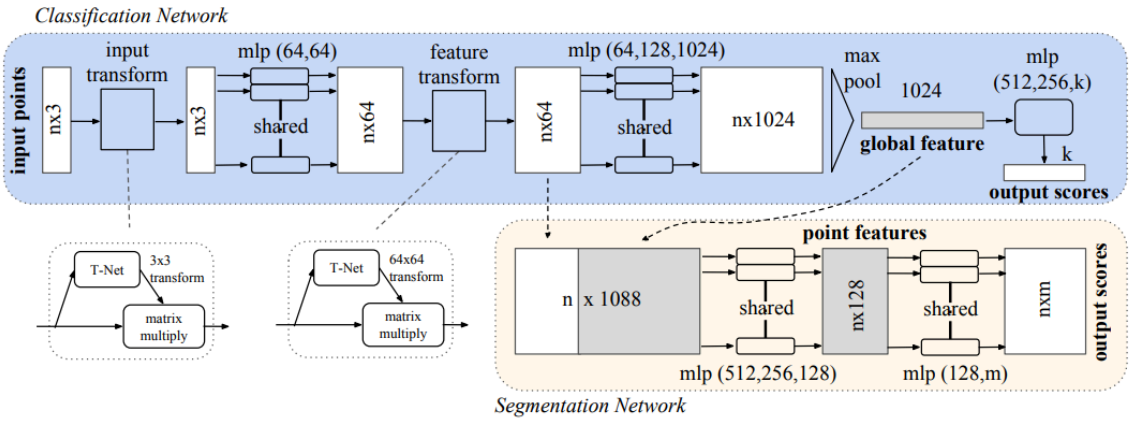



**Figure 10. PointNet Architecture**

Image Source: Qi, Charles R., Hao Su, Kaichun Mo, and Leonidas 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*, pp. 652-660. 2017.

representation, obtained by composing simple but non-linear modules that each transform the representation at one level (starting with the raw input) into a representation at a higher, slightly more abstract level"[18]. People usually refer any machine learning techniques involving multi-layer neuron network as the base of deep learning. Then, other techniques such as optimization, regularization and data representation integrate with the network as training and evaluation processes. Usually, deep learning is more expensive than machine learning, requiring more computation resources and larger dataset. However, it also shows advantages in modeling complex problems and better predict results.

### (3) Point Cloud Classification

Within point-cloud fields, PointNet[19] and PointNet++[20] are two main model architectures widely used to process point cloud data for embedding, classification and detection.

The task of 3D shape classification usually start with learning embedding features for each local points and then consolidate a global embedding with aggregation techniques. Classification is performed by

[18] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." *nature* 521, no. 7553 (2015): 436-444.

[19] Qi, Charles R., et al. "Pointnet: Deep learning on point sets for 3d classification and segmentation." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

[20] Qi, Charles Ruizhongtai, et al. "Pointnet++: Deep hierarchical feature learning on point sets in a metric space." *Advances in neural information processing systems* 30 (2017).

feeding global embedding features with MLPs. The challenge for point-cloud related deep learning methods is that we cannot directly use those techniques for 2D image for 3D point cloud[21]. The reason is that point cloud is unordered. The conventional CNN network cannot be used to extract features from point cloud objects. Within deep learning based point cloud methods, PointNet is the pioneer tackle point cloud objects. Utilizing multiple MLP(multi-layer perceptron) to extract point global features and then learn point-wise feature by pooling, PointNet is able to provide a learnable architecture that directly uses point cloud object. The limitation for PointNet is that it learns features independently for each point. This lead to the problem of the missing of internal relationship between points. For example, for all points within one object, there should be some internal features shared by all those points. To solve this issue, PointNet++ proposes another approach that let each point learn from its neighborhood points. This let the model can learn not only each point with its feature but the geometry and spatial relationship formed by points. The key improvement of PointNet++ is a three layer components: sampling layer, grouping layer and learning layer. The sampling layer sample given point cloud to small patches, and then grouping layer group points to cluster which will remain local geometric structures. Then the learning layer will learn local features and geometric structures from each point with its neighbors.

There are several advantages of PointNet++ over PointNet: 1) Hierarchical structure: PointNet++ uses a hierarchical neural network structure to process point clouds, which allows it to capture both local and global features. This is in contrast to PointNet, which only processes the points in a point cloud independently; 2) Multi-scale feature learning: PointNet++ uses multiple scales and levels of abstraction to process point clouds, which allows it to capture features at different scales. This is particularly useful for tasks such as object classification and segmentation, which often require the detection of both fine-grained and coarse-grained features; 3) Improved performance: In general, PointNet++ has been shown to outperform PointNet on a variety of tasks, including object classification and segmentation; and 4) Greater flexibility: PointNet++ is more flexible than PointNet in terms of the types of input it can accept. It can process point clouds of different sizes and densities, and can also handle input with additional features such as color or texture information.

[21] Guo, Yulan, Hanyun Wang, Qingyong Hu, Hao Liu, Li Liu, and Mohammed Bennamoun. "Deep learning for 3d point clouds: A survey." *IEEE transactions on pattern analysis and machine intelligence* 43, no. 12 (2020): 4338-4364.

Given above reasons, this work takes PointNet++ as backbone for the implementation of all point-cloud methods.

### (4) Point Cloud People Detection

In addition to 3D classification, 3D detection is also a very common task for point cloud object. The detection task can be summarized as assigning 3d bounding boxes to each detected object in each point cloud scene. Artificial intelligence and computer vision have experienced exponential growth in popularity and study over the past few years, and a variety of related applications have been steadily incorporated into daily life. This is the case with autonomous driving, a cutting-edge technology that uses a variety of sensors, including cameras, radars, and GPS, among others, to assess its environment and be able to respond appropriately in each circumstance. For the safety of drivers and pedestrians, it is essential to detect road environments, thus the technologies used in these systems must be accurate, durable, dependable, and independent of outside elements like low light levels. In addition to autonomous driving, detection techniques are also widely used for security, surveying and monitoring.

The photometric data of the RGB images is initially used for people detection. However, depending only on RGB data can result in system failures because this data could not be sufficient, particularly in bad weather conditions. For example, both dark environment and strong sunlight will cause RGB-camera based detection have very high error rate. To overcome these constraints, some methods choose to combine RGB cameras with ultrasonic sensors and radars, while others combine RGB data with the depth (RGB + D) provided by Time Of Flight (TOF)[22] sensors to enhance detection performance[23]. LiDAR sensor is a popular TOF sensor that measures the distance between the sensor and its surroundings by producing laser pulses.

RGB data is essential for comprehending our surroundings. The most often used methods for solving pedestrian detection issues are image-based methods. Although they work well overall, their detections are inconsistent when the images are blurry. To solve these issues, a lot of researchers integrate video

---

[22] Foix, Sergi, Guillem Alenya, and Carme Torras. "Lock-in time-of-flight (ToF) cameras: A survey." *IEEE Sensors Journal* 11, no. 9 (2011): 1917-1926.

[23] Spinello, Luciano, and Kai O. Arras. "People detection in RGB-D data." In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3838-3843. IEEE, 2011.

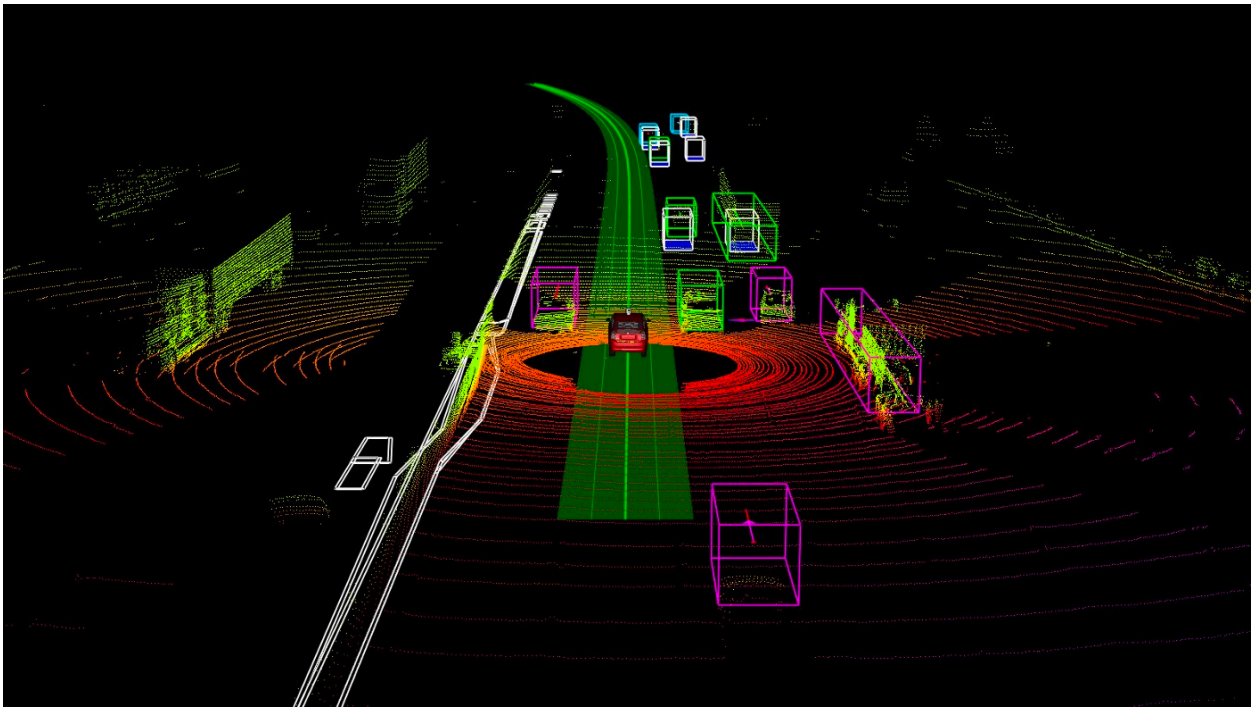


**Figure 11. Point Cloud Detection**
Image Source: https://medium.com/@jhkoh/object-detection-with-lidar-point-cloud-algorithm-94a241fd3f49

cameras and LIDARs. There aren't many LIDAR-based methods, though, that just rely on point clouds' 3D geometric data.

Navarro-Serment et al[24]. present a method for detecting and tracking pedestrians using motion and geometric features. They initially project data subsets onto the 2D plane to perform object detection in order to lessen the computational strain of analyzing the complete point cloud. Following that, a motion score and a set of geometric features in the related 3D point clouds are computed using the detections. As a result, the technique is dependent on 2D domain detections. This suggests mapping the original point cloud onto a 2D plane, which may result in the loss of 3D geometric data.

[24] Navarro-Serment, Luis E., Christoph Mertz, and Martial Hebert. "Pedestrian detection and tracking using three-dimensional ladar data." *The International Journal of Robotics Research* 29, no. 12 (2010): 1516-1528.

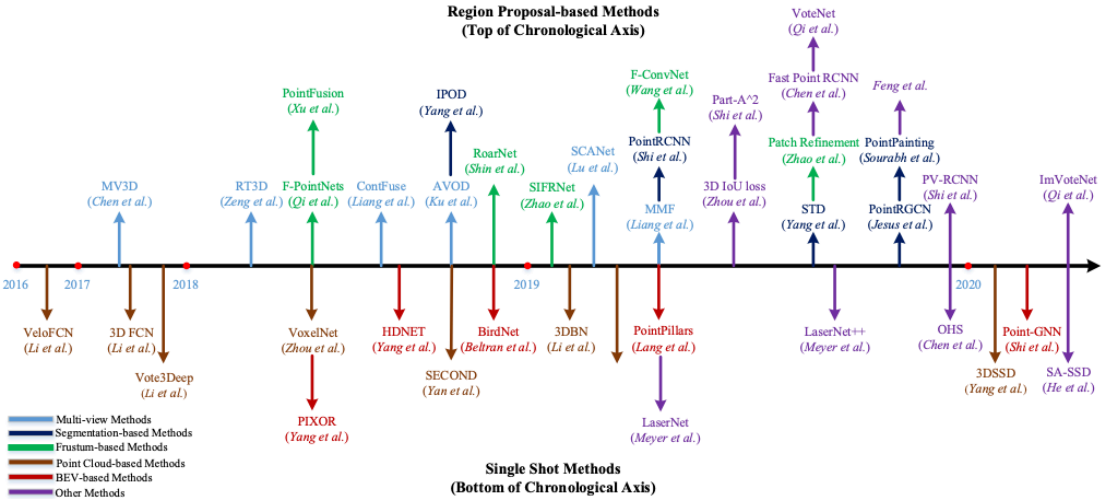



**Figure 12. Chronological overview of the most relevant deep learning-based 3D object detection methods**
Image Source: Guo, Yulan, Hanyun Wang, Qingyong Hu, Hao Liu, Li Liu, and Mohammed Bennamoun. "Deep learning for 3d point clouds: A survey." IEEE transactions on pattern analysis and machine intelligence 43, no. 12 (2020): 4338-4364.

The approaches of Tang et al.[25] and Lin et al.[26] are similar in that they both utilize a Support Vector Machine (SVM) with extracted 2D and 3D features. However, Lin et al. develop a Convolutional Neural Network (CNN) to extract 3D features, whereas features in Tang et al's work are completely manually created. The 3D point cloud data must be discretized into a 3D voxel representation in order to train the CNN, despite the fact that the latter enables the acquisition of pertinent deep features.

Based on PointNet++, the aforementioned hierarchical neural network that extracts deep features from the 3D geometric data of a point cloud, Lorente et al.[27] offer a pedestrian detector in 3D point clouds. Local features are extracted from small neighborhoods, capturing fine-grained geometric structures. These local features are then grouped into larger units and processed to produce higher level features. This process is repeated until the local and global features of the entire point set are obtained. This hierarchical approach

---

[25] Tang, Hsueh-Ling, Shih-Che Chien, Wen-Huang Cheng, Yung-Yao Chen, and Kai-Lung Hua. "Multi-cue pedestrian detection from 3D point cloud data." In *2017 IEEE international conference on multimedia and expo (ICME)*, pp. 1279-1284. IEEE, 2017.

[26] Lin, Tzu-Chieh, Daniel Stanley Tan, Hsueh-Ling Tang, Shih-Che Chien, Feng-chia Chang, Yung-Yao Chen, Wen-Huang Cheng, and Kai-Lung Hua. "Pedestrian detection from lidar data via cooperative deep and hand-crafted features." In *2018 25th IEEE international conference on image processing (ICIP)*, pp. 1922-1926. IEEE, 2018.

[27] Lorente, Òscar, Josep R. Casas, Santiago Royo, and Ivan Caminal. "Pedestrian Detection in 3D Point Clouds using Deep Neural Networks." *arXiv preprint arXiv:2105.01151* (2021).

allows PointNet++ to respect the 3D geometric information of the point cloud as it processes the data. By extracting features at multiple scales and levels of abstraction, PointNet++ is able to capture both the fine details and the overall structure of the point cloud.

Compared with those methods of Navarro-Serment et al., Tang et al., and Lin et al., the deep-learning methods proposed by Lorente et al. has a few advantages: 1) Better performance: Deep learning-based approaches have generally been found to outperform SVMs on a variety of tasks, including object detection and recognition. This is because deep learning models are able to learn complex, non-linear relationships in the data, whereas SVMs are limited to linear decision boundaries; 2) Greater flexibility: Deep learning models are more flexible than SVMs, as they can process input data of various types (e.g. images, point clouds, audio, etc.) and can also handle input with additional features such as color or texture information. SVMs, on the other hand, are limited to processing numerical data; 3) Fewer assumptions: Deep learning models make fewer assumptions about the data than SVMs, which makes them more robust and able to handle a wider range of input data; 4) Better generalization: Deep learning models are able to generalize better to new data, as they are able to learn a wide range of features and patterns in the data. SVMs, on the other hand, rely on the kernel function to determine the similarity between data points, which can be less robust; 5) Automated feature extraction: Deep learning models are able to learn useful features automatically from the data, without the need for manual feature engineering. This can save time and effort, and can also lead to better performance as the model is able to learn more relevant and nuanced features from the data. SVMs, on the other hand, require manual feature engineering.

Above methods are used for static point cloud prediction and classification. However, the LiDAR provides a different dimension for point clouds which is in the form of streaming. This means that instead of having multiple static scenes that are independent, each frames of point-cloud object is related to the previous and the afterwards frames. The challenge of streaming point cloud is compression. ”One hour of point cloud data from the Velodyne HDL-64 sensor mentioned above can result in over 100GB of data, which is too large to realistically share or store using currently available technology. Thus, developing methods of compressing this data has become an indispensable task”[28]. In order to compress point cloud efficiently, multiple approaches has been developed to reduce both space and temporal redundancy. Using

[28] Tu, Chenxi, Eijiro Takeuchi, Alexander Carballo, and Kazuya Takeda. "Real-time streaming point cloud compression for 3d lidar sensor using u-net." *IEEE Access* 7 (2019): 113616-113625.

either 2D tree[29] or 2D matrix[30] is able to reformat unordered point cloud and reduce spatial redundancy. For temporal redundancy, an additional U-Net is widely used to "predict the enclosed frames between each pair of reference/key frames to reduce the temporal redundancy of streaming point cloud data"[31].

### (5) Point Cloud Upsampling

The output of 3D scanning is often the point cloud. Point clouds are becoming more and more common because they are a useful tool for processing 3D geometric objects and a concise way to represent 3D data[32]. However, 3D sensors like LiDAR typically create sparse, noisy, and non-uniform point clouds due to hardware constraints, particularly for small objects or objects far from the camera. This has also been demonstrated in a number of open benchmark datasets, including KITTI. Although 3D sensing technology has advanced significantly in recent years, obtaining a point cloud with high density and comprehensive information is still a costly and time-consuming task. Point cloud applications in areas like 3D form classification, 3D object identification, and 3D object segmentation are impacted by their sparsity and noise. It is obvious that raw point cloud data has to be modified.

Point cloud upsampling refers to the process of generating a dense and uniform point cloud from a starting point that is sparse, noisy, and non-uniform[33]. This issue is quite difficult to solve in the current environment. Point cloud data are represented by a collection of disorganized data points, unlike the image data representation in a computer, which typically encodes the spatial relationship between pixels. As a result, upsampling point clouds is very challenging, and there aren't many works that address it.

---

[29] Gadelha, Matheus, Rui Wang, and Subhransu Maji. "Multiresolution tree networks for 3d point cloud processing." In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 103-118. 2018.

[30] Tu, Chenxi, Eijiro Takeuchi, Chiyomi Miyajima, and Kazuya Takeda. "Compressing continuous point cloud data using image compression methods." In *2016 IEEE 19th international conference on intelligent transportation systems (ITSC)*, pp. 1712-1719. IEEE, 2016.

[31] Tu, Chenxi, Eijiro Takeuchi, Alexander Carballo, and Kazuya Takeda. "Real-time streaming point cloud compression for 3d lidar sensor using u-net." *IEEE Access* 7 (2019): 113616-113625.

[32] Alexa, Marc, Johannes Behr, Daniel Cohen-Or, Shachar Fleishman, David Levin, and Claudio T. Silva. "Computing and rendering point set surfaces." *IEEE Transactions on visualization and computer graphics* 9, no. 1 (2003): 3-15.

[33] Zhang, Yan, Wenhan Zhao, Bo Sun, Ying Zhang, and Wen Wen. "Point Cloud Upsampling Algorithm: A Systematic Review." *Algorithms* 15, no. 4 (2022): 124.

Deep learning has produced new methods for handling point cloud data, including models like PointNet[34]. As a result, academics have started to pay more attention to point cloud upsampling jobs.

The initial point cloud upsampling algorithm was proposed by Alexa et al. in 2003[35], which is a **optimization-based point cloud upsampling** techniques. On the moving least squares (MLS) surface, they upsample a point by interpolating points as the vertices of a Voronoi diagram. It creates a Voronoi diagram using any three points on the plane. The center of a circle that touches three or more points without having any points within is the vertex of a voronoi diagram. The center of the circle with the biggest radius is chosen and projected onto the MLS surface after generating the Voronoi diagram. There is an upsampled point as a result. Until the greatest circle's radius falls below the predetermined threshold, this process is repeated. To boost calculation efficiency, the local approximation method is used. After that, a series of Optimization-Based Point Cloud Upsampling methods were developed[36] [37] [38] [39] [40] [41]. Overall, optimization-based approaches have substantial limits and are not data driven, despite the fact that they can upsample point clouds to some amount. They are predicated on priors like normal estimation and the smoothness hypothesis for surfaces with fewer characteristics. The preservation of multi-scale structures is another challenge for these techniques.

**Learning-based point cloud upsampling methods**, which use deep learning techniques, have gained popularity among researchers in recent years. These methods have several advantages over optimization-

---

[34] Same as 21.

[35] Same as 33.

[36] Lipman, Yaron, Daniel Cohen-Or, David Levin, and Hillel Tal-Ezer. "Parameterization-free projection for geometry reconstruction." *ACM Transactions on Graphics (TOG)* 26, no. 3 (2007): 22-es.

[37] Huang, Hui, Dan Li, Hao Zhang, Uri Ascher, and Daniel Cohen-Or. "Consolidation of unorganized point clouds for surface reconstruction." *ACM transactions on graphics (TOG)* 28, no. 5 (2009): 1-7.

[38] Preiner, Reinhold, Oliver Mattausch, Murat Arikan, Renato Pajarola, and Michael Wimmer. "Continuous projection for fast L1 reconstruction." *ACM Trans. Graph.* 33, no. 4 (2014): 47-1.

[39] Huang, Hui, Shihao Wu, Minglun Gong, Daniel Cohen-Or, Uri Ascher, and Hao Zhang. "Edge-aware point set resampling." *ACM transactions on graphics (TOG)* 32, no. 1 (2013): 1-12.

[40] Wu, Shihao, Hui Huang, Minglun Gong, Matthias Zwicker, and Daniel Cohen-Or. "Deep points consolidation." *ACM Transactions on Graphics (ToG)* 34, no. 6 (2015): 1-13.

[41] Dinesh, Chinthaka, Gene Cheung, and Ivan V. Bajić. "3D point cloud super-resolution via graph total variation on surface normals." In *2019 IEEE International Conference on Image Processing (ICIP)*, pp. 4390-4394. IEEE, 2019.

based approaches. One advantage is that uneven point clouds can now be used directly for training thanks to network models like PointNet[42] and PointNet++[43]. As a result, deep learning-based point cloud upsampling models have produced a range of promising results, and the application of deep learning to point clouds has become a highly active area of research.

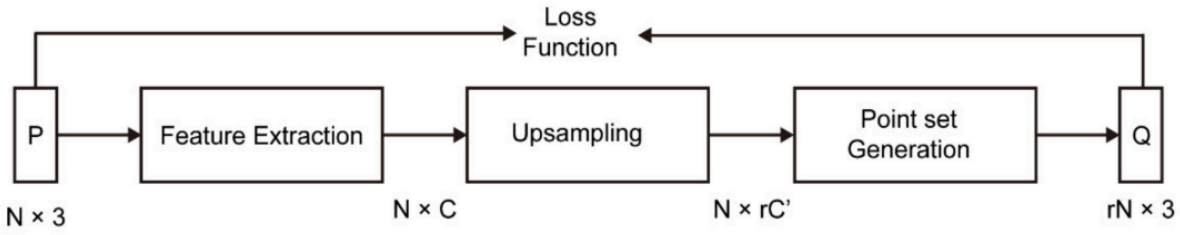



**Fig 13. Schematic diagram of a point cloud upsampling model based on deep learning.**
C and C' are the features of the point, r is the upsampling rate. P is input, Q is output, N is number of points.
Image Source: same as 34.

The first deep learning model for point cloud upsampling, PU-Net[44], was introduced by Yu et al[45], in which hierarchical feature learning and multi-level feature aggregation are employed as feature learning methodologies. As the frontal portion of the network for hierarchical feature learning, PU-Net employs the hierarchical feature learning technique suggested in PointNet++. PU-Net especially employs a somewhat short grouping radius in each layer to get more of the local context. In order to upsample the downsampled point features in hierarchical feature learning and recover the original point features, PU-Net first applies the interpolation approach in PointNet++. Convolution is then used to bring the dimensionality of the interpolated features at various levels down to a single dimension. Multi-branch feature expansion component is used in PU-Net for feature expansion.

---

[42] Same as 21.

[43] Same as 22.

[44] Yu, Lequan, et al. "Pu-net: Point cloud upsampling network." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.

[45] Yu, Lequan, Xianzhi Li, Chi-Wing Fu, Daniel Cohen-Or, and Pheng-Ann Heng. "Pu-net: Point cloud upsampling network." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2790-2799. 2018.

Another method that is widely used is PU-GAN[46]. Using dynamic graph convolution as their model, Wang et al.[47] proposed a multi-step point cloud upsampling network (MPU) to define local neighborhoods in feature space. On the basis of feature similarity, a k-nearest neighbors (kNN) search is used to extract point features from nearby neighborhoods. The long-range and non-local information may be obtained using this approach without the need for point set subsampling. The feature extraction unit, in more detail, is made up of a dense series of blocks where the MPU transforms the input into a fixed number of features, groups the features using a feature-based kNN, then refines each grouped feature using a closely interconnected multilayer perceptron (MLP) chain, before obtaining point features via maxpooling. Dense connections are added by the MPU both inside and between blocks. Since this connection style allows for explicit information reuse, the reconstruction accuracy is increased while the model size is significantly decreased. An up-and-down sampling mechanism is introduced in PU-GAN. It expands the point feature by upsampling it, then downsamples the enlarged feature to calculate the difference between the up- and down-sampled features. To self-correct the extended feature, the difference is upsampled and added to the first step. In order to create a unique 2D vector for each feature-map and attach it to each point feature, PU-GAN first replicates the input features and uses the 2D grid technique in FoldingNet[48]. The output upsampled features are then produced by a self-attention unit and a collection of MLPs. A group of MLPs and a reshape operation make up the down-feature operator.

In summary, learning-based point cloud upsampling outperformes optimization-based methods for greater performance, fewer assumption and better generalization. Within learning-based methods, Pu-GAN has better potential than Pu-Net as the adversarial training process allows the generator network to learn complex, non-linear relationships in the data and produce more realistic and accurate results.

## 2.3. Point Cloud Visualization

---

[46] Li, Ruihui, Xianzhi Li, Chi-Wing Fu, Daniel Cohen-Or, and Pheng-Ann Heng. "Pu-gan: a point cloud upsampling adversarial network." In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 7203-7212. 2019.

[47] Yifan, Wang, Shihao Wu, Hui Huang, Daniel Cohen-Or, and Olga Sorkine-Hornung. "Patch-based progressive 3d point set upsampling." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5958-5967. 2019.

[48] Yang, Yaoqing, Chen Feng, Yiru Shen, and Dong Tian. "Foldingnet: Point cloud auto-encoder via deep grid deformation." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 206-215. 2018.



**Figure 14. Screenshot of point cloud visualization in ESRI**
Image Source: https://mediaspace.esri.com/media/t/1_5awtw0ot

Since point cloud is widely used for survey, analysis and simulation, visualization tools and libraries are also developed accordingly. However, point cloud visualization is a complicated problem. Fabio points out that "wide range of applications implies that the data can have quite different properties that must be considered in the solution of the surface interpolation problem"[49]. Some methods developed meshing techniques to create continuous triangle surface for point cloud objects and use them for visualization[50]. And some other research built visualization methods with point clouds[51]. This research will focus on the second direction. Among all of them, QSplat[52] was the first attempt to render large point sample model.

---

[49] Remondino, Fabio. "From point cloud to surface: the modeling and visualization problem." *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 34 (2003).

[50] Mencl, R. "Reconstruction of surfaces from unorganized 3D points clouds (Doctoral dissertation, PhD Thesis, Dortmund University, Germany)." (2001).

[51] Grossman, Jeffrey P., and William J. Dally. "Point sample rendering." *Eurographics Workshop on Rendering Techniques*. Springer, Vienna, 1998.

[52] Rusinkiewicz, Szymon, and Marc Levoy. "QSplat: A multiresolution point rendering system for large meshes." *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*. 2000.

Following with this, Layered Point Clouds[53] (LPC) was developed mainly to deal with the performance of GPU. Instant Points is the first one that doesn't require pre-processing steps[54]. In addition to those academic experiments, several commercial solution are also available such as Bentley[55], ESRI[56], Meshlab[57], PCLVisualizer[58], and Euclideon[59].

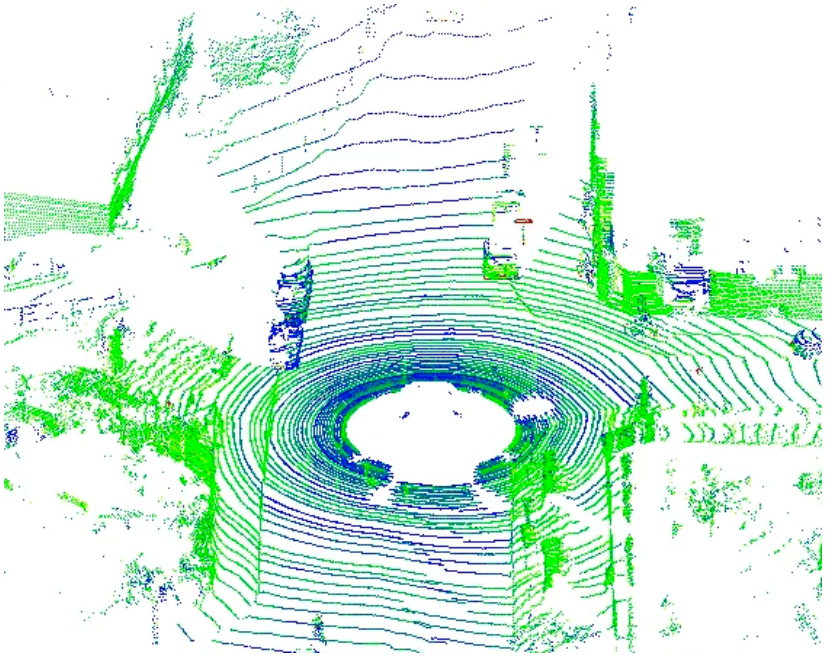


**Figure 15. Typical LiDAR visualization libraries used in research**
Image Source: https://www.researchgate.net/figure/3D-point-cloud-from-data-captured-by-a-LiDAR-sensor_fig1_343969326

---

[53] Gobbetti, Enrico, and Fabio Marton. "Layered point clouds: a simple and efficient multiresolution structure for distributing and rendering gigantic point-sampled models." *Computers & Graphics* 28.6 (2004): 815-826.

[54] Wimmer, Michael, and Claus Scheiblauer. "Instant Points: Fast Rendering of Unprocessed Point Clouds." *PBG@ SIGGRAPH*. 2006.

[55] Bentley Pointools. http://www.bentley.com/en-US/Promo/Pointools/pointools.htm (Accessed 2022-11-08).

[56] ESRI ArcGIS. https://pro.arcgis.com/en/pro-app/latest/help/mapping/layer-properties/point-cloud-scene-layer-in-arcgis-pro.htm (Accessed 2022-11-08).

[57] Meshlab. https://blogs.gre.ac.uk/designsupport/3d-realisation/laser-scanning/meshlab-point-cloud-to-mesh/ (Accessed 2022-11-08).

[58] PCL CloudViewer. https://pcl.readthedocs.io/projects/tutorials/en/latest/cloud_viewer.html (Accessed 2022-11-08).

[59] Euclideon. http://euclideon.com/ (Accessed 2022-11-08).

A common pattern for those visualization tools and their development is the focus on technical challenges. Most of new products and solutions came with new solutions for higher efficient GPU using, memory footprint, data structure and data processing. However, the experience of user is always ignored in development of those tools.

There are some researchers taking the user experience as their research angles for point cloud. Alexiou et al conducted research to evaluate the visual quality of point cloud contents with the medium of virtual reality[60]. In these research, a subjective evaluation method was established to evaluate the quality of point cloud[61]. This evaluation method asks participants to give scores for visualization results, and then use those results to evaluate visualization quality. Su et al conducted similar research but with a focus on perceptual quality assessment[62]. This is done by well-designed questions and grading options. However, still very few research have been done for the comprehensive analysis of user experience with point cloud visualization tools. More importantly, very less research outcome have been incorporated into the development of new visualization tools.

## 2.4. Recognizability

For the majority of people, recognizing the text on this page, the food on their plates, or the person they have just met are simple tasks that don't require much thought. Despite the enormous variation in

---

[60] Alexiou, Evangelos, Nanyang Yang, and Touradj Ebrahimi. "PointXR: A toolbox for visualization and subjective evaluation of point clouds in virtual reality." *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2020.

[61] da Silva Cruz, Luis A., et al. "Point cloud quality evaluation: Towards a definition for test conditions." *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2019.

[62] Su, Honglei, et al. "Perceptual quality assessment of 3D point clouds." *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019.

appearance that each object produces on our eyes[63], we are able to detect and classify objects from among tens of thousands of possibilities [64]with ease and within a fraction of a second[65].

How does the brain carry out this function? What sort of data does the visual system use to create descriptions of collections of object attributes that capture the invariant characteristics of objects? How are these descriptions saved, and how is the viewed item responsible for activating them? Whether they are particular to a behavior or a mental function like learning, planning, or reasoning, object representations can either be universal or specialized.

Scientists from a range of fields, including cognitive psychology, neurobiology, neuropsychology, computing, and engineering, have traditionally investigated these problems. Some of these studies have demonstrated the lack of a true multifunctional general recognition system[66]. Instead, several representations of an item are created during the biological identification process, each tailored to the changes needed for either perception or action. According to the reviewed literature, distinct representations formed in various neural sites or by various interconnectivity patterns are necessary for the recognition of prototypical members of an object category, the encoding of dynamic and plastic transformations of objects or object parts, the identification of individual members of a homogeneous object class, and the planning of movements that are regularly made when interacting with familiar objects.

Usually, we define several categories for object and group them together based on similarity. Similarity is widely also defined regarding to colors, shapes, functions, and even social meanings. Researchers observed that common things are categorized similarly in different society[67]. Though they are named differently according to languages, the subset of those things shows very consistent patterns. Though

---

[63] Logothetis, Nikos K., and David L. Sheinberg. "Visual object recognition." *Annual review of neuroscience* 19, no. 1 (1996): 577-621.

[64] Biederman, Irving. "Recognition-by-components: a theory of human image understanding." *Psychological review* 94, no. 2 (1987): 115.

[65] Thorpe, Simon, Denis Fize, and Catherine Marlot. "Speed of processing in the human visual system." *nature* 381, no. 6582 (1996): 520-522.

[66] Logothetis, Nikos K., and David L. Sheinberg. "Visual object recognition." *Annual review of neuroscience* 19, no. 1 (1996): 577-621.

[67] Brown, Roger. "Words and things." (1958).

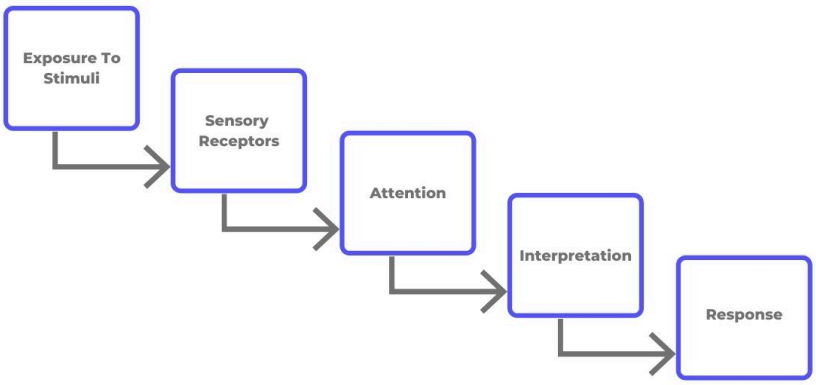



**Figure 16. The recognition process.**
Image Source: https://blog.ambio-media.com/posts/the-perceptual-process

Brown explained it with reason of "community's nonlinguistic purposes"[68]. However, the concept that categories could truly represent more than simply language constructs has been demonstrated by other academics[69]. They demonstrated that some combinations of qualities are more likely to occur together in the actual world than others, and that human conceptual categories are based on this perceptual foundation. In contrast to pairings like "feathers" and "wheels," frequent co-concurrence such as "feathers" and "wings. Bundles of these often recurring characteristics serve as the foundation for a natural categorization of items.

Their work promoted the idea that there are "intrinsically different things"[70] in the universe as well as a taxonomy of objects, where categories are connected by class inclusion. These classifications, sometimes known as "cues," are organic collections of stimuli with various perceptible features. Valid category predictors are cues that occur frequently in one category but infrequently in all others. Therefore, humans are able to categorize things in the world in a systematic manner using groups of attributes that occur naturally.

---

[68] Same as 72.

[69] Rosch, Eleanor, Carolyn B. Mervis, Wayne D. Gray, David M. Johnson, and Penny Boyes-Braem. "Basic objects in natural categories." *Cognitive psychology* 8, no. 3 (1976): 382-439.

[70] Same as 74.

To be more specific, the recognition process, or so called as perceptual process is a series of steps for human to perceive and identify things appearing towards their attention.

The first step in the perception process, **selection**, is when we transform environmental inputs into memorable experiences. As a result of the continual barrage of information we receive in everyday life, any of the following stimuli may be presented to us in a split second: the words we are hearing, an accident witness, or the sound of a ticking clock, to name a few. Our universe is all-encompassing, thus there are innumerable impulses coming at our sensory organs at once that need to be processed. However, we are unable to take in all the information that is accessible to us since doing so would generate chaos and information overload. According to Singer, "We experience everything in the world not as it is—but only as the world comes to us through our sensory receptors,[71]" we selectively take in just a portion of the information from the environment. This selection step of perception has drawn the attention of many scholars. There is an analogy between selection and mapping. While making a map, it is impossible to put everything into a single map. Therefore, selection and abstraction are necessary to capture important information and filter unwanted. Same to mapping, human's perception also select stimuli. A stimulus in physiology is a perceptible alteration in the internal or external surroundings of an organism's physical or chemical composition. Sensitivity is the capacity of an organism or organ to recognize external stimuli and to respond appropriately to them.

**Organization** is the second level of perception. After gathering data from the outside world, we must arrange it somehow by identifying some significant patterns. Some academics sometimes refer to this stage of organization as classification since it involves grouping items or individuals into groups. The social and physical events and things we come into contact with at this level of perception will already have a shape, color, texture, size, etc. For instance, when asked to define what a human being is, some people may do it in terms of skin tone, while others may do so in terms of race or nationality. When we imagine what our university library is like when our eyes are closed, we see a structured space with both an interior and external structure. Because of the function of organization, humans are able to structure and bring coherence to our general knowledge about people and the social world by providing typical

[71] Singer, Marshall R. *Intercultural communication: A perceptual approach*. Prentice Hall, 1987.

patterns of behavior and the range of likely variation between types of people and their defining behaviors and characteristics[72].

The process of giving the selected stimuli meaning is referred to as **interpretation**, which is the third stage of perception. We attempt to interpret the ordered and consistent patterns that have been formed from the chosen stimuli by giving them meanings. However, different individuals may interpret the same signal in various ways. For instance, the arrival of a police officer at the scene of the crime might be perceived in different ways. While the victim may find it comforting and relieving, the criminal will undoubtedly be terrified by it. Interpretation are always conducted differently within individuals. This is due to the cultural background each individual holds[73]. Nearly everyone can agree on the objective meaning of a physical thing or event, but each unique person's interpretation of that object or event depends on their own experiences and cultural upbringing. People with various backgrounds and experiences will interpret the same stimuli differently, leading to the appearance of perceptual variety.

---

[72] Qiong, O. U. "A brief introduction to perception." *Studies in literature and language* 15, no. 4 (2017): 18-28.

[73] Samovar, Larry A., Richard E. Porter, Edwin R. McDaniel, and Carolyn Sexton Roy. *Communication between cultures*. Cengage Learning, 2016.

# 3. Hypothesis and Research Questions

## 3.1. Hypothesis

This research work is built upon a hypothesis that by incorporating stat-of-art point-cloud methods including detection and upsampling, users can show better recognizability than the original one.

To test this idea, original visualization results will be utilized as the baseline against which different visualization possibilities created by new point-cloud approaches. Recognizability will be examined in terms of object, semantic, and geographical recognition. And these aspects can be tested by the accuracy with which users are asked to accomplish specified tasks.

## 3.2. Research Questions

This research attempts to answer the question: "**Can we leverage the recognizability of point-cloud visualization for further engagement**"? This question can be further expanded to three sub questions:

**(1) How to visualize point-cloud?**

To answer this question, this research work dives into existing point-cloud augmentation methods. By incorporating point cloud objects with augmented techniques, we can provide additional assistance for users to recognize point cloud objects. In this work, two learning-based augmentation techniques are selected. Related dataset are prepared, and models are trained. However, if augmented techniques will have positive or negative impacts on recognizability cannot be known in advance. Then, the user-centered study is necessary for the next step.

**(2) How to leverage recognizability?**

To answer this question, a study of recognizability is done to further understand what is recognizability and how can be evaluated it. Based on the understanding of recognizability, this work creates a sub-category of recognizability regarding three aspects: object, semantic, and spatial. The reason these three aspects are selected is due to their close relationship with point-cloud objects and tasks. Then, several user study questions and tasks are designed to evaluate them with users. Through results, we are able to at least shed light on if recognizability is enhanced through the new visualization tools.

**(3) How to foster new engagement?**

Engagement refers to the ability of inviting more people to involve in point-cloud related tasks or work. In this work, it is presented by a series of design scenarios to imagine the using of point cloud in real life.

## 3.3. Research Goal

The goal of the work should be presented in two levels.

First, a demo of point cloud visualization libraries is completed to study how visualization techniques will impact people's recognizability and then engagement. This part of the work will contribute to the current study of point cloud visualization with an angle of measuring its recognizability. It will enrich the criteria used for point cloud visualization from merely efficiency and accuracy to more human-sensible metrics.

Second, a development workflow of visualization tools is presented that connects both technical development and user-centered research. This will contribute to the current work of visualization libraries development with a new perspective that invites users to the front end. Currently, the development of visualization libraries is completely technical-driven with less consideration for users. This work will shed a light on how to introduce user-related consideration into the development stages. The well-documented development process can also be used as a reference for any further development following this direction.

# 4. Methodologies

| Componenets | Methods | Requirment | Results |
|---|---|---|---|
| technical implementation | Evaluate with developed implementation and metrics | Technical implementation, metrics, data, and available comparison from previous research | Quantitative results (scores, error rate, etc.) |
| user study | Evaluate with users through testing | Available user interface, designed tasks and questions, volunteers for testing | Quantitative results (scores, reaction time, etc.) and qualitative results (summary, interview, etc.) |

**Table 1: Research Methodology**

This research aims to build an end-to-end research inquiry including technical demo and user research. Therefore, this work will extend methodologies in two parts: *technical implementation* and *user study*. For the technical implementation part, this work mainly builds on current research papers for implementation and evaluation. The approach is mainly quantitative and scientific driven, including building metrics to evaluate the results and compare them with results reported in papers. For user research, the user test methodology will be applied to engage with users and derive direct feedback that can be quantitatively and qualitatively evaluated with the work.

# 5. Technical Implementation

## 5.1. Framework

The overall technical implementation contains three parts: dataset, augmentation models, and user interface. First, datasets are prepared for the training and evaluation of augmentation models; second, two models are developed to generate augmented point-cloud object for visualization; third, a user interface is designed to display visualization result and conduct user test.

This framework is developed with intention for good extensibility. Since all components are relatively independent, the framework is capable to replace or add new components.

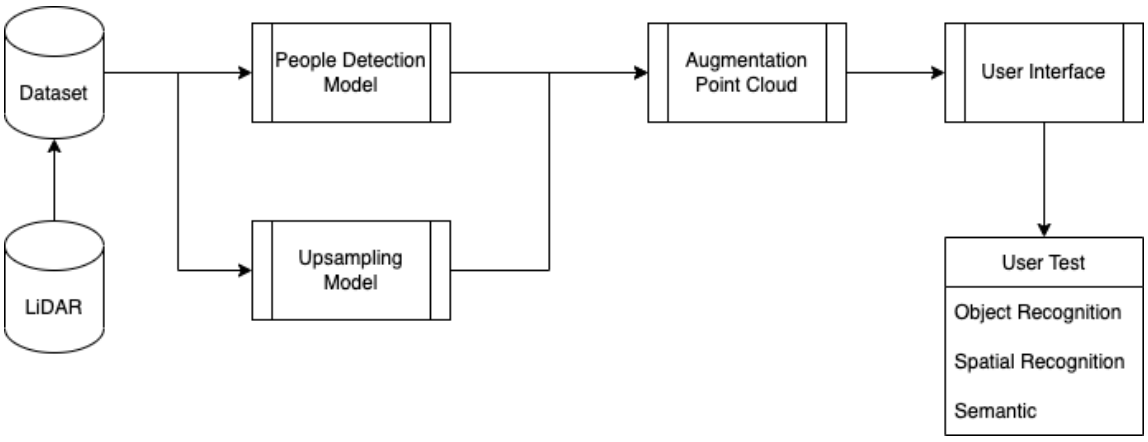



**Figure 17. Technical Framework**
Image Source: drawn by author

## 5.2. Dataset

Dataset is an important part of learning-based techniques. Since all machine learning models, it needs to train and evaluate data. An adequate amount of high-quality data can ensure the model learns adequate features from data. Usually, in supervised learning, each data will have a label associated with the data entity. This ensures that we can feed data to the model and let the model learn features based on given

labels. In most cases, the dataset will be split into two parts: training and evaluation. The training dataset will be for the training process. So that model will directly learn from all features of the training dataset. The training dataset will enable the model to fit all learnable parameters in the training process. The evaluation dataset is used for the evaluation process, which is the post-training stage. It provides opportunities to tune the hyperparameters of the model and to evaluate overfitting. A common pattern for overfitting is that the training result keeps increasing but the evaluation result is not changed.

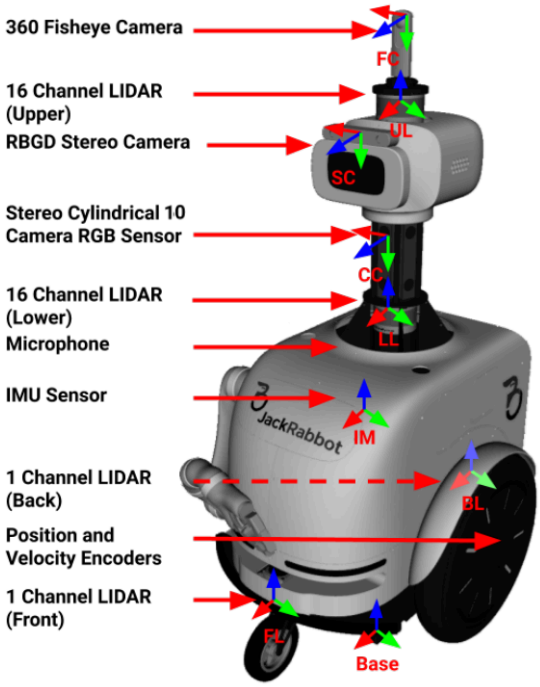



**Figure 18. JackRabbot**
Image Source: same as 45.

For this research, two dataset are used for two augmentation techniques.

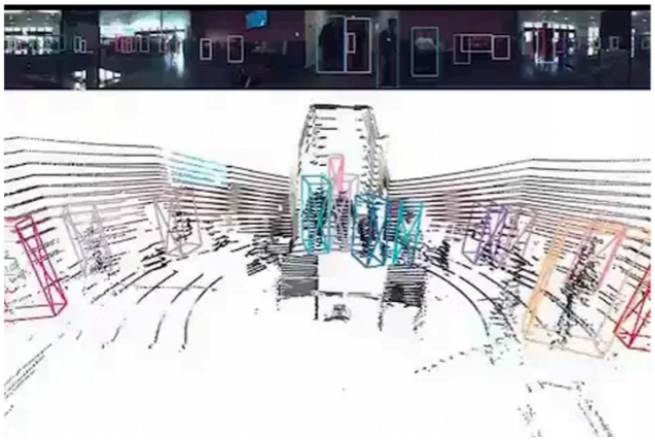
First, for people detection model, the JRDB dataset is used[74]. This dataset was captured by small robot JackRabbot that is "equipped with 4 LIDAR sensors, 3 cameras, 2 movement encoders, an IMU sensor, and a microphone"[75]. Due to the flexibility of the robot, this dataset includes 3d data with rich context including both indoor and outdoor environment.

[74] Martín-Martín, Roberto, et al. "Jrdb: A dataset and benchmark for visual perception for navigation in human environments." *arXiv preprint arXiv:1910.11792* (2019).

[75] Same as 28.

This research also looked into another popular dataset, which is KITTI[76]. However, this dataset of KITTI was captured by LiDAR mounted on the top of the car. Due to this limitation, most environments are road and detected objects are cars.

Compared with KITTI, the JRDB dataset is able to illustrate the human environment with more human-related activities. This provides opportunities for the model to learn the rich context for detection. Also, the environment itself will be much more complex, involving outdoor and indoor places such as plazas, auditoriums, offices, and classrooms. These features will be valuable to be captured in the model.



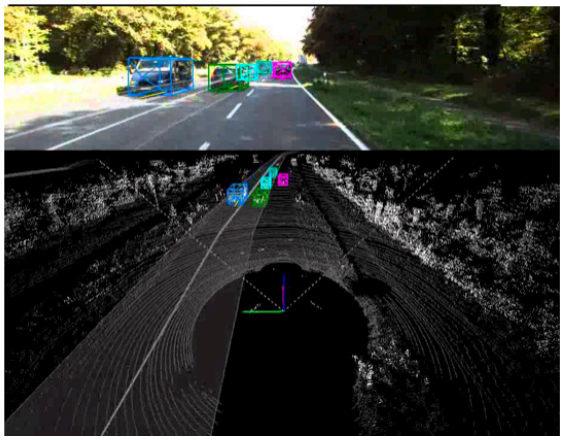


**Figure 19.(a) typical JRDB scene, captured in a indoor auditorium.**
Image Source: Same as 45
**Figure 19.(b) typical KITTI scene, captured on road.**
Image Source: Same as 47

The JRDB dataset contains over 60,000 annotated frames. Within each frame, there are several bounding box for human detection. Each bounding box is a 3D box illustrating a human's spatial location  In total, 1.8 million human bonding box annotations are available for training. In this dataset, all data are stored with following vector format: [x, y ,z, d]. Each vector data denote a point's spatial information. In here, "xyz" are the spatial location of a point in Cartesian system, and "d" is the density of the point. Each file contains multiple vectors as for all points for a frame. And the frame is captured per second.

## 5.3. People Detection Model

[76] Geiger, Andreas, Philip Lenz, Christoph Stiller, and Raquel Urtasun. "Vision meets robotics: The kitti dataset." *The International Journal of Robotics Research* 32, no. 11 (2013): 1231-1237.

The first augmentation technique this work employed is people detection. The goal of people detection is to detect people from given point cloud data and label these people with bounding boxes. This can be seen as a visual assistance for user to recognize content of the point cloud visualization.

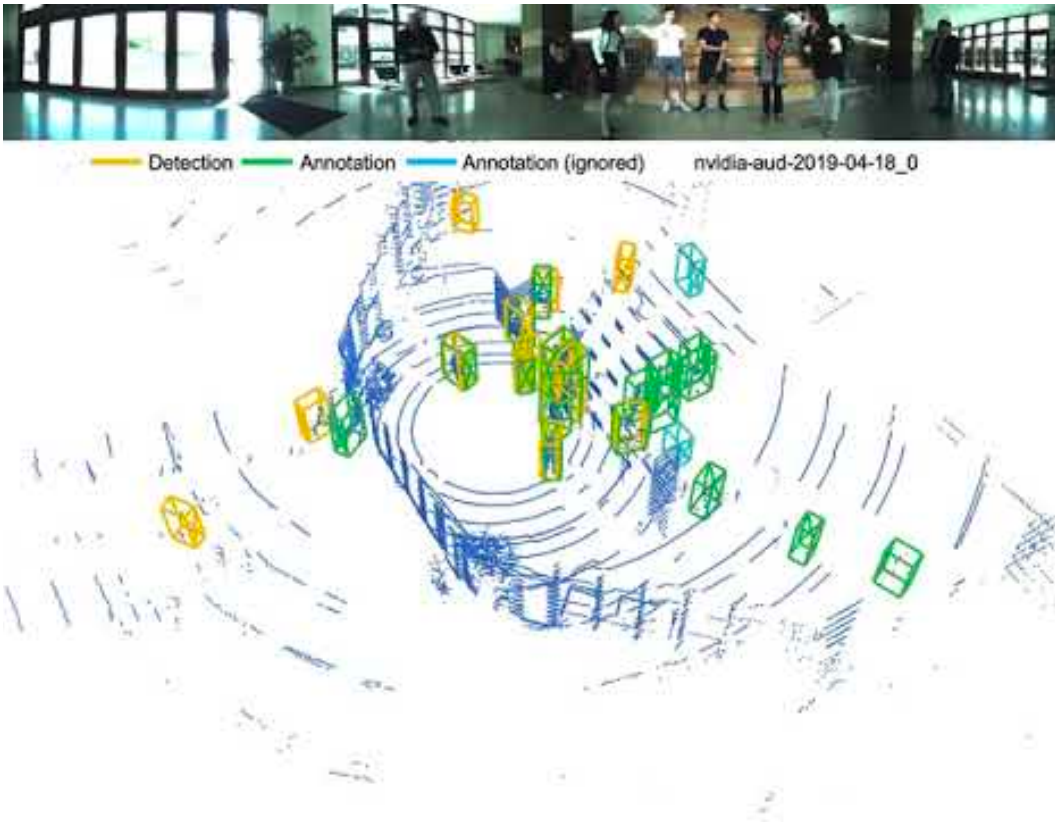



**Figure 20. People detection result of Person-MinkUNet**
Image Source: screenshot of animation in https://www.youtube.com/watch?v=RnGnONoX9cU

Numerous people detection research have been carried out over the past few years fusing RGB data with that point cloud from a LIDAR[77] [78]. However, this research will focuses on those research only use LiDAR point cloud data for people detection.

[77] Premebida, Cristiano, Joao Carreira, Jorge Batista, and Urbano Nunes. "Pedestrian detection combining RGB and dense LIDAR data." In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 4112-4117. IEEE, 2014.

[78] Schlosser, Joel, Christopher K. Chow, and Zsolt Kira. "Fusing lidar and images for pedestrian detection using convolutional neural networks." In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2198-2205. IEEE, 2016.

This work uses uses a existing model architecture Person-MinkUNet from JRDB leader board[79]. Person-MinkUNet was built on sub-manifold sparse convolution to people detection. It uses Minkowski Engine[80] and U-Net architecture[81].

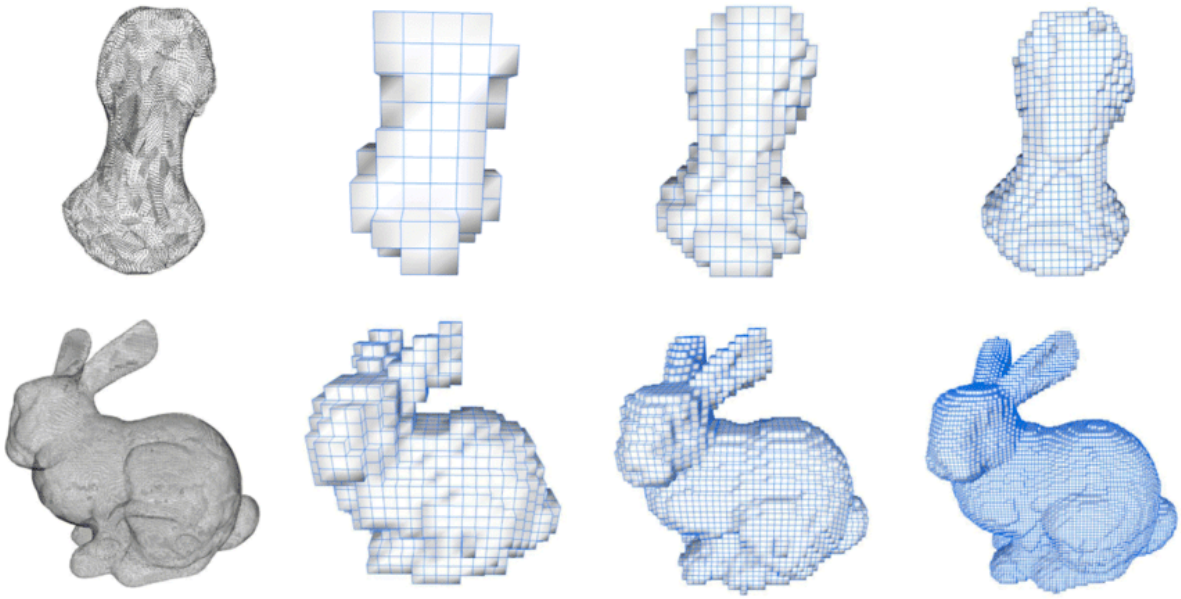


**Figure 21. Point cloud voxel model in different scales**
(from left to right: original point cloud object, scale=4, scale=2, scale=1)
Image Source: https://www.researchgate.net/figure/Point-cloud-voxel-model-in-different-scales-a-Point-clouds-model_fig4_340823107

The first step of this model is voxelization. It is a process of transferring discrete point cloud data into a grid-like system. The backbone for voxelization is based on the Sparse Point-Voxel Convolution

---

[79] Jia, Dan, and Bastian Leibe. "Person-MinkUNet: 3D Person Detection with LiDAR Point Cloud." *arXiv preprint arXiv:2107.06780* (2021).

[80] Choy, Christopher, JunYoung Gwak, and Silvio Savarese. "4d spatio-temporal convnets: Minkowski convolutional neural networks." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019.

[81] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." *International Conference on Medical image computing and computer-assisted intervention*. Springer, Cham, 2015.

(SPVConv) proposed in Tang et al[82]. This arhitecture was developed to overcome two limitations in previous works of Point-Voxel Convolution[83] and Sparse Convolution[84] [85].

Point-Voxel Convolution that performs the convolutions in voxels to increase locality and reduce erratic data access while representing the 3D input data in points to conserve memory. In particular, its voxel-based branch convolves over the voxelized input from the point-based branch while its point-based branch changes each point separately. However, model built upon this approach "takes 35 seconds to process a single scene, which is not affordable for most real-time applications"[86].

It has always been believed that scaling up volumetric convolution is impractical and inefficient. Sparse Convolution, which skips the non-activated regions to dramatically lower memory consumption, has recently been proposed by researchers. To be more precise, it first identifies all synapses that are active between the input and output points (referred to as the kernel map), and then it executes the convolution using this kernel map. Only these output points that also make up the input are taken into account in order to maintain the sparse activation.

As a result, compared to volumetric convolution as is, sparse convolution can afford a substantially greater resolution. Due to the constrained computation resources, the network cannot be particularly deep. In order to achieve a sufficiently big receptive field, the network must downsample very aggressively, which is quite lossy, and "this resolution is too coarse to capture the small instances"[87].

Sparse PointVoxel Convolution (SPVConv) integrates the aforementioned two methods to get over the restrictions of both modules. First, the high-resolution representation is always maintained by the point-

---

[82] Tang, Haotian, et al. "Searching efficient 3d architectures with sparse point-voxel convolution." *European conference on computer vision*. Springer, Cham, 2020.

[83] Liu, Zhijian, Haotian Tang, Yujun Lin, and Song Han. "Point-voxel cnn for efficient 3d deep learning." *Advances in Neural Information Processing Systems* 32 (2019).

[84] Graham, Benjamin, Martin Engelcke, and Laurens Van Der Maaten. "3d semantic segmentation with submanifold sparse convolutional networks." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 9224-9232. 2018.

[85] Same as 57.

[86] Same as 59.

[87] Same as 59.

based branch. Second, the sparse voxel-based branch models over various receptive field sizes using sparse convolution. Through sparse voxelization and devoxelization, two branches can communicate for very little cost. It extracts features and forms voxels for next step training. After that, "a fully connected layer is used to regress 3D bounding boxes from extracted features, and these boxes are used for detections"[88].

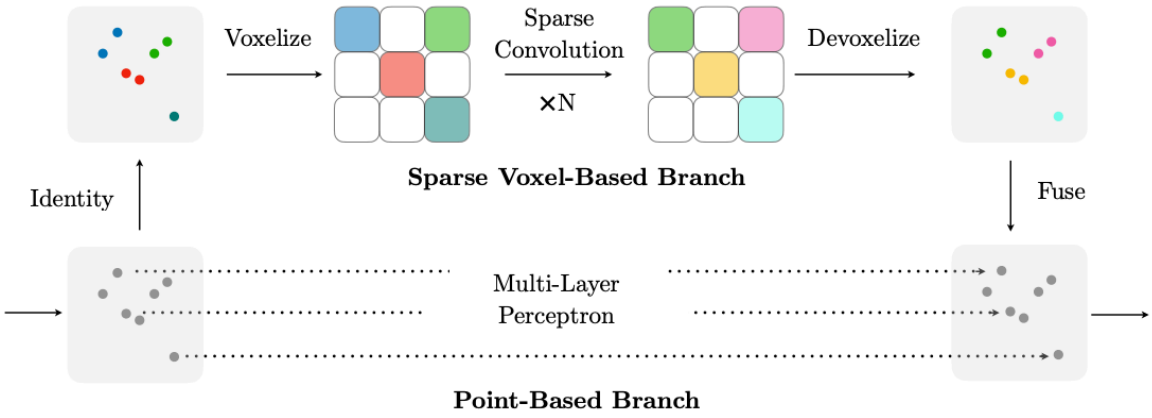



**Figure 22. Overview of Sparse Point-Voxel Convolution (SPVConv)**
Image Source: same as 59.

The output bounding box has 7 parameters (3 for location, 3 for dimensions, and 1 for z-rotation). Also, it contains 1 classification score for the detection probability. The probability is evaluated by measure loss by using Center-Point[89]. And the final loss function uses cross-entropy loss for classification and L2 loss for regression.

The training of this model is done by Google CoLab. It took about 24 hours to train the model with a validation result of 70.3% AP. The training is done with 100 epochs, Adam Optimizer and 1e-3 init learning rate. Regularization techniques are implemented with 0.3 dropout and 1e-7 weight decay.

[88] Same as 30.

[89] Yin, Tianwei, Xingyi Zhou, and Philipp Krahenbuhl. "Center-based 3d object detection and tracking." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2021.

| Result | Hyperparameter | epochs | scores |
|---|---|---|---|
| Result from Jia et al. | Learning rate 1e-3 for first 15 epochs, and then exponentially decay 1e-6; random scaling (0.95, 1.05), random rotation (z-aixs) | 40 | 71.5% AP |
| Result from this work | same as above. | 60 | 70.3% AP |

**Table 2: People-detection model training result**

## 5.4. Upsampling Model

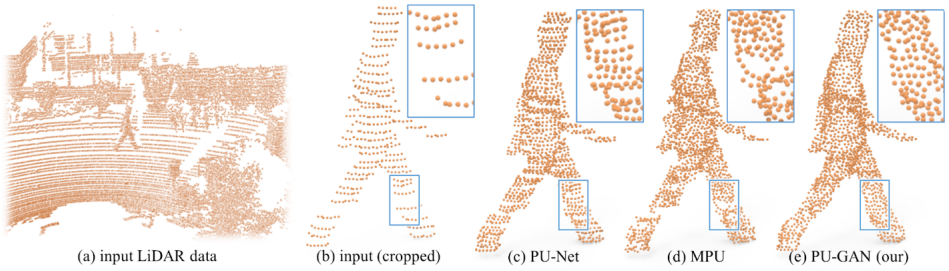



**Figure 23. Result of Pu-gan's Upsampling Result**
Image Source: Li, Ruihui, et al. "Pu-gan: a point cloud upsampling adversarial network." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019.

Another model deployed for this research is a upsampling model. Raw point cloud generated from LiDAR are often sparse, non-uniform and full of noise points. Upsampling is a process that can "generate a point cloud that is dense, complete, and uniform, as a faithful representation of the underlying surface"[90]. Early approaches for point cloud upsampling are usually math-based[91] [92]. So that they

[90] Li, Ruihui, et al. "Pu-gan: a point cloud upsampling adversarial network." *Proceedings of the IEEE/CVF international conference on computer vision*. 2019.

[91] Alexa, Marc, et al. "Computing and rendering point set surfaces." *IEEE Transactions on visualization and computer graphics* 9.1 (2003): 3-15.

[92] Lipman, Yaron, et al. "Parameterization-free projection for geometry reconstruction." *ACM Transactions on Graphics (TOG)* 26.3 (2007): 22-es.

interpolate new points or filter out noise points by computing the spatial surface with given point cloud sets. However, recent work is done focusing on learning-based methods[93] [94] [95].

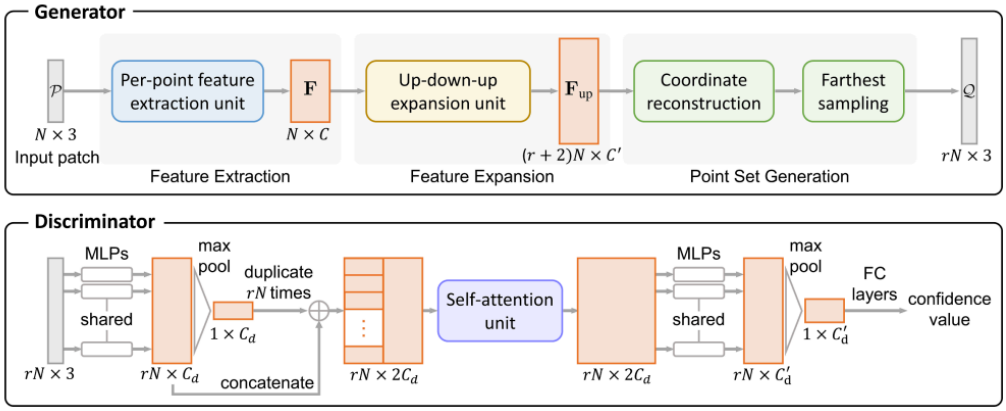



**Figure 24. Pu-GAN's generator and discriminator architecture.**
Image Source: Li, Ruihui, et al. "Pu-gan: a point cloud upsampling adversarial network." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019.

This research takes Pu-Gan, the next version of Pu-Net for implementation. The goal of this model is for any given sparse point cloud set *P* with *N* points, it is able to generate a dense *Q* point cloud set with *rN* points[96]. There are two rules the generated *Q* point set should meet: 1) "*Q* should describe the same underlying geometry of a latent target object as *P"*[97], 2) "points in *Q* should be uniformly-distributed on the target object surface, even for sparse and non-uniform input *P"*[98].

Pu-Net includes two parts: generator and discriminator. The goal of **generator** is to generate an augmented dense point cloud set as the result of upsampling. The goal of **discriminator** is to distinguish

[93] Same as 46.

[94] Yu, Lequan, et al. "Ec-net: an edge-aware point set consolidation network." *Proceedings of the European conference on computer vision (ECCV)*. 2018.

[95] Yu, Lequan, et al. "Pu-net: Point cloud upsampling network." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.

[96] Same as 47.

[97] Same as 47.

[98] Same as 47.

whether a given dataset is produced by generator. The whole process involve a series of feature expansion, interpolation and generation.

The **generator** includes three components: feature extraction component, feature expansion component, and point set generation component.

**The feature extraction component** will extract features from given dataset. For a given points set *P* of *N x d*, it will learn features ***F*** as a tensor. In here, N is the number of points and d is the number of dimensions (x, u, z, density, color, etc.). The feature extraction method is built upon the implementation in Wang et. al[99]. It uses dense connections to learn features across layers.

**The feature expansion component** will expand feature ***F*** to expanded features $\boldsymbol{F_{up}}$. This method develops an up-down-up expansion unit to improve the variation in $\boldsymbol{F_{up}}$. Hence, the generated points set can be presented with more various distribution. This method will be processed with following steps. First, it will copy features ***F*** and then fed them into separated MLPs. This will result in multiple independent expanded features $\boldsymbol{F_1}$. Second, it will upsample $\boldsymbol{F_1}$ to $\boldsymbol{F'_{up}}$, and then downsample it back to $\boldsymbol{F_2}$. Then the difference between $\boldsymbol{F_1}$ and $\boldsymbol{F_2}$ as Δ. Then, will be upsampled Δ to $\Delta_{up}$ and added together with $\boldsymbol{F'_{up}}$. This process is called as self-correction which can produce fine-grained features and avoid extra training[100].

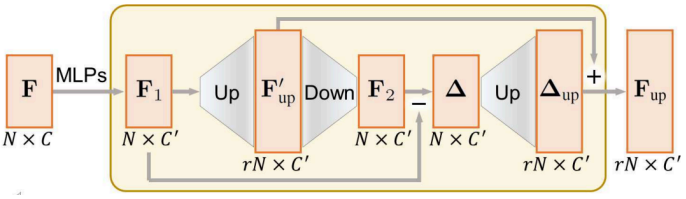



**Figure 25. The up-down-up expansion unit.**
Image Source: Li, Ruihui, et al. "Pu-gan: a point cloud upsampling adversarial network." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019.

For **the point set generation component** will feed $\boldsymbol{F'_{up}}$ to another MLPs to generate points. However, since $\boldsymbol{F'_{up}}$ is created as local information, this means most of expanded features will be close to original

[99] Same as 48.

[100] Same as 48.

inputs. To have a better distribution from global context, this work generate additional points ($r + 2$ in total) then select farthest $r$ point.

The **discriminator** is used to evaluate if a input is produced by generator. This requires extraction of global features. This work used a backbone from Yuan et al.[101] to achieve this goal. In addition to the backbone, self-attention unit is included. The attention unit will be able to achieve better feature fusion and extraction. After attention unit, MLPs and max pooling are implemented to generate global features and finally make regression values as the confident score. The confident score is in range from 0 to 1, and the higher value means the more likely a point sets is generated from generator.

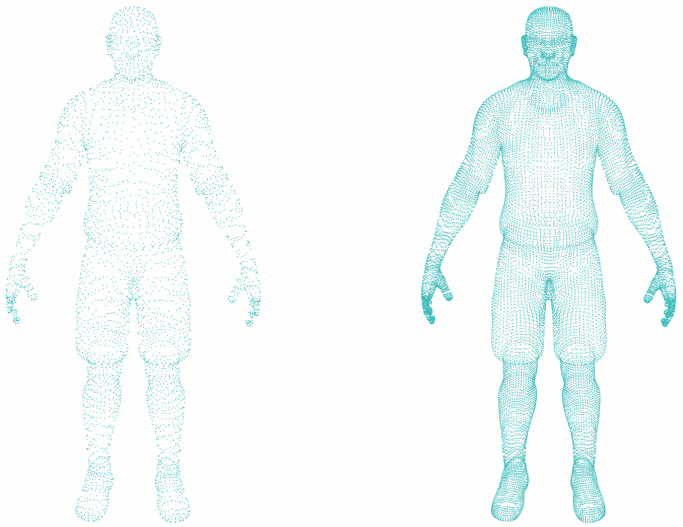


**Figure 26.(a) upsampled result, Figure 26.(b) original point cloud object**
Image SourceL created by author

| Result | Hyperparameter | epochs | scores (1.0%) |
|---|---|---|---|
| Result from Li et al. | Adam and two time-scale update rule, learning rate: 0.001 (generator), 0.0001 (discriminator), batch size 28. | 100 | 3.91 |

[101] Yuan, Wentao, Tejas Khot, David Held, Christoph Mertz, and Martial Hebert. "Pcn: Point completion network." In *2018 International Conference on 3D Vision (3DV)*, pp. 728-737. IEEE, 2018.

| Result | Hyperparameter | epochs | scores (1.0%) |
|---|---|---|---|
| Result from this work | same as above. | 120 | 4.02 |

**Table 3: Upsampling model training result**

The training of this model is done by Google CoLab. It took about 20 hours to train the model.

## 5.5. Combination of Detection and Upsampling

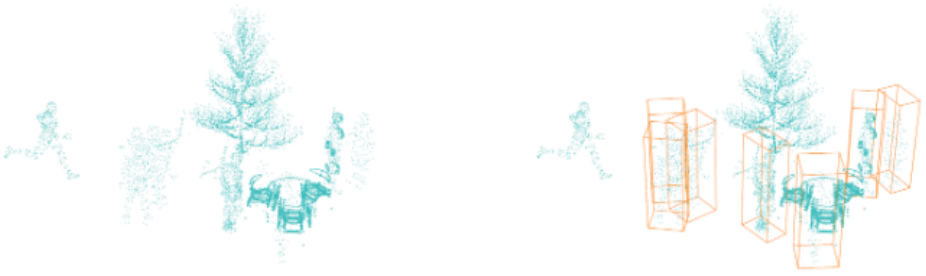


**Figure 27.(a) original result**
**Figure 27.(b) detection result**
Image Source: drawn by author



**Figure 28.(a) upsampling result**
**Figure 28.(b) upsampling + people detection result**
Image Source: drawn by author

After getting results of both upsampling and detection from the aforementioned models, this research synthesizes them together to create combined visualization results. Here, several combinations are tested to further understand proper visualization methods regarding recognization.

This research implemented the following visualization combinations: 1) detection bounding boxes, 2) upsampling point cloud, and 3) a combination of both upsampling and detection bounding boxes. All these possible combinations will be further tested in user test stages.

## 5.6. Design of User Interface

The design of the interface is shown below (Figure 29). The overall structure is simple and clear with only the necessary elements required. In the design, there will be two layouts to tackle two types of questions.

First, a page with a single image is displayed. On the right side, there will be a text box for questions regarding this image. The goal of this page is to ask questions specified to one visualization mode (original, upsampling, bounding box, mixed).

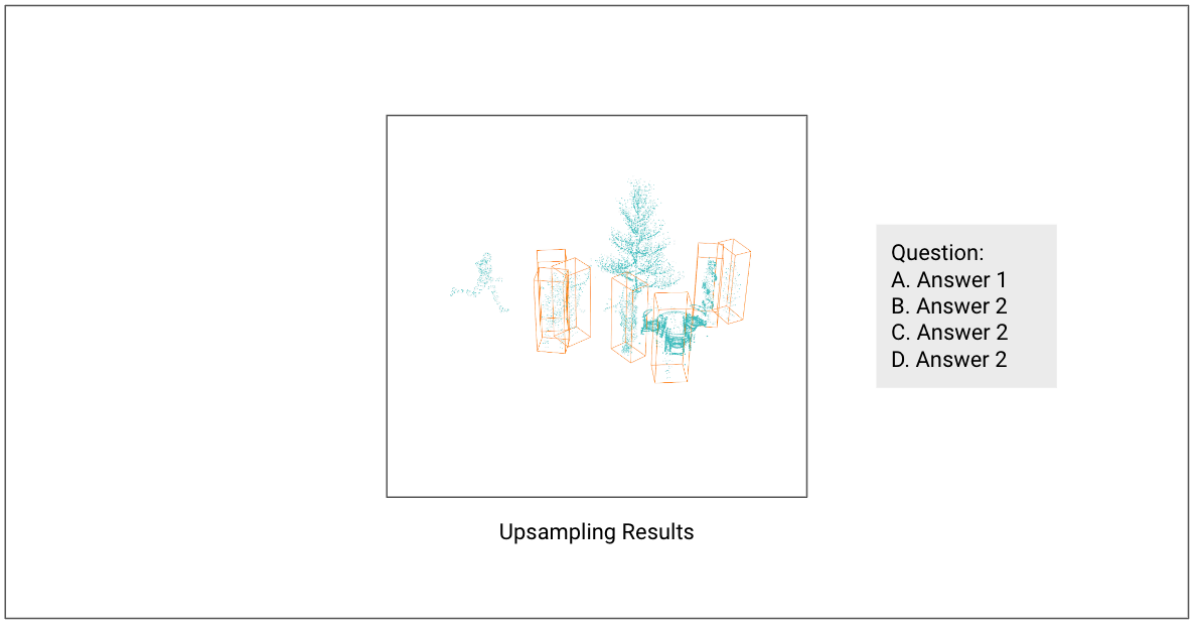



**Figure 29. User test interface design page 1: single result**
Image Source: drawn by author

Second, there will be another page that shows two images in juxtaposition. On the first page, there will be a text block with related questions. The goal of this layout is to ask a question regarding the comparison of two visualization modes. In the user test, these two pages will be arranged in a sequential order with several page1 appearing first, and then followed by page2.

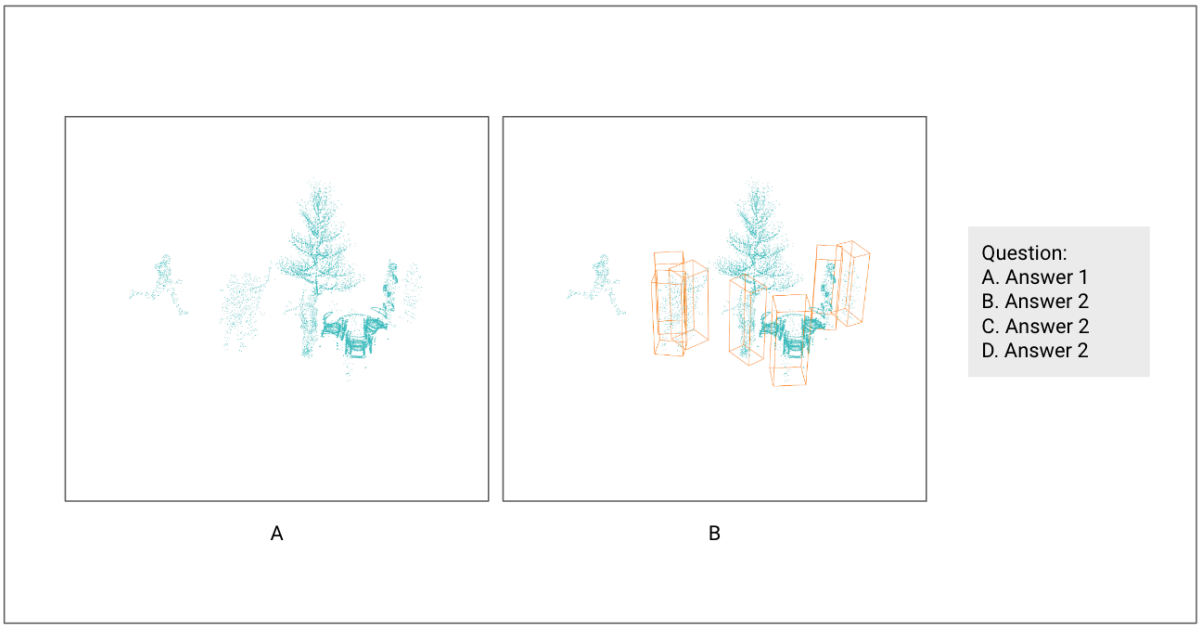



**Figure 30. User test interface design page 2: comparison**
Image Source: drawn by author

The next question this user interface design tries to tackle is the color palette used for visualization. In this research, three possible color combinations are tested for points' and boxes' visualization.

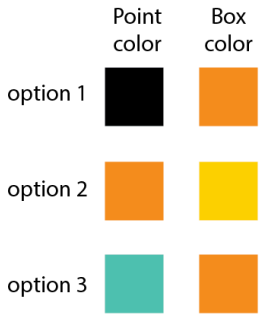



**Figure 31. Color palette**
Image Source: drawn by author



**Figure 32. Color palette option 1**
Image Source: drawn by author

The first combination is black points with orange boxes. For this visualization method, I notice that it may make users think boxes are way more important than points. However, this is not the focus of this research. Therefore, it may cause confusion for users to rely more on information from boxes for their recognization.



**Figure 33. Color palette option 2**
Image Source: drawn by author

The second combination is orange points with yellow boxes. This option was developed after option 1 with an idea of balance the visual importance of points and boxes. Therefore, in this option, two similar colors were chosen to visualize points and boxes. However, in this option, I notice that due to the similar colors it may confuse user since it is really hard to distinguish between boxes and points. This lead to harder recognization for user to understand visualization results.



**Figure 34. Color palette option 3**
Image Source: drawn by author

The third combination is cyan points with orange boxes. This option tries to cover the limits of both option 1 and option 2. On the one hand, both points and boxes are colored. So the user will not have the same feeling as option 1 that points are like background and boxes are foreground. On the other hand, points and boxes are chosen with contrasting colors so that users can still distinguish boxes and points. After using these three options to test with different users, finally, the third option was chosen as the color palette.

# 6. User Study

## 6.1. Overview

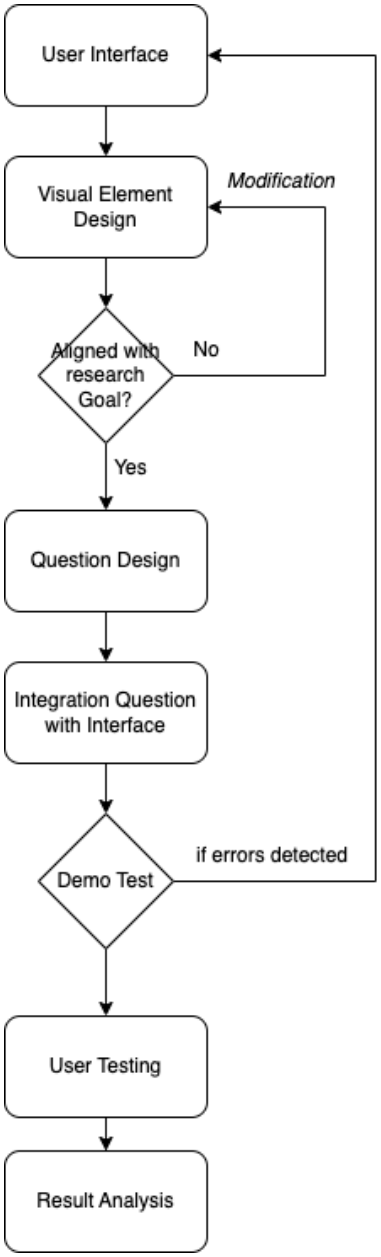



**Figure 35. User Study Implementation Methodology**
Image Source: drawn by author.

The user study methodology will mainly focus on testing visualization tools with users and collect feedbacks that can be used for further analysis.

The user study starts with iterations with interface design. Since the visualization tools developed from the last step are able to generate multiple visualization results. Then, how to select those results and present them to users for testing purposes is the key question. To achieve this, there are several critical components designed in this methodology.

First, the visual element design will be conducted to determine color palate and plot weight for visualization results. In this process, it is necessary to consider the hearty of information layers the result wants to deliver to users. By iterating with those visual elements, it is able to achieve a better information perceiving process.

Second, a series of questions will be developed with user testing. By letting users answers question related to visualization results, this research can collect data reflecting the level of recognizability. In this section, the question will be surrounded by three sub-categories: object recognizability, semantic recognizability, and spatial recognizability. How to integrate questions into the interface is also an important part. Especially for questions asking for a comparison between different results. Therefore, multiple iterations between interface design and question design are also going to be conducted.

Finally, the research will conduct user tests with volunteers. The result will be captured in the format of quantitative data and some qualitative comments from users. The analysis against those collected data will be applied to further understand how different visualization methods will impact users' recognizability.

## 6.2. Types of Recognizability

Effective visualizations make it easier to understand data sets by using the right visual metaphor. There are several presentation tactics available in the visualization industry, many of which may be used with data of a similar nature. These varied methods frequently provide distinctive imagery that highlights certain data properties or visualization objectives. Several rendering approaches are usually adequate, but

some may convey important information more quickly and precisely. It might be challenging to select the optimal visualization method for a given data collection. The visualization expert must make sure the methodology chosen will provide answers to the queries put forth by domain experts in addition to choosing an acceptable technique for the type of data. Therefore, user-centered study for visualization methods are necessary.

| Recognizability Type | Description | How to test? |
|---|---|---|
| Object Recognition | Ability to recognize objects in given scene. | Object counting |
| Semantic Recognition | Ability to recognize semantic of objects in a scene or the scene itself. | Identify objects |
| Spatial Recognition | Ability to tell scale and orientation of a space from given scene. | Spatial estimation, place guessing |

**Table 4: Research Methodology**

Several works have been done to reveal how visualization can impact recognition. Per Sheppard's research, including material textures, lighting, and climatic conditions in photorealistic representations appears to enhance identification, facilitate comprehension, and enable evaluation of the visualization result's aesthetic and emotive aspects[102]. Sheppard's work is mainly focusing on understanding the relationship between recognization and landscape visualization. For this work, the categories of recognizability will be expanded related to point-cloud visualization. By evaluating purposes of point point-cloud visualization, this work creates three sub-categories for point-cloud recognizability.

Given this using of point cloud, this work identifies three types of recognition emerging from point cloud application. By categorizing recognition into three types, this research is able to design questions specifically towards each type. Thus, forming a more straightforward user testing process.

**Object recognition** is a fundamental ability for human to recognize environment. For example, when we are perceiving a scene, we are able to to identify each individual object. In real word, this is done by a

[102] Sheppard, Stephen RJ. "Landscape visualisation and climate change: the potential for influencing perceptions and behaviour." *Environmental science & policy* 8, no. 6 (2005): 637-654.

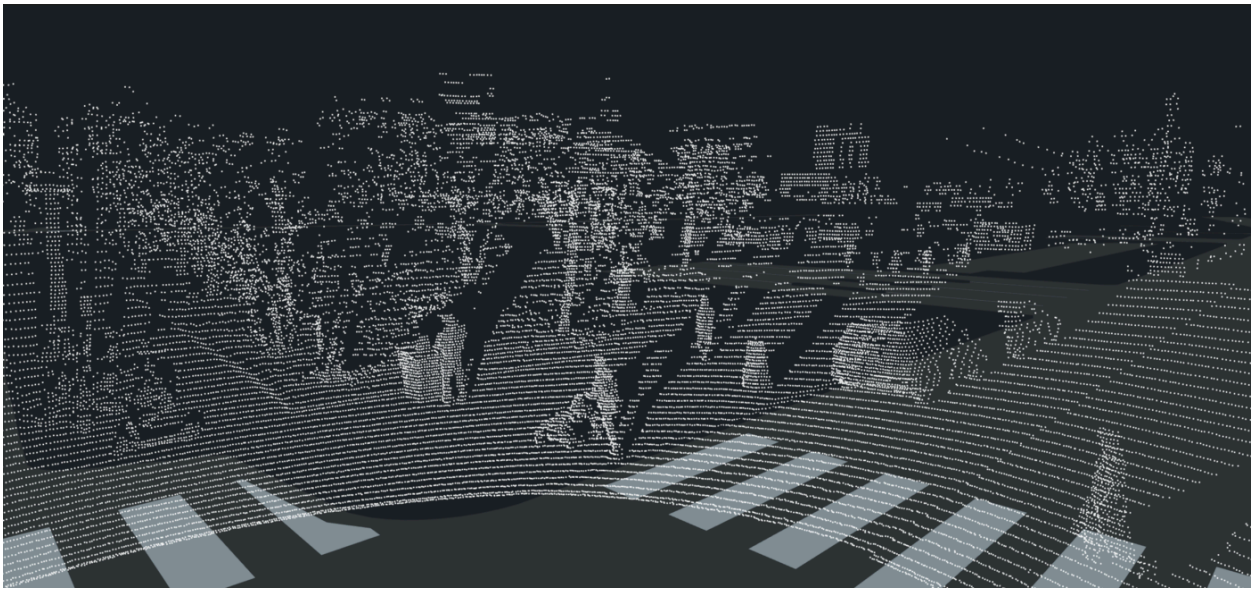


**Figure 36. For point cloud view, even recognize all objects is not an easy task.**
Image Source: https://medium.datadriveninvestor.com/lidar-3d-perception-and-object-detection-311419886bd2

series of recognition process based on objects' shapes, colors and textures, as well as environments light, etc. However, in point cloud, due to the level of abstraction, event the basic live object recognition is not as easy as in the task in real world. In this research, object recognition will be tested by basic ability for user to count number of objects on visualization results.

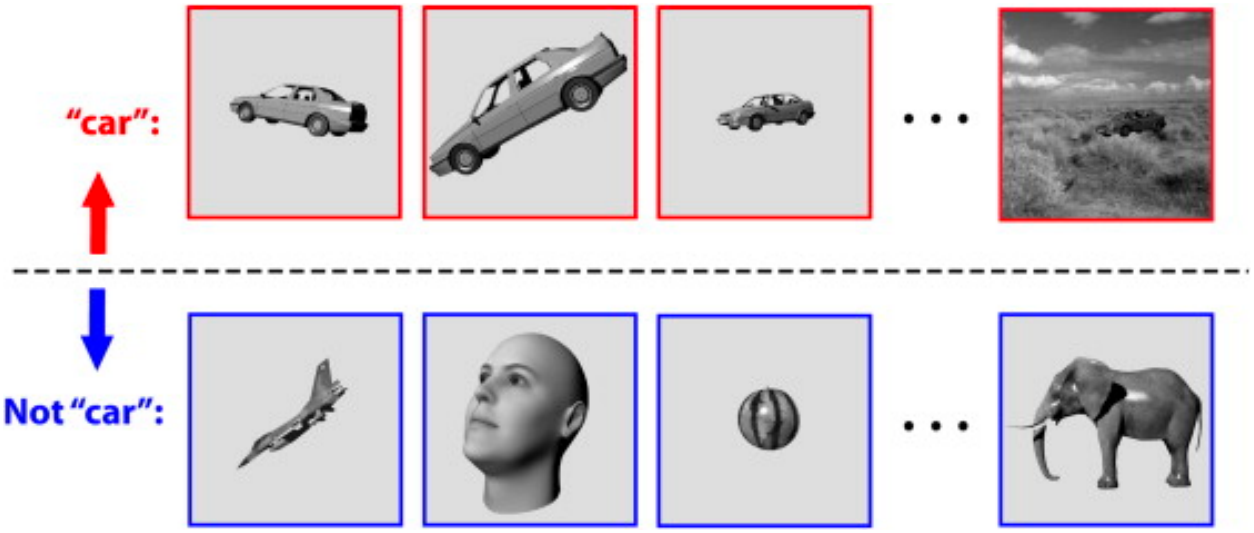



**Figure 37. Semantic recognition.**
Image Source: https://www.sciencedirect.com/science/article/pii/S089662731200092X

**Semantic recognition** is the second level of recognition that identifies semantic meaning for objects that are recognized. For each object recognized by human, we will attempt to give it a semantic meaning. It is defined "as the ability to accurately discriminate each named object ('identification') or set of objects ('categorization') from all other possible objects, materials, textures other visual stimuli, and to do this over a range of identity-preserving transformations of the retinal image of that object (e.g. image transformations resulting from changes in object position, distance, and pose)."[103]. This is the core part for recognition since it help human assign meaning to each objects and then those meaning can be used for communication and thinking. Thus, we are able to connect our internal brain activities with external physical stimuli. It is easy to notice that, in our interactions with the outside world, external objects are the same, but our reactions to and interpretations of them are not. Because perception is culturally determined, this is the case. In other words, "cultural factors provide some of the meaning involved in perception and are, therefore, intimately implicated with that process"[104]. However, in this research, we tries to eliminate the impact of cultural background of users towards the final results. This will be helpful to eliminate the workload of user recruiting and further analysis. To make it possible, the test of this research will focus on semantic recognition of some basic objects such as human, chairs, and trees.

**Spatial recognition** is the ability to recognition spatial related information. For example, when we view the image of a room, we are able to make quick and rough estimation of the room's size. This is a type of recognition directly related to point cloud application. In most of cases for point cloud application, we review point cloud for work related to spatial survey, detection, and analyzing. Therefore, for this research, the spatial recognition will be emphasized by designing questions directly regarding to it. In user test, questions such as estimating scale and describing environment will be added for this purpose.

## 6.3. Question Design

---

[103] DiCarlo, James J., and David D. Cox. "Untangling invariant object recognition." *Trends in cognitive sciences* 11, no. 8 (2007): 333-341.

[104] Samovar, Larry A., Richard E. Porter, and Nemi C. Jain. *Understanding intercultural communication*. Wadsworth Publishing Company, 1981.

| Question Type | Questions | Sample Ansers |
|---|---|---|
| **Basic Questions** | Have you every work with point cloud? | (A) Yes, frequently, (B) Yes, for 1 or 2 projects, (C) Yes, have seen some visualization with point cloud, (D) only heard of it, (E) no idea what it is |
| | Have you every work with visual objects based analysis? *(based on photos, drawings, maps)* | (A) Yes, frequently, (B) Yes, for 1 or 2 projects, (C) Yes, have seen some visualization with point cloud, (D) only heard of it, (E) no idea what it is |
| | What is your major? | text input |
| **Object Recognition** | How many objects can you recognize from the image? | (A) count1, (B) count2, (C) count3, (D) count4 |
| | How easy for you to get this answer? *(1: very difficult, 5: very difficult)* | (A) 1, (B) 2, (C) 3, (D) 4, (E) 5 |
| | Did you change your answer for (1) after seeing the new image? *If so, select the new answer. If not, select the original answer*. | (A) count1, (B) count2, (C) count3, (D) count4 |
| | Given your experience, do your thing this new image improve your judgement? *(1: add difficulties for my recognition, 5: make my recognition easier)* | (A) 1, (B) 2, (C) 3, (D) 4, (E) 5 |
| **Semantic Recognition** | Select all objects shown on the images. | (A) object1, (B) object2, (C) object3, (D) object4 |
| | Can you give a brief description for this image? | text input |
| | How easy for you to get this answer? *(1: very difficult, 5: very difficult)* | (A) 1, (B) 2, (C) 3, (D) 4, (E) 5 |

| | | |
|---|---|---|
| | Did you change your answer for (1) after seeing the new image? *If so, select the new answer. If not, select the original answer*. | (A) object1, (B) object2, (C) object3, (D) object4 |
| | Given your experience, do your thing this new image improve your judgement? *(1: add difficulties for my recognition, 5: make my recognition easier)* | (A) 1, (B) 2, (C) 3, (D) 4, (E) 5 |
| **Spatial Recognition** | Can you estimate the scale of this space? | (A) size1, (B) size2, (C) size3, (D) size4 |
| | Can you guess this is an indoor or outdoor space? | (A) indoor, (B) outdoor, (C) not sure |
| | How easy for you to get this answer? *(1: very difficult, 5: very difficult)* | (A) 1, (B) 2, (C) 3, (D) 4, (E) 5 |
| | Did you change your answer for (1) after seeing the new image? *If so, select the new answer. If not, select the original answer*. | (A) size1, (B) size2, (C) size3, (D) size4 |
| | Given your experience, do your thing this new image improve your judgement? *(1: add difficulties for my recognition, 5: make my recognition easier)* | (A) 1, (B) 2, (C) 3, (D) 4, (E) 5 |
| **Summary** | Could you leave some thoughts about this test? (what you find challenging or interesting) | |

**Table 5: Question Design**

Questions for this research is designed to capture three cognition types defined previously. There are five sections for questions.

First, a section of basic questions will collect basic information regarding to users. In this section, questions will be asked to have better understanding for users' familiarity and experience of working with point cloud.

Second, question regarding to object recognition will be asked. For this section, it mainly measure and evaluate users' ability to recognize objects shown on image. Therefore, questions will ask users to count

number of objects. It will show original image, image with upsampling, image with bounding box, and finally image with both upsampling and bounding box. Questions will be asked sequentially for all 4 steps. Also, time used to answer the question will also be recorded.

The third section is about semantic recognition. In this section, users will be asked to identify each object with its semantic meaning. To do so, the first question will ask users to select all objects they have recognized in images. After that, the second question asked user to briefly summarize about what the image is about. To conduct these two questions, we are able to understand the ability for user to understand each individual object's semantic meaning and the ability to understand the whole scene. Same as the second section, comparison between original image and augmented (upsampling and bounding box) will also be conducted.

The fourth section will test users' ability for spatial recognition. This is done by let users estimate each scene and give rough guessing of the scale. Usually, scale guessing is able to be done by human's perception and recognition with understanding of perspective and size consistence. Since point cloud scene is captured with less "line" elements, capturing perspective clue is not as easy as through camera photos. Same as the second and the third section, for spatial recognition, users will be asked to recognize spatial characters regarding to augmented results.

Finally, a summary text box will let user to briefly conclude challenge and fun for the whole process. The aim of having this section is to leave some flexibility to let user express their thoughts. This research expect to get inspirations for next step's or future's from those texts.

# 7. Results

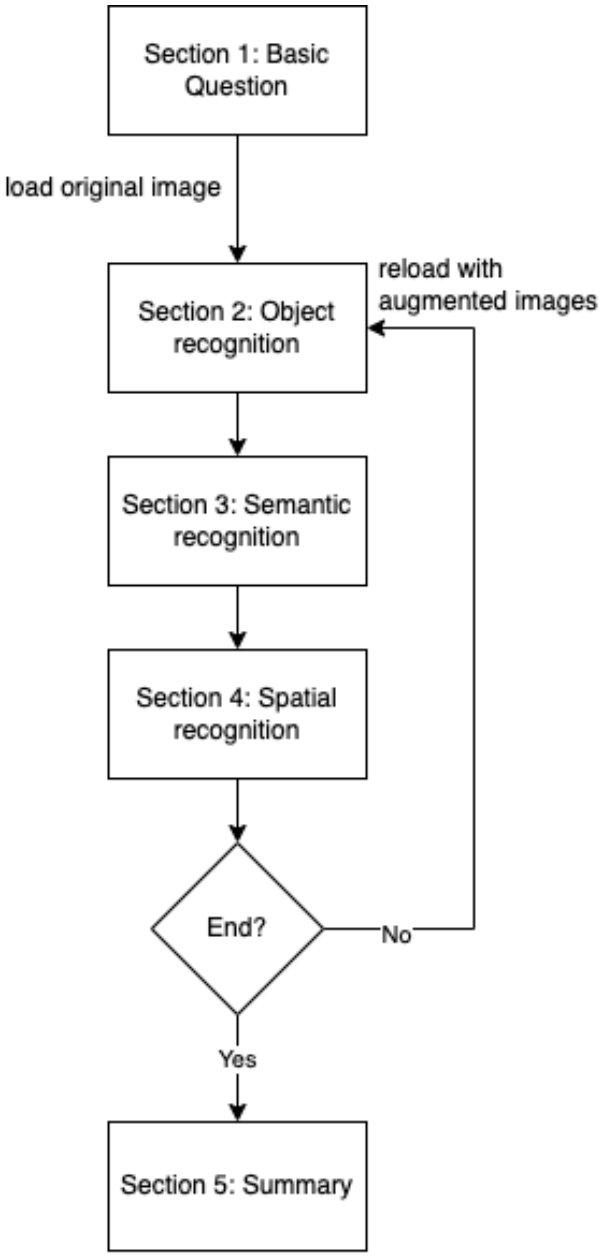



**Figure 38. Question Design**
Image Source: drawn by author

## 7.1. User Study Result Overview

The user test was conducted with 25 users in total. Users come from different background including design, management, marketing, and engineer. All users take relatively similar period of time to complete

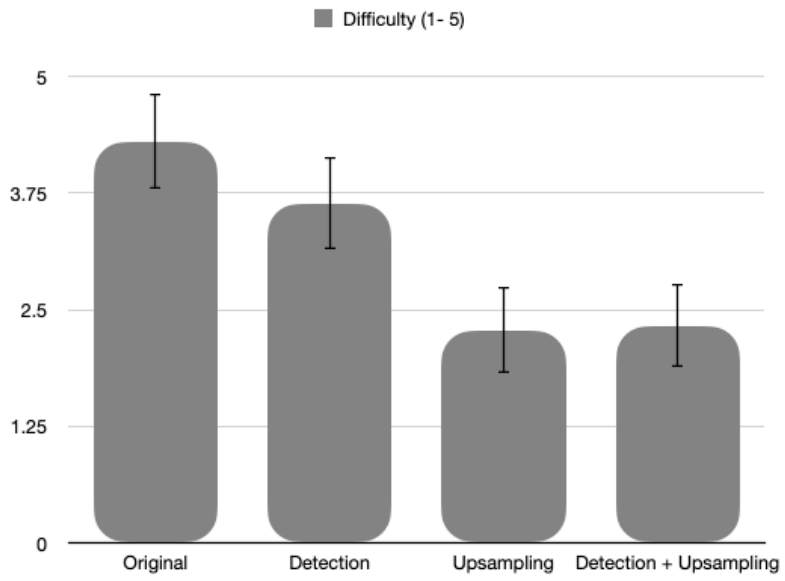



**Figure 39. Results of user test regarding to level of difficult of for each visualization option.**
Image Source: drawn by author

the research (approx. 5mins).

| Options | Average | Standard diviation | Samples |
|---|---|---|---|
| Original | 4.31 | ±0.51 | 25 |
| Detection | 3.64 | ±0.49 | |
| Upsampling | 2.28 | ±0.46 | |
| Detection + Upsampling | 2.33 | ±0.44 | |

**Table 6: Results of user test regarding to level of difficult of for each visualization option**

Overall, augmentation strategies have improved user recognition significantly. The average difficulty rating given by users for the original point cloud visualization is 4.31, whereas the results of three different augmentation choices far exceed this rating. However, people' reactions to the three enhancement procedures vary. In general, consumers do better in upsampling approaches than in

detection. However, as compared to upsampling alone, the combination of detection and upsampling does not demonstrate a significant improvement.

## 7.2. Analysis

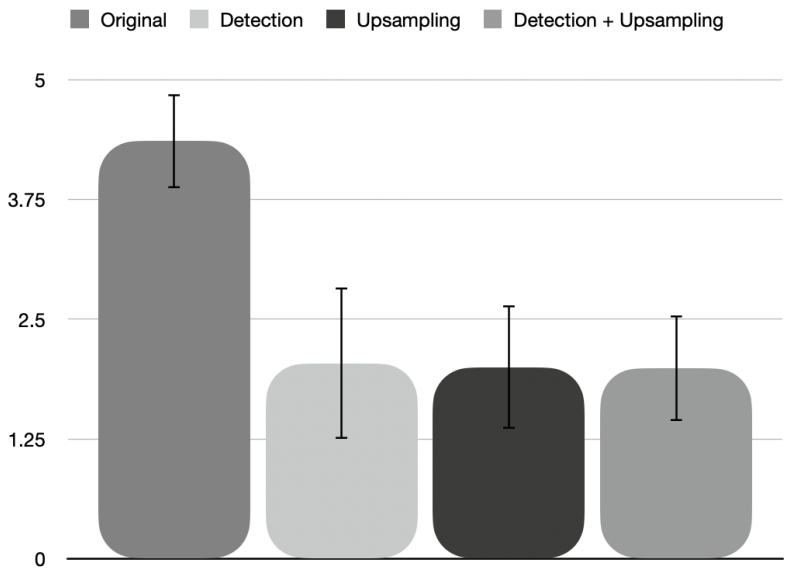



**Figure 40. Object Recognition results**
Image Source: drawn by author

| Options | Average | Standard diviation | Samples |
|---|---|---|---|
| Original | 4.36 | ±0.49 | 25 |
| Detection | 2.04 | ±0.79 | |
| Upsampling | 2 | ±0.65 | |
| Detection + Upsampling | 1.99 | ±0.55 | |

**Table 7: Object Recognition results**

Figure 40 depicts the outcome of object recognition. As a consequence, we can observe that both Detection and Upsampling may improve identification capacity for object detection. However, no

substantial difference exists between upsampling and detection. And the combination of both produces comparable outcomes as each choice alone.

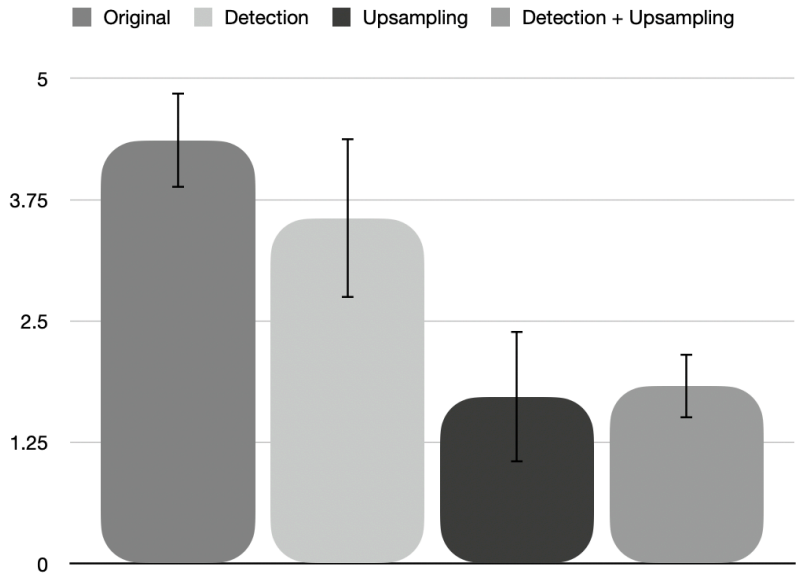



**Figure 41. Semantic Recognition results**
Image Source: drawn by author

| Options | Average | Standard diviation | Samples |
|---|---|---|---|
| Original | 4.36 | ±0.49 | 25 |
| Detection | 2.04 | ±0.79 | |
| Upsampling | 2 | ±0.65 | |
| Detection + Upsampling | 1.99 | ±0.55 | |

**Table 8: Semantic Recognition results**

Figure 41 shows the results of semantic recognition. From the result, we can see both detection and upsampling are able to enhance the ability of recognition. Upsampling has outperformance than Detection. To further understand the reason, one scene shows great difference (Detection avg. score 4.03, Upsampling avg. score 1.15) for detection and upsampling rates is shown here.

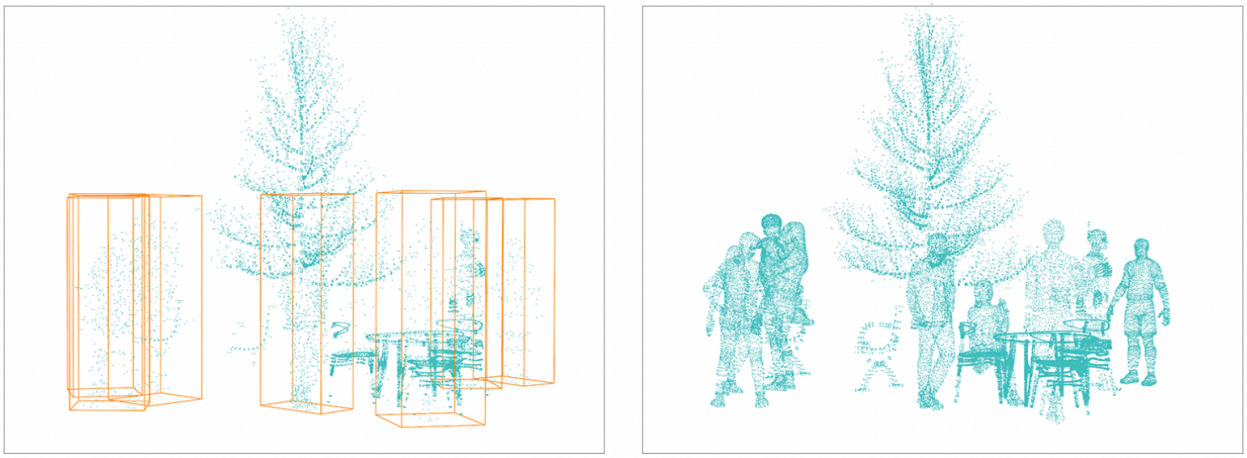


**Figure 42. Left: detection result, Right: upsampling result**
Image Source: drawn by author

It is clear from this example that augmentation of detection bounding boxes is ineffective in recognition help. On the other hand, upsampling methods capture geometrical features powerfully and allow users to deeper appreciate their semantic connotations.

## 7.3. Summary

We may form a preliminary conclusion on the built visualization framework based on the user study.

First, augmentation tools can improve user recognition. However, because of the nature of various augmentation procedures, this capability varies.

Second, each augmentation approach may have distinct impacts for different recognition modes. For instance, detection boxes are ineffective for semantic recognition.

Third, varied scene and view angles will have an influence on augmentation technique detection. In a cluttered scene with many objects overlapping, for example, detection approaches outperform upsampling.

Fourth, the synthetic augmentation approach is not necessarily superior. In some circumstances, additional augmentation information may raise the burden of recognition and, as a result, become distracting.

# 8. Speculative Design

## 8.1. Speculative Design

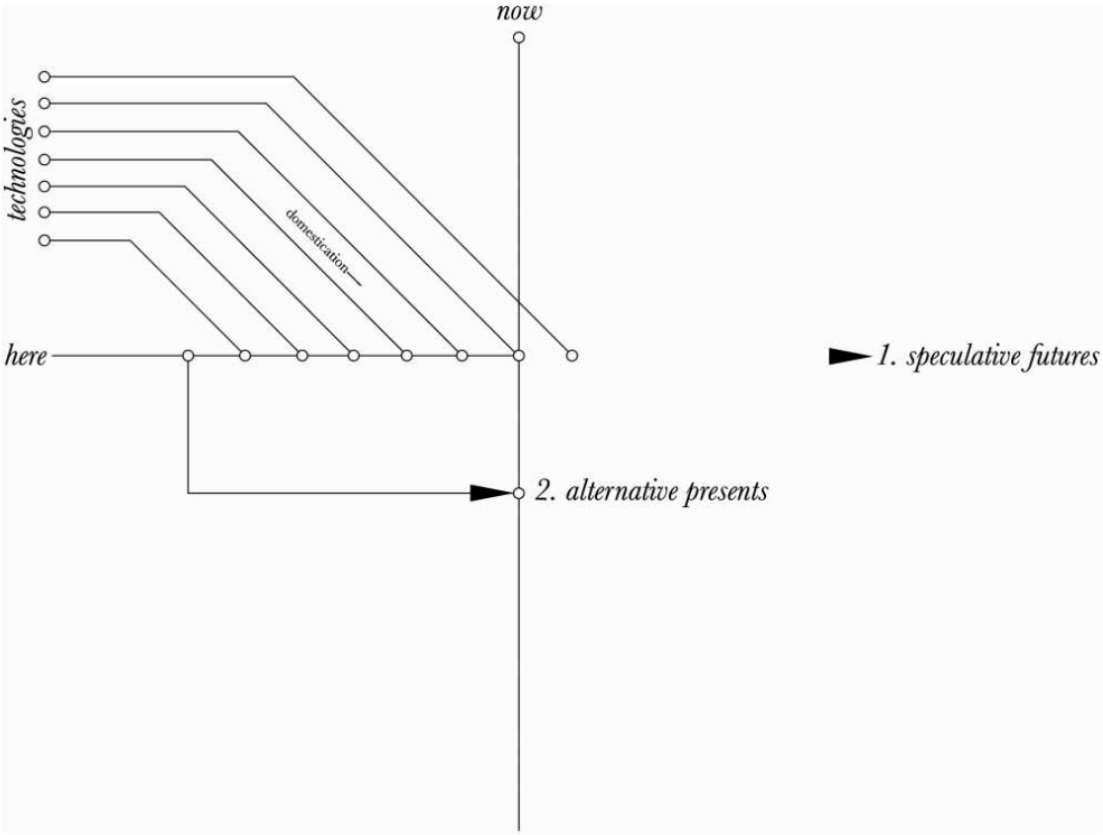



**Figure 43. the goal of speculative design.**
Image Source: Auger, James. "Speculative design: crafting the speculation." *Digital Creativity* 24, no. 1 (2013): 11-35.

Since the early 2000s, speculative design, along with allied techniques such as critical design and design fiction, has increased in significance in HCI[105]. Speculative design was originally developed as a practice for divining "new genres" of technology use, but it has come to describe critically oriented research practices that create artifacts, representations, or depictions of possible and often alternate futures, away from immediate practical concerns of implementation and commercial viability. Speculative design in

[105] Wong, Richmond Y., and Vera Khovanskaya. "Speculative design in HCI: from corporate imaginations to critical orientations." In *New Directions in Third Wave Human-Computer Interaction: Volume 2-Methodologies*, pp. 175-202. Springer, Cham, 2018.

HCI may take many forms, from design concepts to actual objects, and is used to conceive other sociotechnical configurations of the world in order to probe problems about values and politics through design. As shown in the figure, speculative design presents *alternative presents* with a projections of new technologies to nowadays.

## 8.2. Design 1: Night Vision Glasses



**Figure 44. night vision for wildlife research.**
Image Source: https://www.researchgate.net/figure/observed-wildlife-using-night-vision-technique_fig4_320299022

Night vision is a capacity to see in low-light circumstances, either naturally with scotopic vision or using a night-vision apparatus, is referred to as night vision. Night vision needs both a wide spectral range and a narrow intensity range. Humans have poor night vision in comparison to many animals. Therefore, several techniques used to assist human for better night vision capability.

This speculative design projects the developed visualization tools for a night vision glasses. Since the development of this tools aims at make point could more accessible for everyday using. This design incorporate into glasses with multiple modes. For example, a original mode can display point cloud directly processed by LiDAR. Then, other modes can incorporate point cloud with augmented methods to

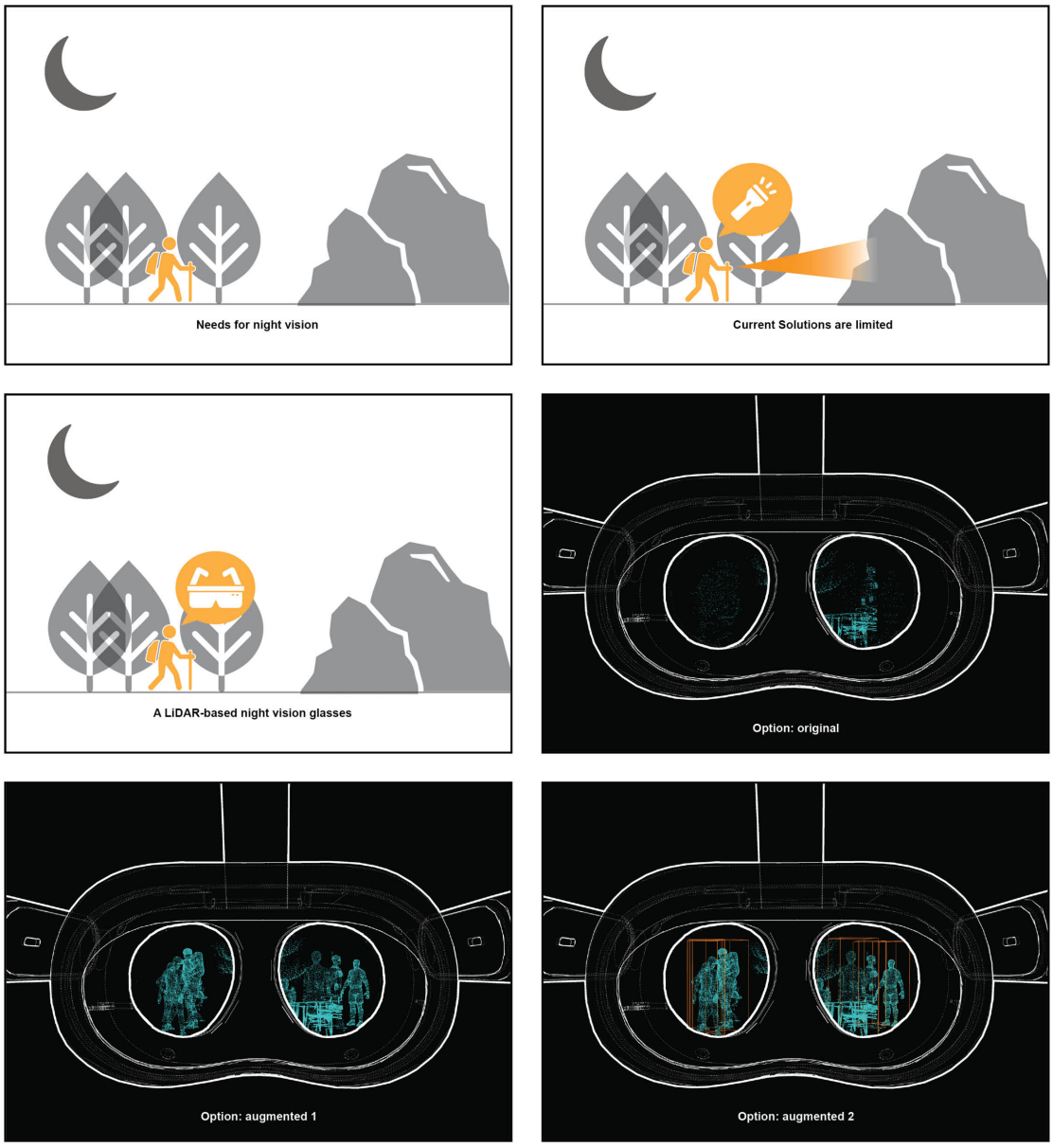



**Figure 45. Storyboard of Night Vision Glasses.**
Image Source: Drawn by author.

enhance people's recognizability for specific conditions. For example, for condition requires with focus on people, people detection bounding box can be applied. For other situations, new augmented methods can also be developed.

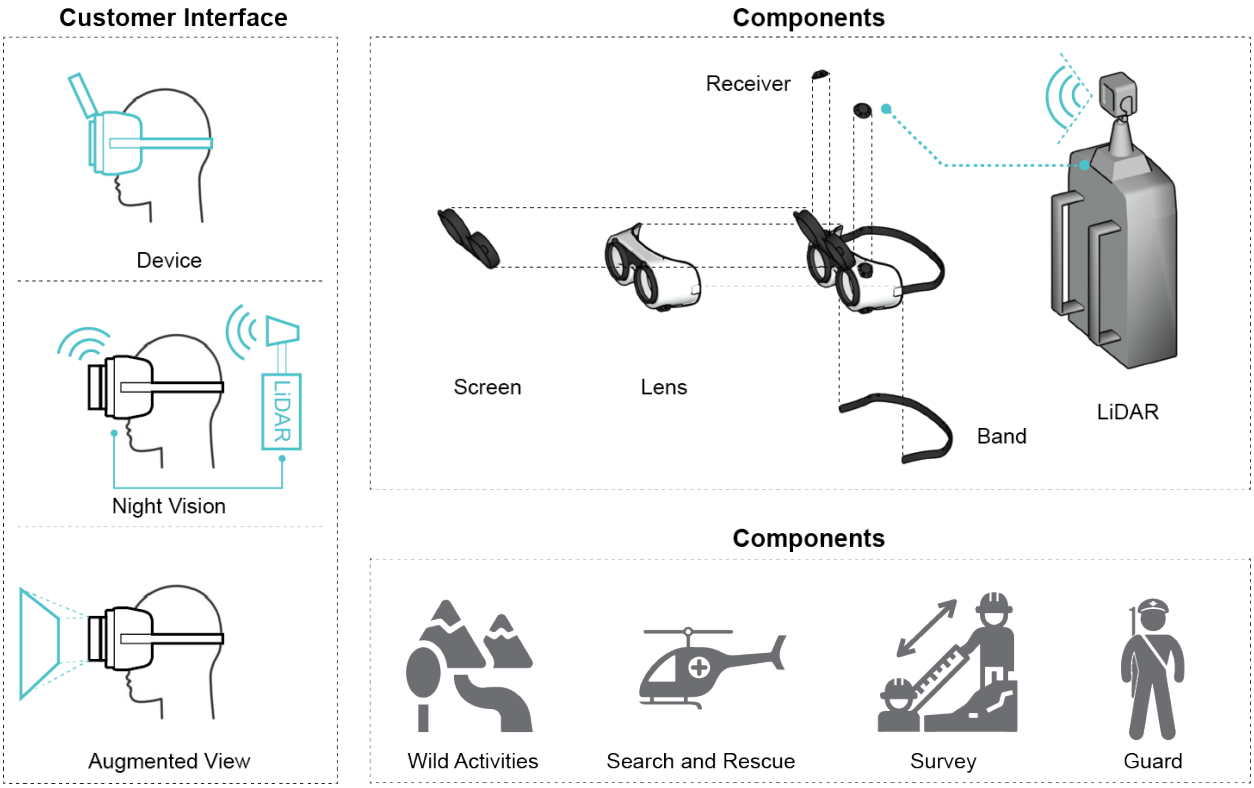



**Figure 46. Design Diagrams of Night Vision Glasses.**
Image Source: Drawn by author.

The physical product of this design is composite of two part: a glasses that is able to let user see point cloud scene through lens, a portable LiDAR device that can be mounted to backpack or other places.

The user interface is designed with a prototype of goggles that provides an on/off control. Then, after the device is on, the LiDAR device starts to produce point cloud scenes of the surrounding. It can be sent through Bluetooth and received by the glasses' receiver. The result can be visualized through a lens. Multiple modes will be developed to augment results for different use cases such as wild activities, rescue, survey, and guard.

## 8.3. Design 2: Dance with Cloud

Body performance is a type of performance that involves using bodies to express movement, rhythm, force, and form. It creates aesthetic and cultural meaning by arranging body movement with music, lights,



**Figure 47, *Fase, Four Movements to the Music of Steve Reich* by Anne Teresa De Keersmaeker.**
A performance including people, shadows, and light.
Image Source: http://www.bureaudesrecommandations.com/dance

space design, and narratives. Consequently, the visual representation of the stage is one of the most important key elements for body performance. For some experimental performances, performers used representation methods to reinterpret the body's presence and create additional layers of meaning for the performance. For example, in Anne Teresa De Keersmaeke's work (Figure 47), she creates shadows to help express the performance. The presence of shadow can be seen as another representation layer adding an essential narrative for the performance. This work inspired several experimental performance afterwards.

Whitley's work (Figure 48) is a eight-minute virtual reality film that explores movement on human and celestial proportions using modern dance and motion capture technologies. The user may flip between two gorgeous cosmic realms and experience the choreography in various ways in this interactive Daydream experience. It was developed on Guardian VR platform. This work demonstrates value of synthesizing of physical presents and digital representation in creating new performance forms.



**Figure 48, Celestial Motion by Alexander Whitley**
Image Source: https://www.ft.com/content/e48c0458-beb0-11e9-9381-78bab8a70848

This speculative design proposes using of point cloud visualization for artistic expression. This point cloud visualization work research into the relationship between point cloud visualization realists and human recognizability. Therefore, by expanding various augmentation features, point clouds are able to create a new representation layers like Anne Teresa De Keersmaeker work but is more rigorous and rich. By incorporating multiple augmentation methods with body movement, rhythm and narrative, point cloud can further modify the appearance of human body and provide alternative understanding between performance and representation.

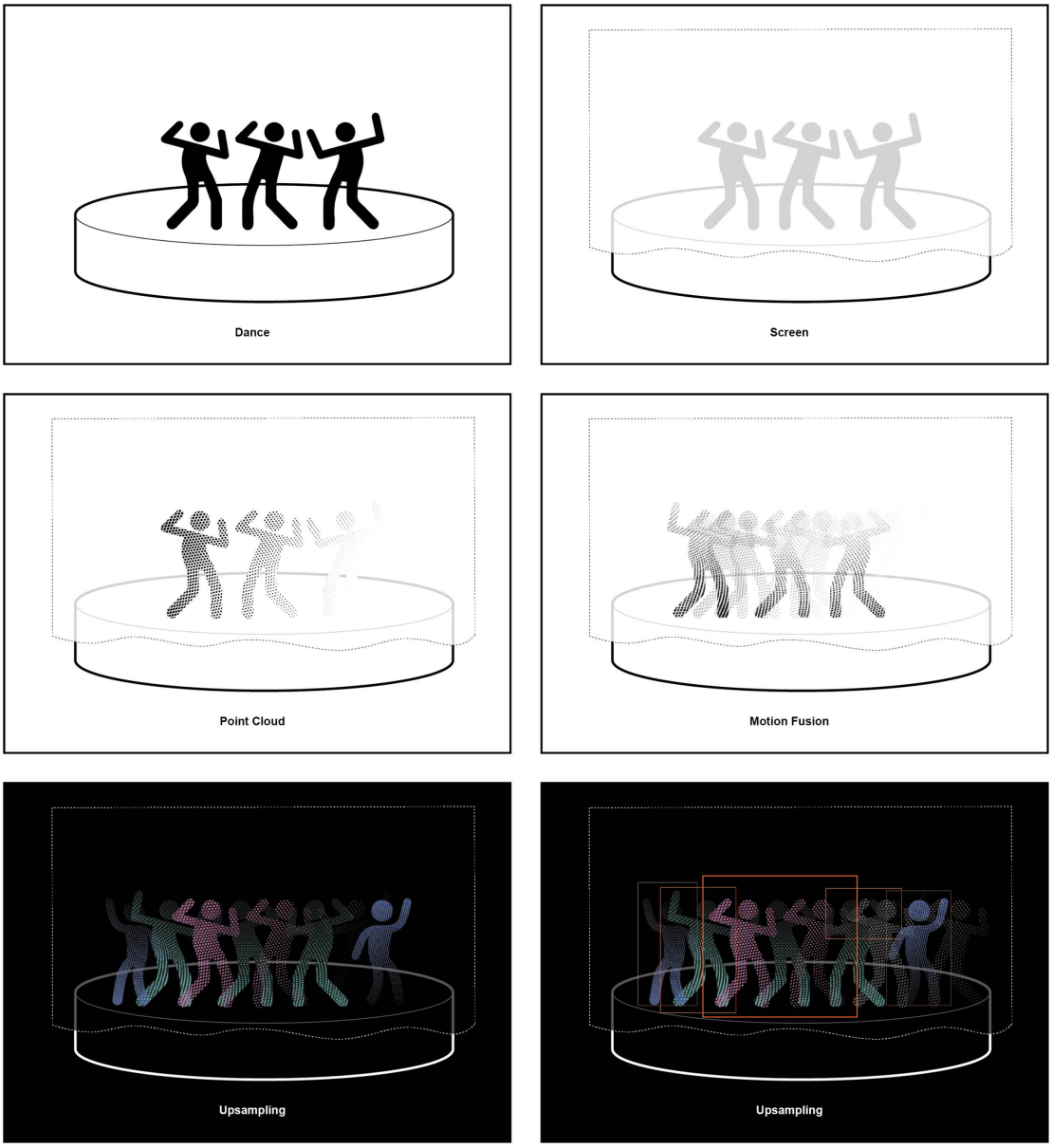



**Figure 49. Storyboard of Dance with Cloud.**
Image Source: Drawn by author.

## 8.4. Design 3: Automated construction assistant

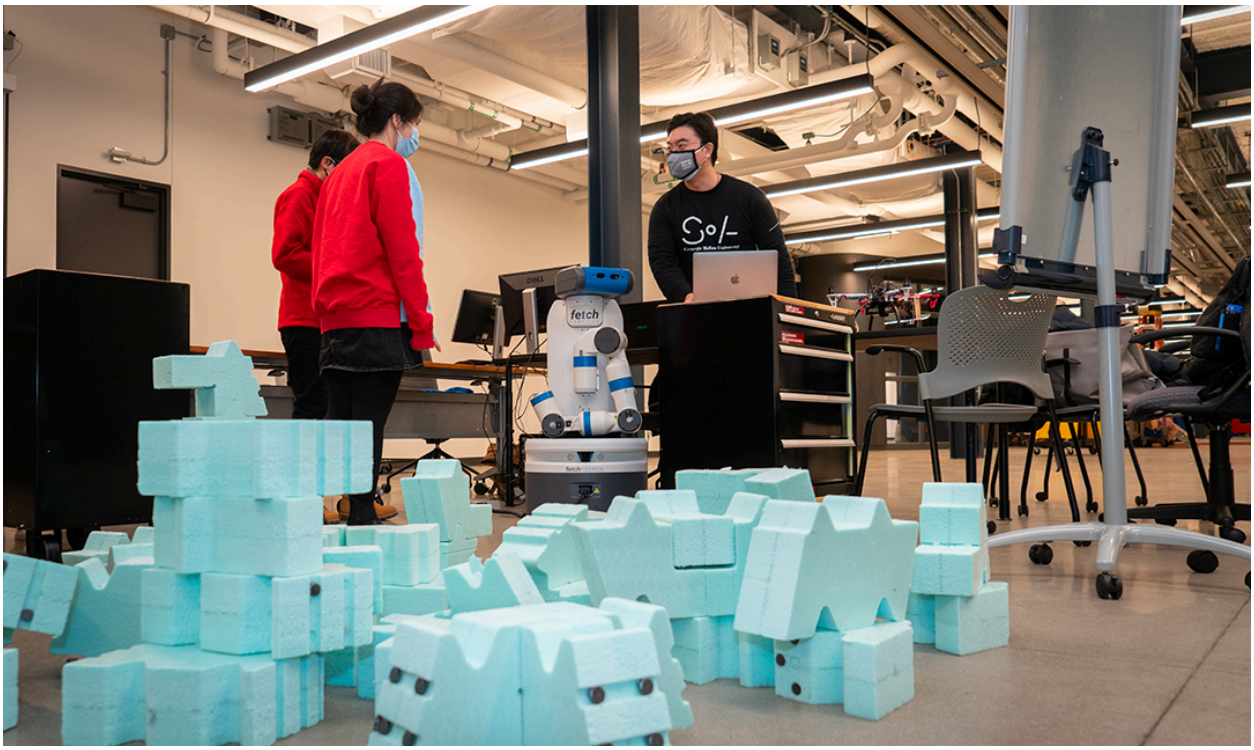



**Figure 50. ReAC project**
Image Source: https://engineering.cmu.edu/mfi/research/projects/rethinking-automation-in-construction.html

Research on using robots as assistants in construction is an important field that aims to make construction work more efficient and accessible through the use of robotics technology. By incorporating robots into construction processes, it is possible to enhance the productivity and capabilities of human workers, improving the efficiency and effectiveness of construction projects. As such, the use of robot assistants in construction is an important area of study that has the potential to significantly impact and improve the construction industry.

During my study, I was working for the ReAC project. The ReAC (Rethinking Automation in Construction) project aims to use recent advances in artificial intelligence and robotics to support cooperative construction. By studying construction practices on sites and in modular factories and conducting technical research into reinforcement learning and robotics, the project aims to create robot-in-the-loop systems that can enhance and safely interact with construction workers in both construction

sites and modular factories. Since the LiDAR device is a necessary component for the robot development, the visualization can also be incorporated into tasks performed by those robots.




**Figure 51. Storyboard of Automated construction assistant**
Image Source: Drawn by author.

This speculative design proposes the use of point cloud visualization as a component of robot-assisted construction. The goal of this approach is to use robots to assist human workers in construction tasks, and

a well-designed visualization interface can enhance the process of human-robot collaboration. In particular, tasks that require human decision-making can benefit from improved visualization, especially in situations where traditional cameras are unable to capture sufficient visual information. By incorporating point cloud visualization into the design of robot assistants, it is possible to improve the efficiency and effectiveness of construction projects, and facilitate more successful human-robot collaboration.

# 9. Conclusion

## 9.1. Contribution

The contribution of this project can be summarized to following points:

- Develops a new point-cloud visualization framework incorporating state-of-art point cloud techniques for augmentation
- Adds an user-centric perspective towards visualization tools development
- Thickens point cloud and visualization research by introducing recognizability analysis
- Demonstrates a feasible development workflow extending current technical-driven development to a further step of user study.
- Departs from a focus on ‘precision’ and instead places the perceptions and experiences of the user at the center of the validation and development processes

## 9.2. Limitation and Future Work

Due to the limitation of time and efforts, there are some room left within the research question for next step work.

First, no consideration is given to the person's background when conducting the user research. In other words, this study presupposes that all users have the same level of recognition skill. "There is abundant evidence in the graphics community that individual variations matter," Ottley says. These previous investigations emphasize a variety of personality qualities and cognitive talents that can modify visualization system usage and demonstrate a measurable effect on speed, accuracy, process, and

attention."[106]. For example, students with architecture backgrounds may show better cognition ability towards point cloud. However, any conclusions require further steps of detailed user study regarding to them.

Second, the research only looked at a small number of visualization approaches. Only two approaches, detection and upsampling, are implemented into point cloud visualization and evaluated with users in this study. Given that various approaches such as coloring and texturing can have an impact on visualization outcomes and cognition, this work plan will incorporate other strategies for technical development and user testing.

Finally, only web-based user interfaces are evaluated. However, the interface can also impact people's recognition. For example, the capacity of spatial recognition in a virtual world may differ from that of a web-interface. To obtain a whole view, it is vital to investigate several interactive mediums.

[106] Ottley, Alvitta. "Adaptive and personalized visualization." *Synthesis Lectures on Visualization* 7, no. 1 (2020): 1-117.

## bibliography


[1] Alexiou, Evangelos, Nanyang Yang, and Touradj Ebrahimi. "PointXR: A toolbox for visualization and subjective evaluation of point clouds in virtual reality." *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2020.

[2] Alexa, Marc, Johannes Behr, Daniel Cohen-Or, Shachar Fleishman, David Levin, and Claudio T. Silva. "Computing and rendering point set surfaces." *IEEE Transactions on visualization and computer graphics* 9, no. 1 (2003): 3-15.

[3] Biederman, Irving. "Recognition-by-components: a theory of human image understanding." *Psychological review* 94, no. 2 (1987): 115.

[4] Brown, Roger. "Words and things." (1958).

[5] Card, Mackinlay. *Readings in information visualization: using vision to thin*k. Morgan Kaufmann, 1999.

[6] Cafaro, Francesco, and Jessica Roberts. "Data through Movement: Designing Embodied Human-Data Interaction for Informal Learning." *Synthesis Lectures on Visualization* 8.3 (2021): 1-127

[7] Choy, Christopher, JunYoung Gwak, and Silvio Savarese. "4d spatio-temporal convnets: Minkowski convolutional neural networks." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019.

[8] Corner, James. *The agency of mapping: speculation, critique and invention*. na, 1999.

[9] Cosgrove, Denis E. *Social formation and symbolic landscape*. University of Wisconsin Press, 1998.

[10] da Silva Cruz, Luis A., et al. "Point cloud quality evaluation: Towards a definition for test conditions." *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2019.

[11] Dinesh, Chinthaka, Gene Cheung, and Ivan V. Bajić. "3D point cloud super-resolution via graph total variation on surface normals." In *2019 IEEE International Conference on Image Processing (ICIP)*, pp. 4390-4394. IEEE, 2019.

[12] Foix, Sergi, Guillem Alenya, and Carme Torras. "Lock-in time-of-flight (ToF) cameras: A survey." *IEEE Sensors Journal* 11, no. 9 (2011): 1917-1926.

[13] Gadelha, Matheus, Rui Wang, and Subhransu Maji. "Multiresolution tree networks for 3d point cloud processing." In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 103-118. 2018.

[14] Gallagher, Richard, and Tim Appenzeller. "Beyond reductionism." *Science* 284, no. 5411 (1999): 79-79.

[15] Geiger, Andreas, Philip Lenz, Christoph Stiller, and Raquel Urtasun. "Vision meets robotics: The kitti dataset." *The International Journal of Robotics Research* 32, no. 11 (2013): 1231-1237.

[16] Gobbetti, Enrico, and Fabio Marton. "Layered point clouds: a simple and efficient multiresolution structure for distributing and rendering gigantic point-sampled models." *Computers & Graphics* 28.6 (2004): 815-826.

[17] Graham, Benjamin, Martin Engelcke, and Laurens Van Der Maaten. "3d semantic segmentation with submanifold sparse convolutional networks." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 9224-9232. 2018.

[18] Grossman, Jeffrey P., and William J. Dally. "Point sample rendering." *Eurographics Workshop on Rendering Techniques*. Springer, Vienna, 1998.

[19] Guo, Yulan, Hanyun Wang, Qingyong Hu, Hao Liu, Li Liu, and Mohammed Bennamoun. "Deep learning for 3d point clouds: A survey." *IEEE transactions on pattern analysis and machine intelligence* 43, no. 12 (2020): 4338-4364.

[20] Huang, Hui, Dan Li, Hao Zhang, Uri Ascher, and Daniel Cohen-Or. "Consolidation of unorganized point clouds for surface reconstruction." *ACM transactions on graphics (TOG)* 28, no. 5 (2009): 1-7.

[21] Huang, Hui, Shihao Wu, Minglun Gong, Daniel Cohen-Or, Uri Ascher, and Hao Zhang. "Edge-aware point set resampling." *ACM transactions on graphics (TOG)* 32, no. 1 (2013): 1-12.

[22] Jia, Dan, and Bastian Leibe. "Person-MinkUNet: 3D Person Detection with LiDAR Point Cloud." *arXiv preprint arXiv:2107.06780* (2021).

[23] Keim, Daniel A., Florian Mansmann, Jörn Schneidewind, and Hartmut Ziegler. "Challenges in visual data analysis." In *Tenth International Conference on Information Visualisation (IV'06)*, pp. 9-16. IEEE, 2006.

[24] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." *nature* 521, no. 7553 (2015): 436-444.

[25] Li, Ruihui, et al. "Pu-gan: a point cloud upsampling adversarial network." *Proceedings of the IEEE/CVF international conference on computer vision*. 2019.

[26] Liu, Songbin. *3D Point Cloud Analysis : Traditional, Deep Learning, and Explainable Machine Learning Methods*. Cham, Switzerland: Springer, 2021. Print.

[27] Lin, Tzu-Chieh, Daniel Stanley Tan, Hsueh-Ling Tang, Shih-Che Chien, Feng-chia Chang, Yung-Yao Chen, Wen-Huang Cheng, and Kai-Lung Hua. "Pedestrian detection from lidar data via cooperative deep and hand-crafted features." In *2018 25th IEEE international conference on image processing (ICIP)*, pp. 1922-1926. IEEE, 2018.

[28] Lipman, Yaron, Daniel Cohen-Or, David Levin, and Hillel Tal-Ezer. "Parameterization-free projection for geometry reconstruction." *ACM Transactions on Graphics (TOG)* 26, no. 3 (2007): 22-es.

[29] Liu, Zhijian, Haotian Tang, Yujun Lin, and Song Han. "Point-voxel cnn for efficient 3d deep learning." *Advances in Neural Information Processing Systems* 32 (2019).

[30] Logothetis, Nikos K., and David L. Sheinberg. "Visual object recognition." *Annual review of neuroscience* 19, no. 1 (1996): 577-621.

[31] Lorente, Òscar, Josep R. Casas, Santiago Royo, and Ivan Caminal. "Pedestrian Detection in 3D Point Clouds using Deep Neural Networks." *arXiv preprint arXiv:2105.01151* (2021).

[32] Manovich, Lev. "What is visualization." *paj: The Journal of the Initiative for Digital*.–2010 (2010).

[33] Martín-Martín, Roberto, et al. "Jrdb: A dataset and benchmark for visual perception for navigation in human environments." *arXiv preprint arXiv:1910.11792* (2019).

[34] Navarro-Serment, Luis E., Christoph Mertz, and Martial Hebert. "Pedestrian detection and tracking using three-dimensional ladar data." *The International Journal of Robotics Research* 29, no. 12 (2010): 1516-1528.

[35] Mencl, R. "Reconstruction of surfaces from unorganized 3D points clouds (Doctoral dissertation, PhD Thesis, Dortmund University, Germany)." (2001).

[36] Ottley, Alvitta. "Adaptive and personalized visualization." *Synthesis Lectures on Visualization* 7, no. 1 (2020): 1-117.

[37] Pomerleau, François, Francis Colas, and Roland Siegwart. "A review of point cloud registration algorithms for mobile robotics." *Foundations and Trends® in Robotics* 4.1 (2015): 1-104.

[38] Premebida, Cristiano, Joao Carreira, Jorge Batista, and Urbano Nunes. "Pedestrian detection combining RGB and dense LIDAR data." In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 4112-4117. IEEE, 2014.

[39] Preiner, Reinhold, Oliver Mattausch, Murat Arikan, Renato Pajarola, and Michael Wimmer. "Continuous projection for fast L1 reconstruction." *ACM Trans. Graph.* 33, no. 4 (2014): 47-1.

[40] Purchase, Helen C., Natalia Andrienko, Thomas J. Jankun-Kelly, and Matthew Ward. "Theoretical foundations of information visualization." In *Information Visualization*, pp. 46-64. Springer, Berlin, Heidelberg, 2008.

[41] Qi, Charles R., et al. "Pointnet: Deep learning on point sets for 3d classification and segmentation." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

[42] Qi, Charles R., et al. "Pointnet++: Deep hierarchical feature learning on point sets in a metric space." *Advances in neural information processing systems* 30 (2017).

[43] Qiong, O. U. "A brief introduction to perception." *Studies in literature and language* 15, no. 4 (2017): 18-28.

[44] Remondino, Fabio. "From point cloud to surface: the modeling and visualization problem." *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 34 (2003).

[45] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." *International Conference on Medical image computing and computer-assisted intervention*. Springer, Cham, 2015.

[46] Rosch, Eleanor, Carolyn B. Mervis, Wayne D. Gray, David M. Johnson, and Penny Boyes-Braem. "Basic objects in natural categories." *Cognitive psychology* 8, no. 3 (1976): 382-439.

[47] Rusinkiewicz, Szymon, and Marc Levoy. "QSplat: A multiresolution point rendering system for large meshes." *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*. 2000.

[48] Samovar, Larry A., Richard E. Porter, Edwin R. McDaniel, and Carolyn Sexton Roy. *Communication between cultures*. Cengage Learning, 2016.

[49] Schlosser, Joel, Christopher K. Chow, and Zsolt Kira. "Fusing lidar and images for pedestrian detection using convolutional neural networks." In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2198-2205. IEEE, 2016.

[50] Spinello, Luciano, and Kai O. Arras. "People detection in RGB-D data." In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3838-3843. IEEE, 2011.

[51] Svalec, Marian, Lubos Takac, and Michal Zabovsky. "Performance enhancing of storage system for point cloud geographic data." *2015 10th System of Systems Engineering Conference (SoSE)*. IEEE, 2015.

[52] Su, Honglei, et al. "Perceptual quality assessment of 3D point clouds." *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019.

[53] Tang, Haotian, et al. "Searching efficient 3d architectures with sparse point-voxel convolution." *European conference on computer vision*. Springer, Cham, 2020.

[54] Tang, Hsueh-Ling, Shih-Che Chien, Wen-Huang Cheng, Yung-Yao Chen, and Kai-Lung Hua. "Multi-cue pedestrian detection from 3D point cloud data." In *2017 IEEE international conference on multimedia and expo (ICME)*, pp. 1279-1284. IEEE, 2017.

[55] Thorpe, Simon, Denis Fize, and Catherine Marlot. "Speed of processing in the human visual system." *nature* 381, no. 6582 (1996): 520-522.

[56] Tu, Chenxi, Eijiro Takeuchi, Alexander Carballo, and Kazuya Takeda. "Real-time streaming point cloud compression for 3d lidar sensor using u-net." *IEEE Access* 7 (2019): 113616-113625.

[57] Tu, Chenxi, Eijiro Takeuchi, Chiyomi Miyajima, and Kazuya Takeda. "Compressing continuous point cloud data using image compression methods." In *2016 IEEE 19th international conference on intelligent transportation systems (ITSC)*, pp. 1712-1719. IEEE, 2016.

[58] Urech, Philipp RW, et al. "Point cloud modeling as a bridge between landscape design and planning." *Landscape and Urban Planning* 203 (2020): 103903.

[59] Wang, Qian, and Min-Koo Kim. "Applications of 3D point cloud data in the construction industry: A fifteen-year review from 2004 to 2018." *Advanced Engineering Informatics* 39 (2019): 306-319.

[61] Wimmer, Michael, and Claus Scheiblauer. "Instant Points: Fast Rendering of Unprocessed Point Clouds." *PBG@ SIGGRAPH*. 2006.

[62] Wong, Richmond Y., and Vera Khovanskaya. "Speculative design in HCI: from corporate imaginations to critical orientations." In *New Directions in Third Wave Human-Computer Interaction: Volume 2-Methodologies*, pp. 175-202. Springer, Cham, 2018.

[63] Wu, Shihao, Hui Huang, Minglun Gong, Matthias Zwicker, and Daniel Cohen-Or. "Deep points consolidation." *ACM Transactions on Graphics (ToG)* 34, no. 6 (2015): 1-13.

[64] Yang, Yaoqing, Chen Feng, Yiru Shen, and Dong Tian. "Foldingnet: Point cloud auto-encoder via deep grid deformation." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 206-215. 2018.

[65] Yifan, Wang, Shihao Wu, Hui Huang, Daniel Cohen-Or, and Olga Sorkine-Hornung. "Patch-based progressive 3d point set upsampling." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5958-5967. 2019.

[66] Yin, Tianwei, Xingyi Zhou, and Philipp Krahenbuhl. "Center-based 3d object detection and tracking." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2021.

[67] Yu, Lequan, et al. "Ec-net: an edge-aware point set consolidation network." *Proceedings of the European conference on computer vision (ECCV)*. 2018.

[68] Yu, Lequan, et al. "Pu-net: Point cloud upsampling network." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.

[69] Yuan, Wentao, Tejas Khot, David Held, Christoph Mertz, and Martial Hebert. "Pcn: Point completion network." In *2018 International Conference on 3D Vision (3DV)*, pp. 728-737. IEEE, 2018.

[70] Zhang, Yan, Wenhan Zhao, Bo Sun, Ying Zhang, and Wen Wen. "Point Cloud Upsampling Algorithm: A Systematic Review." *Algorithms* 15, no. 4 (2022): 124.