Exploring the Possibilities of (Near) Real-Time Semantic Segmentation with 3D Point Cloud Data and Effective User-Centric Visualizations for First Responders

Master's Thesis

Author

D.R. van Kleef

Supervisors

ir. E. Verbree

ir. R. Voûte

Responsible Professor

Prof.dr.ir. P.J.M. van Oosterom

University

Delft University of Technology

Abstract

First responders operate in complex and dynamic indoor environments where accurate real-time spatial information is crucial for situational awareness and decision making. The aim of this study is to explore the possibilities of (near) real-time segmentation of 3D point cloud data using state-of-the-art deep learning models and evaluate different visualization techniques to improve the situational awareness of first responders in indoor environments.

The study first evaluates nine segmentation models, considering key factors such as accuracy and inference speed. Although models like Point Transformer V3 + PPT and Point-SAM achieve high segmentation accuracy, real-time performance remains a challenge, particularly on personal devices. Applying these models to self-acquired point cloud data revealed not just preprocessing needs, but deeper compatibility issues. In practice, the entire point cloud was labeled as "clutter," likely due to a combination of model limitations and the author's limited programming experience, highlighting key barriers to real-world deployment.

Beyond segmentation, this research applies cartographic principles and cognitive theories to develop visualization techniques for effectively communicating the segmented point cloud data. Although the concept of (near) real-time segmentation serves as a guiding principle for this research, the achievement of (near) real-time segmentation of self-acquired point cloud data was not achieved. To address this, the self-acquired point clouds were manually segmented to approximate the expected model output, allowing the evaluation of different visualization techniques. Several proof-of-concept visualizations were created, testing different color schemes and levels of detail to assess their impact on interpretability and situational awareness. The findings indicate that structured visualizations, particularly those using functional color schemes, which assign colors according to their functional significance (e.g., green for floors, yellow for doors, red for hazards, dark gray for barriers), improve situational awareness and decision-making, whereas excessive complexity hinders usability in high-pressure scenarios.

The results highlight the progress of deep learning in indoor segmentation while also emphasizing the need for improved data integration and processing workflows. Furthermore, a single structured visualization approach was preferred over role-specific adaptations, reinforcing the importance of clarity, simplicity, and user-centered cartographic design in operational contexts. In the future, optimizing real-time processing and refining visualization techniques will be essential to improve situational awareness for first responders in critical situations.

Acknowledgments

For the past seven months, I have worked on this thesis, which is now finally in front of you. Although this project has been as much of a challenge as it has been rewarding, it has allowed me the opportunity to explore a topic that was totally new to me. Throughout this project, I have met many great people who made the process both insightful and enjoyable. They have provided me with invaluable support for everything from writing code, sharpening arguments, designing visualizations, and expanding my understanding of the subject.

First and foremost, a special thank you to my supervisors Edward Verbree (TU Delft) and Robert Voûte (CGI) for their excellent guidance, constructive feedback, and unwavering support. Their expertise and support have proved invaluable in navigating the complexities of this research. Additionally, I would like to thank Peter van Oosterom (TU Delft) for his supervision and insightful feedback as the responsible professor.

I thank my colleagues and fellow students at CGI, especially Jop Smeets, Bart-Peter Smit, and Algan Yasar, for inspiring discussions, ideas, and technical advice during my work. Moreover, I would like to express my gratitude to CGI, where I carried out this research project for the environment and setting to generate this work, as well as for granting me access to all the necessary resources to complete this thesis.

In addition, I acknowledge the assistance of OpenAI's ChatGPT, which helped improve the readability and clarity of this thesis through proofreading and language suggestions.

And last but not least, I thank my family, girlfriend, friends, and fellow students for their constant support and encouragement. This has provided me with motivation all along, knowing that they believed in me.

Thanks to everyone who made it possible.

Reinier van Kleef - Rotterdam 25-03-2025

Contents

Acronyms

2D	Two Dimensional
3D	Three Dimensional
AI	Artificial Intelligence
AR	Augmented Reality
CaCO	Calamity Coordinator
CNN	Convolutional Neural Networks
COP	Common Operational Picture
CV	Computer Vision
DL	Deep Learning
FCN	Fully Convolutional Networks
GNSS	Global Navigation Satellite Systems
IoU	Intersection over Union
LiDAR	Light Detection and Ranging
mIoU	Mean Intersection over Union
MLP	Multi-Layer Perceptron
MRG	Multi-Resolution Grouping
MSA	Multi-Head Self-Attention
MSG	Multi-Scale Grouping
OvD	Officer of Service
PCA	Principal Component Analysis
PPT	Position Pooling Transformer
PTv3	Point Transformer V3
RGB	Red, Green, Blue
RF	Random Forest
S3DIS	Stanford Large-Scale 3D Indoor Spaces
SA	Situational Awareness
SAGAT	Situational Awareness Global Assessment Technique
SART	Situational Awareness Rating Technique
SfM	Structure from Motion
SLAM	Simultaneous Localization and Mapping
SPCs	Smart Point Clouds
TTA	Test Time Augmentation
ViT(s)	Vision Transformer(s)

Chapter 1

Introduction

In recent years, advances in Computer Vision (CV) and Artificial Intelligence (AI) technologies, such as convolutional neural networks (CNN) and transformer-based models, have significantly improved the accuracy and efficiency of indoor semantic segmentation (Long et al., 2015; Sultana et al., 2020). These advancements, in combination with the rise of more affordable technologies, such as low-cost cameras, drones, and even LiDAR-equipped iPhones, have brought disaster relief and emergency response to the brink of being completely transformed by these technological innovations (Oh et al., 2019).

A heavily relied upon process that makes this transformation possible is that of semantic segmentation, which, simply put, consists of assigning a semantic label or class to each pixel of an image (or point in a point cloud). In addition to labeling objects, this process adds a semantic layer to the original data, which enriches the existing data with semantic information, creating so called smart point clouds (Poux et al., 2016; Yu et al., 2018). In the 1970s, in the early days of semantic segmentation, it was a traditional computer vision task that followed manual extraction and selection-based methods that relied on hand-crafted features, edge detection algorithms, and region-based techniques. However, these early approaches were limited in scope and accuracy due to the challenges of manually defining features (Ohta et al., 1978).
Recent advances in deep learning changed this workflow, as illustrated in Figure 1.1, which compares a more traditional CV workflow to that of the modern deep learning approach. In traditional workflows, raw input data (such as an image or point cloud) are first processed through feature engineering, where human experts manually extract and select relevant features. These features serve as structured input to a shallow architecture classifier, which then predicts an output label, such as object categories or segmentation masks.
In contrast, deep learning models, particularly Convolutional Neural Networks (CNNs), remove the need for manual feature extraction. Instead, they automatically learn hierarchical representations from raw input data, identifying patterns across multiple layers. This allows the model to capture complex relationships and spatial structures, resulting in a more robust and scalable approach. In this case, the final output is a refined prediction (such as segmented regions in an image or classified objects in a 3D point cloud) generated directly from the learned feature representations. This transition allowed for the building of deeper and more powerful architectures, like Fully Convolutional Networks (FCNs), which provide better pixel- or pointwise predictions. Consequently, CNN-based models have defined a new state-of-the-art in semantic segmentation because of their improved accuracy and performance in large-scale tasks.

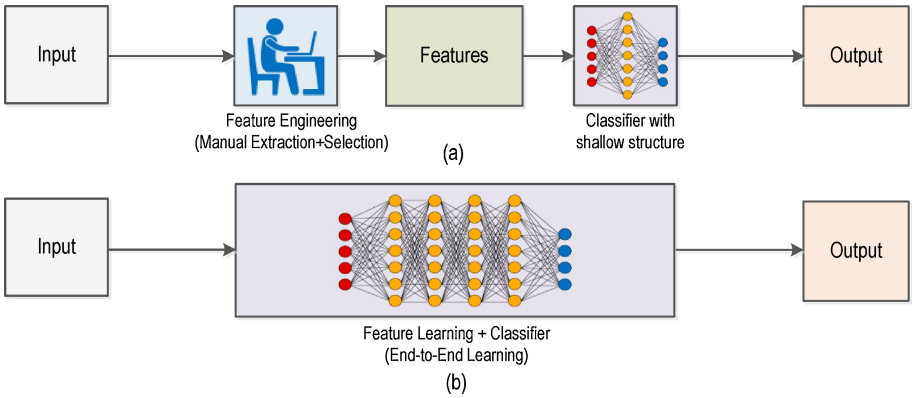


Figure 1.1: (a) Traditional computer vision workflow (b) Deep learning workflow (Wang et al., 2018).

Furthermore, the recent rise of transformer-based methods has greatly improved the semantic segmentation performance on indoor datasets like S3DIS and ScanNet compared to more traditional CNN-based models. Vision Transformers (ViTs), a type of transformer-based model with self-attention mechanisms, allow the model to capture long-range dependencies within an image or a point cloud. This mitigates one of the major limitations of CNNs that can occur in practice, where the global context must be modeled but is inefficient with CNNs. As a result, transformer-based approaches excel in segmenting complex and large-scale scenes, where understanding the relationships between distant pixels or points is essential (Zhang et al., 2022).
Moreover, ViT-based models like PointTransformer V3 + PTT, MaskFormer, and Superpoint Transformer simplify the segmentation process by removing the need for task-specific elements (like non-maximum suppression, predefined anchor boxes) that are commonly used in CNN-based models. These components usually need to be manually tuned and introduce complexity into the segmentation pipeline (Cheng et al., 2022; Robert et al., 2023; Wu et al., 2024). On the other hand, Vision Transformer-based models use self-attention mechanisms that inherently model spatial relationships across the entire scene, without the need for these additional components. This simplification results in more flexible, scalable, and efficient models for varying segmentation tasks by increasing accuracy and reducing computational overhead (Strudel et al., 2021).

In the current era of semantic segmentation, labels can represent common semantic categories or identify pixels as parts of separate objects. This workflow transformation allows (near) real-time scene understanding and paves the way for new applications, such as live data visualizations for disaster relief and emergency response (Wang et al.,2018). For this research, (near) real-time processing is defined as a 30-second threshold, as this allows for rapid situational awareness (SA) while accounting for the computational complexity of segmenting large 3D point cloud datasets. This (near) real-time visualizations help first responders instantly assess critical environments, helping them quickly identify hazards, navigate complex indoor spaces, and make informed decisions on the fly. By continuously processing and visualizing segmented 3D point cloud data in (near) real-time, these systems, whether visualized in a GIS, augmented reality, or a game engine, provide clear insights and significantly improve the efficiency and effectiveness of emergency operations (Li et al., 2014).
This (near) real-time visualization for first responders attempts to increase SA. The concept of SA refers to the ability of individuals and teams to perceive, comprehend, and anticipate environmental changes (Kapucu and Garayev, 2011). Through (near) real-time visualization, decision makers can make data-informed decisions in inherently dynamic environments. This could ultimately improve responsiveness, resource allocation, and overall mission success in high-stakes situations.

In order to optimally increase the SA for first responders, they must be able to quickly interpret the segmented point cloud. Therefore, it is important to keep the visualizations of these segmented point clouds clear and simple, ensuring that critical information is immediately recognizable without overwhelming the user. By focusing on clarity and usability, these visualizations can effectively support rapid decision making under pressure, allowing first responders to focus on their tasks rather than deciphering complex data representations (Endsley et al., 2003).

Two important questions arise from this. How can segmented 3D point cloud data be effectively visualized, and how can these visualizations be made usable for different kinds of users? Addressing these challenges involves not only rendering complex, high-density point clouds but also the process of selecting visual characteristics that are simple and immediately understandable for first responders. By overcoming these challenges, there is an opportunity to create innovative visual solutions that improve decision making in critical scenarios.
Therefore, the uniqueness behind this research is not just about how to visualize complex, high-density segmented point clouds in (near) real-time but also in selecting appropriate visual characteristics that can be easily interpreted by different kinds of users under stress. It is important to keep the following phrase in mind: "How do I say what to whom and is it effective?" This phrase can be broken down into four keywords:

- **How**: The visualization through cartographic methods and techniques.
- **What**: The (near) real-time segmented point cloud data.
- **Whom**: The map audience (first responders) and the purpose of the map.
- **Effective**: Does the segmented point cloud provide the information it needs to provide to increase situational awareness?

The art of map making involves applying these four keywords to ensure that visualizations are clear, purpose-driven, and user-centered to communicate critical information in a way that users in the field, such as first responders, can quickly interpret and act upon during high-pressure situations (van der Meer, 2018). Effective visualization in (near) real-time needs to balance simplicity and detail. Key elements, such as walls, doors, and potential hazards, must be easily distinguishable, while non-essential information should be excluded (Schmidt and Götze, 1998). Additionally, visualizations must be fine-tuned for different types of users. Selecting the right visualization is crucial, as it directly impacts decision-making. Poor visualization choices can reduce the effectiveness of (near) real-time visualization, ultimately affecting the overall outcome of an emergency.

Chapter 2

Research Objectives

The objective of this research is to further explore the possibilities of (near) real-time indoor semantic segmentation by using an AI-driven model to process and visualize 3D point cloud data generated by an iPhone 11 Pro equipped with a LiDAR scanner. However, the aim of this research is not only to segment the point cloud data in (near) real-time, but also to explore how the segmented point cloud can be effectively visualized to increase SA for first responders. By addressing both technological and user-centered challenges, this research seeks to answer the following research question.

2.1 Main Research Question

To what extent is it possible to segment (near) real-time 3D point cloud data using an existing deep learning model, and how can this segmented point cloud be effectively visualized to increase situational awareness for first responders?

To answer the main research question, the research is divided into three distinct phases. The first phase focuses on researching whether it is possible to use an existing semantic segmentation model to segment point clouds in (near) real-time on self-gathered data from an iPhone 11 Pro equipped with a LiDAR scanner. The second phase of this research focuses on how to effectively visualize the segmented point cloud to increase SA for first responders. Finally, the last phase involves evaluating the effectiveness of the different visualizations of the segmented point cloud. Throughout the three phases, partial findings from literature studies, model integration, interviews, and the developed proof-of-concept are generated, facilitating an iterative refinement of the research and gradually addressing the main research question.

2.2 Research Sub-Questions

2.2.1 Phase 1: Model Selection and Integration

The first phase of this research focuses on exploring theoretical concepts and technical aspects related to computer vision, smart point clouds, and semantic segmentation. The primary objective of this phase is to identify a suitable existing semantic segmentation model capable of performing (near) real-time segmentation, implement the selected model on a personal device, and evaluate its performance on the S3DIS dataset and self-acquired data collected with an iPhone 11 Pro equipped with a LiDAR scanner.

To address the first sub-question, a comprehensive review of current state-of-the-art semantic segmentation models is conducted. This review prioritized models based on their ability to achieve (near) real-time segmentation while maintaining high segmentation accuracy. Following a detailed comparison of the identified models, the first sub-question is answered.

Sub-Question 1: Which existing deep learning segmentation model is best suited for (near) real-time point cloud segmentation in indoor environments?

After selecting one of the existing segmentation models, the chosen model is implemented on a personal device and evaluated using S3DIS training data to validate its performance. The goal is not only to successfully integrate the model, but also to replicate the accuracy reported by the authors on the same

dataset (S3DIS). This results in the following sub-question:

Sub-Question 2: To what extent is it possible to integrate the chosen segmentation model on a personal device and reproduce the accuracy results reported by the authors on the training data?

To complete the first phase, the model is further tested on self-collected data. This step involves preprocessing the data, ensuring compatibility with the model's requirements, and conducting segmentation to evaluate its robustness under real-world conditions. The focus is on assessing the adaptability of the model to less controlled scenarios, such as those represented by self-acquired datasets, and understanding its limitations. This leads to the final sub-question of the first phase:

Sub-Question 3: How can the selected segmentation model be integrated and tested on self-acquired data?

2.2.2 Phase 2: Visualization and Interaction

After completing the first phase, the focus shifts towards the visualization of the segmented point cloud. In this phase, 3D visualizations are used for mapping and displaying classes, with each class being represented by specific colors based on interactions/interviews with the first responders and on scientific literature. Multiple visualizations are made; the focus here lies in how each user interacts with the different visualizations. This leads to the fourth, fifth, and sixth sub-question:

Sub-Question 4: What are the key information needs during emergencies to support decision making and situational awareness for various stakeholders?

Sub-Question 5: How can scientific principles and cartographic methods be applied to effectively communicate segmented 3D point cloud data to diverse user groups?

Sub-Question 6: How can a proof-of-concept visualization be developed to demonstrate the effective communication of segmented point cloud data, tailored to different user needs?

2.2.3 Phase 3: Reflection and Evaluation

After the implementation of the model and the visualization of the segmented point cloud, the last phase focuses on reflecting on the quality of the different visualizations. This involves evaluating the effectiveness of the visualizations. The reflection also explores how different types of users, control room personnel, local responders, and CoPI commanders interpreted and interacted with the different visualizations. The results collected will help identify strengths, limitations, and opportunities for further improvement. This leads to the final sub-question:

Sub-Question 7: How effective are the proof-of-concept visualizations in increasing situational awareness through the communication of segmented point cloud data?

2.3 Scope

2.3.1 Scope

The scope of this research focuses on selecting, integrating and evaluating an existing semantic segmentation model for (near) real-time 3D point cloud data within indoor environments. Since it was not possible to segment self-acquired data, this first part remains largely theoretical. Thereafter, the focus shifts towards analyzing how to effectively visualize a segmented point cloud through specified visualizations for different types of users.

2.3.2 Scope limitations

The following points are beyond the scope of this research:

- **Model improvement:** The research will not address the improvement of the architectures or algorithms of the segmentation models. It will only integrate a pre-existing model onto the author's device.
- **Drift in point cloud data:** Although drift may occur during the data collection process when collecting data for sub-question 3, this research will not attempt to mitigate or correct drift.
- **Outdoor Environments:** The research will focus exclusively on indoor environments.
- **Integration in the CGI architecture:** This research does not address the integration of the segmentation model into the CGI architecture. The focus remains on testing and evaluating the model independently, without considering its full integration into the broader system architecture.
- **Focus on observer:** The research focuses only on the visualization for the observer, not on the visualization for the explorer (data gatherer).
- **LiDAR data:** This research focuses solely on 3D point cloud data generated by an iPhone Pro 11 equipped with a LiDAR scanner. Other methods for generating point clouds, such as Structure from Motion (SfM) or photogrammetry, are outside the scope of this research.
- **(Near) real-time segmentation:** While the concept of (near) real-time segmentation serves as a guiding principle for the research, the achievement of (near) real-time segmentation of point cloud data is considered out of scope due to time constraints and technical limitations.

2.4 Research Structure

Figure 2.1 provides a conceptual overview of the research structure. The research follows a structured approach, beginning with the identification phase, where a literature review is conducted to establish the research problem and objectives. Based on this foundation, the main research question is formulated, which is further divided into sub-questions to answer the main research question. Following this, the methods phase involves the execution of the research methodology, applying literature research, model testing, data processing, visualization development, and expert interviews to generate findings for each sub-question. These findings form the basis for the evaluation phase, where the results are analyzed, conclusions are drawn, and the main research question is answered. The study concludes with a discussion that reflects on the findings and their implications.

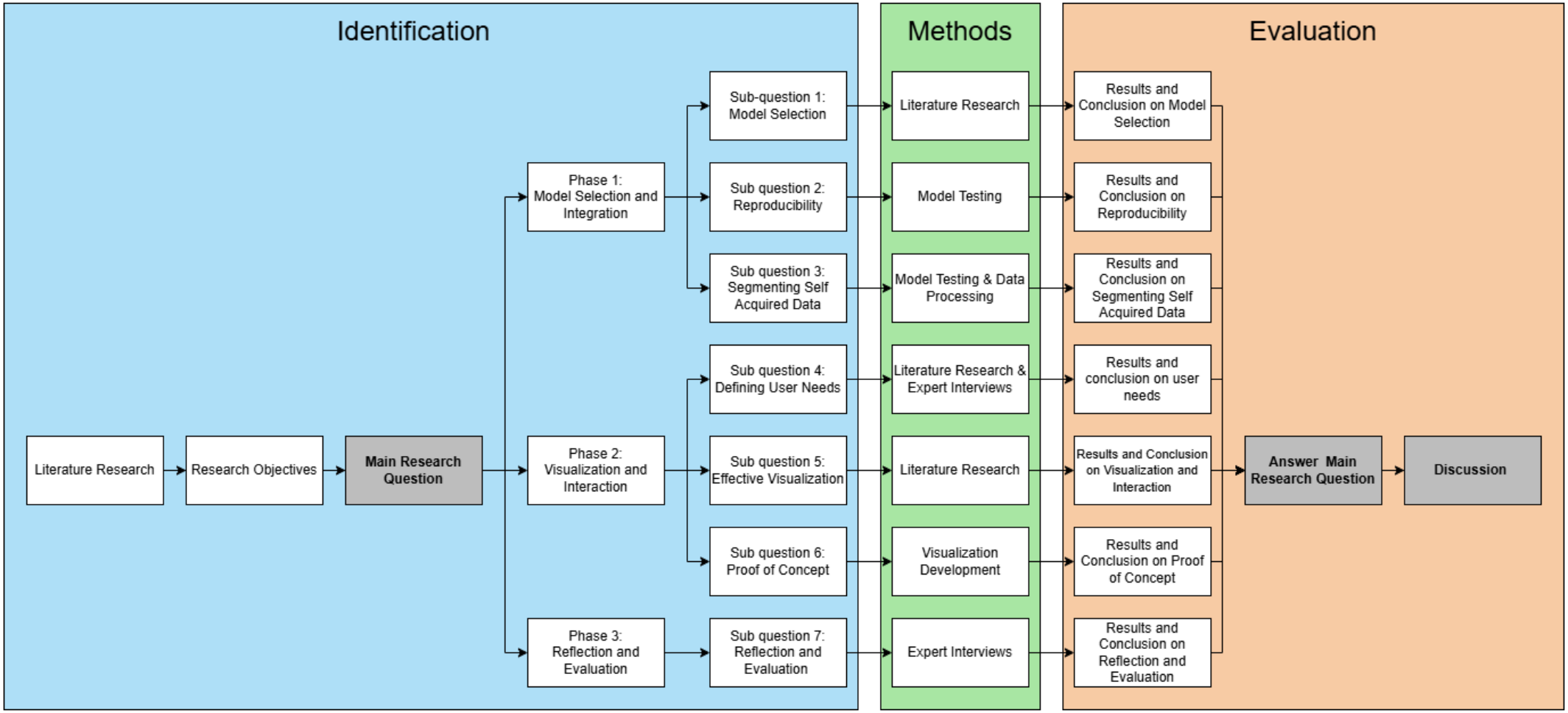


Figure 2.1: Research Structure

Chapter 3

Related Work

The related work section sets the groundwork for this research by reviewing the relevant academic literature in order to contextualize the main research question. First, it explores the basics of computer vision, various segmentation models, and methods to address the first sub-question. Building on this exploration, the related work section further investigates the principles of cartography, SA, and different visualization techniques.

3.1 Computer Vision and Segmentation

The first section introduces computer vision, different types of segmentation, various segmentation approaches, and different segmentation models. It includes more traditional methods such as CNN-based models as well as more advanced techniques using graph-based and transformer-based models.

3.1.1 Computer Vision

The multidisciplinary scientific field of computer vision focuses on enabling computers to interpret and make decisions based on visual data from the real world, it aims to provide machines with human-like vision capabilities (Khan and Al-Habsi, 2020). The concept of creating a system that simulates the human brain led to the initial development of neural networks. In the 1940s, McCulloch and Pitts (1943) sought to understand how the brain processes and recognizes complex patterns through the interaction of interconnected basic cells, known as neurons. For humans, identifying specific objects through subtle patterns in shape and light is an easy task. However, for computers, understanding the visual world with all its complexity is a tough job. Using neural networks, computers can begin to interpret visual data by mimicking the way the human brain processes complex information (Voulodimos et al., 2018).

This computational approach to visual perception has led to the development of structured frameworks to process images and extract meaningful information. Forsyth and Ponce (2002), for example, introduce three stages of vision processing.
First, there is the early vision. This stage focuses on low-level tasks such as feature extraction and stereopsis (perception of depth). The primary goal of this stage is to process raw visual input data and extract basic information, such as edges, textures, and depth, from stereo images, helping to build a representation of the visual scene.
The second stage, mid-level vision, focuses on tasks like segmentation and tracking. The goal of this stage is to group and organize the raw features extracted during the early vision into meaningful structures. A key task in mid-level vision is image segmentation, where similar regions of an image are clustered under meaningful labels. The output of mid-level vision provides a structured representation of objects and regions in the scene, making it easier for high-level vision to process and interpret (Forsyth and Ponce, 2002).
The third stage of vision processing is high-level vision. This stage involves more abstract tasks such as registration and recognition. In this stage, the system interprets and associates specific meanings or labels with objects in the scene, including identifying objects or patterns and aligning them with predefined models. The goal of this high-level vision is to interpret and recognize the content of the visual data. This ultimately tries to understand the scene at the conceptual level, enabling actions such as

identification, scene understanding, and decision making based on visual input (Forsyth and Ponce, 2002).

The integration of deep learning (DL) has been of significant importance in the advancement of CV. Traditional CV algorithms heavily relied on manual extraction and selection using handcrafted features and rules to process images as can be seen in Figure 1.1 (Wang et al., 2018). However, with the integration of DL, CV has seen remarkable progress, made possible by automatically learning visual features from raw data, making tasks such as image classification, object detection, and facial recognition more accurate and efficient. Due to this progress, CV plays a key role in modern technology such as real-time traffic management, autonomous driving, smart surveillance systems, and vision systems to inspect and classify fruits and vegetables (Bhargava and Bansal, 2021; Janai et al., 2020; Osman et al., 2017).

3.1.2 Segmentation

Segmentation serves as one of the tasks within the broader field of CV, it involves separating visual data (whether 2D images or 3D point clouds) into distinct regions, allowing computers to better understand and interpret complex scenes (Minaee et al., 2021). The primary goal of segmentation is to accurately classify or label each pixel in an image or a point in a point cloud, thus facilitating the identification and understanding of various elements within a scene. By assigning labels to pixels or points, segmentation helps to make more precise analysis and decision making in a wide range of applications (Minaee et al., 2023. Through this segmentation process, machines can extract critical information from complex visual data.

Segmentation can be categorized into three main types as can be seen in Figure 3.1. The first type of segmentation is semantic segmentation; hereby each pixel is classified into semantic categories. However, it does not distinguish between different instances of the same object. In figure 3.1 b, all cars are colored blue, regardless of their individual identity, with this form of segmentation, no distinction is made between the different cars.
The second type of segmentation is instance segmentation. Previously, each pixel was assigned to only one class. However, with instance segmentation, a distinction is made between pixels that belong to the same object class but represent different instances of that object, as can be seen in Figure 3.1 c.
The more recent form of segmentation is panoptic segmentation, which combines both semantic and instance segmentation approaches. Panoptic segmentation provides per-pixel class labels along with instance-level distinctions as can be seen in Figure 3.1 d. (Kirillov et al., 2019).

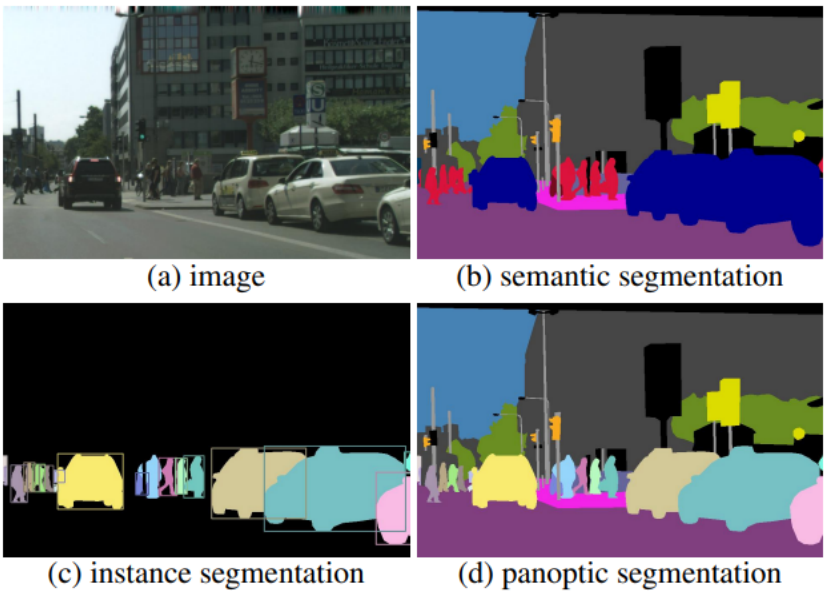


Figure 3.1: (a) image, ground truth (b) semantic segmentation, per-pixel class labels (c) instance segmentation, per-object mask and class label (d) panoptic segmentation, per-pixel class plus instance labels (Kirillov et al., 2019).

Smart Point Clouds

Building upon segmentation, the concept of Smart Point Clouds (SPCs), introduced by Poux et al. (2016) emerges as a innovative concept for managing, interpreting, and using point clouds. SPCs integrate semantic, geometric, and topological information directly into the structure of the point cloud. SPCs have therefore emerged to overcome the insufficiencies of traditional point clouds by embedding intelligence directly into the point cloud, allowing applications that require high levels of automation, accuracy, and domain-specific adaptability.

SPCs extend segmentation techniques by integrating enriched attributes that improve the usability of point clouds across various applications. The concept of SPCs can be divided into four key characteristics. First, there is the Semantic Enrichment, each point or cluster within the point cloud is augmented with metadata and classifications that describe its attributes and relationships. For example, in indoor environments, SPCs can label structural elements such as walls and doors, as well as movable objects such as a desk chair. This semantic layer allows for deeper context-driven analysis (Poux et al., 2016).
Second, SPCs are interoperable and modular. SPCs are designed to integrate easily across different workflows and domains, making them highly adaptable for use in fields such as urban planning, robotics, and emergency response. Their modular architecture facilitates extensions and customization to meet specific application requirements (Poux, 2019).
Third, SPCs are advanced structurations. Using hierarchical data structures such as octrees and voxels, SPCs efficiently manage large datasets, ensuring that queries and analysis can be performed at high speeds. These structures preserve the spatial and topological relationships inherent in the data, making SPCs ideal for real-time applications.
Lastly, automation and scalability of SPCs. Using machine learning algorithms, SPCs automate segmentation, classification, and clustering processes, reducing the dependence on manual innervation. This capability is particularly advantageous for large-scale datasets, such as urban or architectural scans.
By incorporating these characteristics into point cloud structures, SPCs enhance the functionality of traditional segmentation workflows, enabling more accurate, scalable, and context-aware analyses. They address challenges such as noise, occlusion, and varying point densities, ensuring robust performance across diverse applications. Their ability to integrate domain-specific knowledge further expands their potential, making SPCs a cornerstone in intelligent decision-making systems (Poux et al., 2016).

Furthermore, SPCs increase the utility of segmentation techniques by adding semantic, geometric, and topological intelligence to the point cloud structure. This integration transforms raw spatial data into actionable datasets that support advanced analysis and decision making in various applications. SPCs leverage segmentation to combine semantic and instance-level information, and with that address the key limitations of traditional workflows.
According to Poux (2019) the enriched capabilities of SPCs extend their applicability across several domains, making them integral to decision-making workflows. In disaster management, SPCs could improve SA by identifying critical infrastructure elements, such as emergency exits or structural vulnerabilities, allowing first responders to act quickly and effectively. In autonomous navigation, the semantic and topological real-time data provided by SPCs improve the accuracy and reliability of navigation systems for robotics and vehicles. Similarly, in urban planning, SPCs support the extraction of detailed environmental insights, such as building footprints and road networks, helping to design and develop infrastructure (Poux, 2019).
By providing structured, semantically enriched data, SPCs also improve operational efficiency and accuracy. They streamline data processing workflows, reduce errors associated with manual interpretation, and enable real-time analysis and visualization. These capabilities are particularly beneficial in dynamic environments, where timely and accurate decision making is critical. Thus, SPCs represent a transformative advance in the use of 3D spatial data for high-stakes applications (Poux et al., 2016).

Despite their potential, SPCs face challenges related to computational complexity, standardization, and scalability. Developing efficient algorithms to integrate semantic intelligence, improving indexing methods for large-scale datasets, and standardizing SPC frameworks for interoperability are critical areas for future research. According to Poux (2016) addressing these challenges will further unlock the potential of SPCs across domains and applications.
By extending the capabilities of segmentation techniques, SPCs redefine the processing and application of 3D spatial data. They embed semantic, geometric, and topological intelligence directly into point cloud structures, enabling advanced workflows that integrate segmentation outputs with actionable insights. From disaster management to autonomous navigation and urban planning, SPCs facilitate enhanced con-

textual understanding, operational efficiency, and real-time analysis. Their modular and interoperable design supports domain-specific adaptations, ensuring their scalability and utility across diverse fields. As SPCs continue to evolve, they are poised to become a cornerstone of intelligent decision-making systems, addressing the demands of increasingly complex and dynamic environments (Poux, 2019).

3.1.3 Image Segmentation

Building on the concept of Smart Point Clouds, which enhance segmentation with enriched attributes, and with a clear understanding of what semantic segmentation is and its role in classifying pixels or points into distinct categories, it is essential to explore the various methods used to achieve this task. Over time, several approaches have been developed, ranging from more traditional techniques to more advanced deep learning techniques.

Traditional Techniques

First, there are the more traditional techniques, such as region proposals, which categorize pixels according to their inherent properties, such as color, intensity, and texture. By recognizing regions with similar characteristics, region-based segmentation attempts to produce logical and meaningful segments within an image (Karthick et al., 2014). Although the basic principles can be dated back to the 1970s and 1980s, region-based segmentation techniques are still widely used and continue to serve as a valuable component in many modern applications. Another more traditional method is the graph cut algorithm; this method can be traced back to the late 1990s and early 2000s, but is also still used today. This approach involves modeling the image as a graph, where pixels (or superpixels) are represented as nodes, and edges represent the similarity between neighboring pixels (Boykov and Funka-Lea, 2006).
Another more traditional approach that gained popularity in the early 2010s involves the use of handcrafted geometric features in combination with classical machine learning classifiers. A common example is Principal Component Analysis (PCA) for local feature extraction followed by random forest classification (RF). In this approach, geometric features are derived from the eigenvalues of the local neighborhood of a point, capturing structural signals such as planarity and linearity. These derived features are then used to train an RF classifier that assigns semantic labels to each point based on local geometry (Poux, 2019; Wolf et al., 2015). This technique has proven effective in recognizing large structural components in 3D point clouds, such as floors, ceilings, and walls, due to their distinct geometric characteristics. However, the approach shows significant limitations in cluttered environments and for detecting smaller and critical objects such as fire extinguishers or exit signs. These objects often lack clear geometric signatures and require contextual understanding or color-based reasoning, which traditional pipelines do not capture. Although Bassier et al. (2019) achieved impressive precision and recall rates above 85% for major structural classes using RF, their experiments also showed that a large portion of indoor patches, especially smaller or cluttered ones, could not be reliably classified, often misidentified as walls or ceilings. Furthermore, Nurunnabi et al. (2016) demonstrated that PCA-based descriptors are highly sensitive to noise and outliers, which can significantly degrade performance in cluttered indoor environments of the real world. Therefore, although useful as a baseline, traditional techniques such as PCA with RF are not suitable for applications that require rich feature representation, fine-grained classification, or high segmentation accuracy under real-time constraints. More modern approaches that can be used for (near) real-time segmentation include CNN-based approaches, updated graph-based approaches, and transformer-based approaches. These more modern approaches are discussed in greater detail in the following sections.

CNN-based Semantic Segmentation

CNN-based semantic segmentation models have become pretty common in the field of deep learning. They are widely used for various computer vision tasks, including image classification, object detection, and image generation. CNN models usually include three types of layers (Garcia-Garcia et al., 2017).
First, there are convolutional layers; these layers use a set of filters (or kernels) to perform convolutions over the input data, extracting important features such as edges or textures from the images or point clouds. The whole input image is transformed into feature maps; these feature maps are the output of the convolutional layers and represent the presence of specific patterns or features detected by the filters

in different regions of the image (Yu et al., 2014).
Second, there are the nonlinear layers; these layers apply an activation function (typically element-wise) to the feature map, allowing the network to capture and model complex, nonlinear relationships in the data. Without these layers, the network would only be able to combine features in a linear way, which limits its ability to capture the complexities of the images or point clouds. (Minaee et al., 2021).
Third, there are the pooling layers; these layers reduce the spatial resolution of the feature maps by summarizing small regions within the data using average pooling and max pooling (Yu et al., 2014. The operation performed by the pooling layers is often referred to as subsampling or downsampling because it reduces the size of the data, resulting in some loss of information. However, this reduction is necessary for the network because it reduces the total computational load for subsequent layers and helps to prevent overfitting (Voulodimos et al., 2018).
After following several layers of convolutional and pooling, high-level reasoning in the neural network is carried out by fully connected layers, where each neuron is connected to all activations from the previous layers. This ultimately contributes to making predictions in the pixel or region that generate the segmented output (Voulodimos et al., 2018).

Figure 3.2 by Minaee et al. (2021) shows a general architecture of a standard CNN, showing the different stages of processing an image to classify or recognize features.
First, there is the input, in this case an image of a fingerprint. The first block of green squares represents the convolutional layers. Here, small filters (or kernels) scan over the input image to extract important features such as edges, textures, or patterns. These extracted features are represented by the green layers. The next block, shown in blue, represents the pooling layers. By using pooling layers, the size of the feature maps are reduced by summarizing smaller regions. Thereafter, the figure shows a second convolution and pooling layer by repeating the process with another set of convolutional and pooling layers. More complex and abstract features are extracted from the initial patterns, further refining the important parts of the image. Finally, there are fully connected layers; at this stage, all the extracted features are combined and processed to make a final prediction or classification. The fully connected layer takes the high-level features and determines the final output (Minaee et al., 2021).

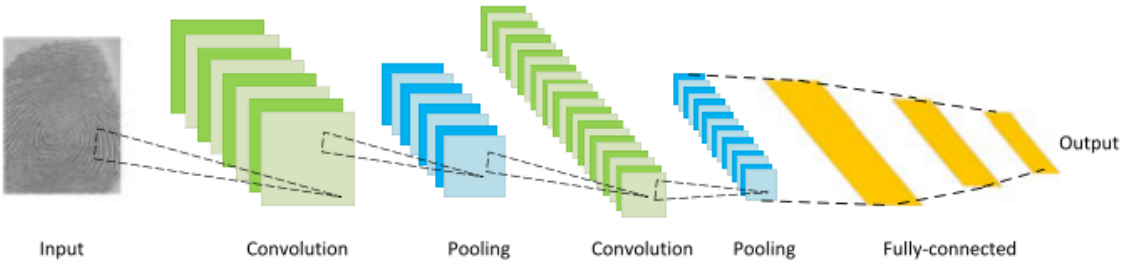


Figure 3.2: Architecture of a Convolutional Neural Network (Minaee et al., 2023).

Graph-based Semantic Segmentation

Second, there are graph-based segmentation models, in which graph theory is applied to create a representation of an image as a graph. In this approach, each pixel in the image is represented as a node and the edges connecting these nodes indicate the degree of similarity between the corresponding pixels, based on characteristics such as color or intensity (Le et al., 2008).
Graph-based models typically follow a series of standard steps for image segmentation. In the first step, a graph is constructed to represent the image, with each pixel serving as a node. The weights of the edges connecting the nodes are defined based on the similarity or dissimilarity between pixels. For example, pixels with similar colors will have a stronger connection, while dissimilar pixels will have a weaker connection. The graph is then partitioned into disjoint regions using a graph partitioning algorithm. This algorithm aims to minimize the weights of the edges that connect different segments, effectively separating the image into meaningful areas.
Finally, the segments are refined by merging or splitting them based on various criteria, such as size, shape, and texture. This step ensures that the resulting segments are coherent and relevant for analysis

(Camilus and Govindan, 2012).

Figure 3.3 illustrates the process of graph-based segmentation models in image analysis. First, there is the original image consisting of a grayscale representation of an original image. Each pixel in this image represents a data point that can be analyzed for segmentation.Second, there is the graph representation. Here, each pixel of the original image is represented as a node in a graph. The connections between these nodes illustrate the relationships between pixels based on their similarities, such as color, intensity, or texture. The graph is then partitioned into distinct segments or clusters. This is achieved by using a graph partitioning algorithm that aims to minimize the connection weights between different segments. This means that pixels within the same segment are more similar to each other than to pixels in other segments. Finally, the last image shows the result of the segmentation process. The segmented areas are displayed in different colors, indicating the different segments of the image that have been identified through the graph-based approach (Helmy, 2019).

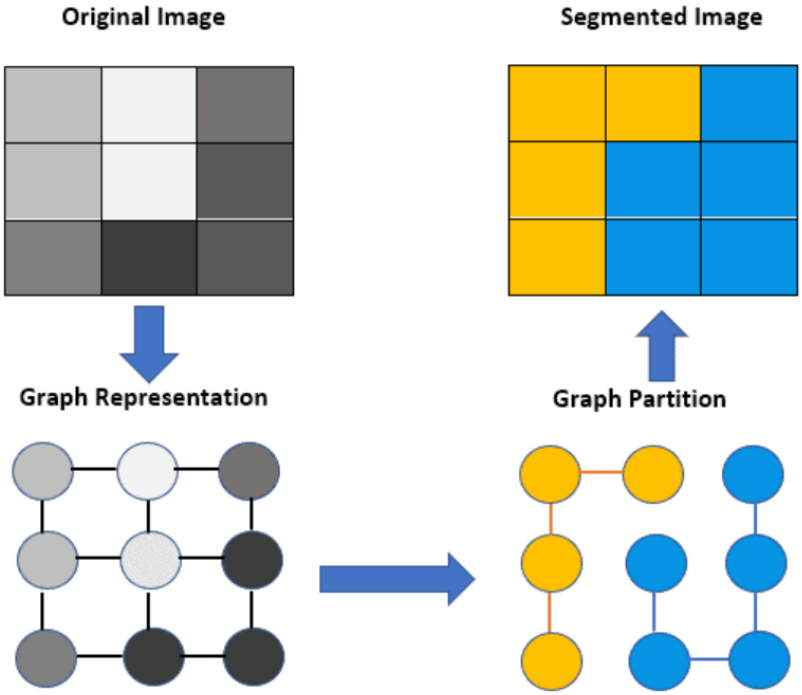


Figure 3.3: Association between image segmentation and graph partitioning (Helmy, 2019).

Transformer Based Semantic Segmentation

The last more modern models that are discussed are transformer-based models, which have gained significant traction in the field of semantic segmentation, offering a robust alternative to traditional CNNs. Transformer-based models are originally developed for natural language processing; these transformers excel in capturing long-range dependencies and contextual information, allowing them to understand complex patterns in data (Vaswani et al., 2017). The implementation of vision transformers (ViT) extends this capability to various CV tasks, including semantic segmentation (Li et al., 2024). Transformer models typically consist of two main components: the encoder and the decoder. The encoder operates on a sequence of input tokens, which, in the case of images, are small sections (or patches) of the image that have been flattened into a one-dimensional format. These tokens are supplemented with positional embeddings, vectors that convey information about the location of each patch within the image, ensuring that spatial relationships are preserved. Some transformer models include a decoder to reconstruct more detailed information about the image. The decoder processes the encoded features and refines them, often using upsampling, to generate detailed outputs such as segmentation maps or object predictions (Strudel et al., 2021).

Figure 3.4 shows ViT, which utilizes the encoder component of the transformer architecture. Instead of processing a sequence of words, as in traditional transformers, it takes this sequence of image patches (Vaswani et al., 2017). Within the transformer encoder block, layers of Multi-Head Self-Attention (MSA) and Multi-Layer Perceptron (MLP) modules are present, with layer normalization applied prior to both types of modules. MSA allows the model to consider all image patches simultaneously, capturing various features and relationships within the image. To mitigate issues such as dead spots in the gradient (areas where the model stops learning effectively due to the attention mechanism), residual connections are implemented within the transformer encoder. These connections help maintain the flow of gradients during training. The output generated by the Transformer encoder consists of a set of characteristics that can be further processed by an MLP for class predictions, as illustrated in Figure 3.4. ViT has shown superior performance compared to CNNs, particularly when trained on large datasets or when applied to low-resolution image datasets, highlighting its effectiveness and scalability in the handling of visual data (Dosovitskiy et al., 2020).

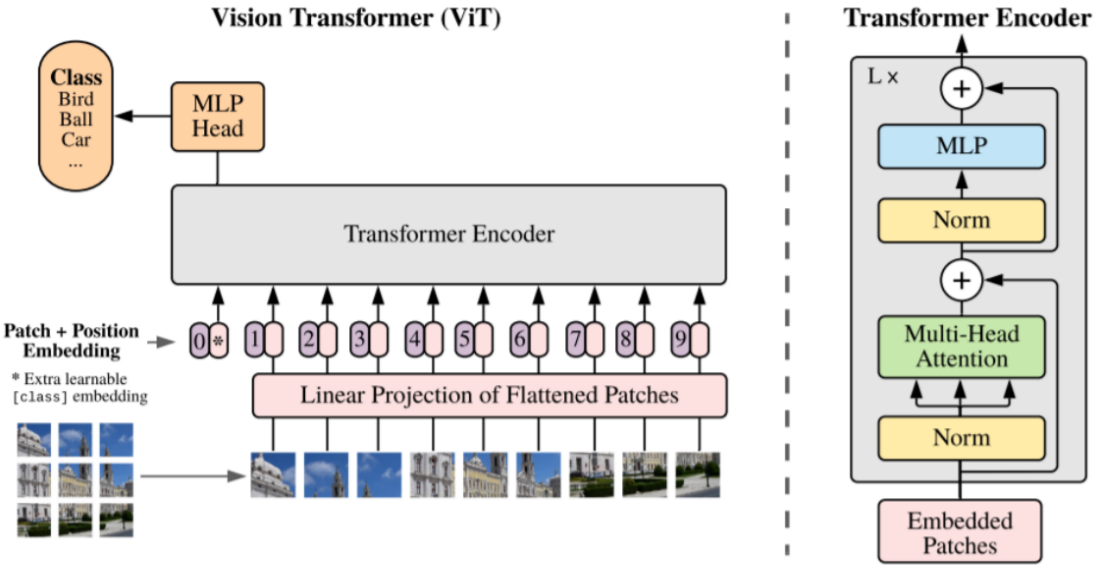


Figure 3.4: Model overview of ViT inspired by Vaswani et al. (2017) and Dosovitskiy (2020).

3.1.4 Point Cloud Segmentation

After reviewing the various types of segmentation methods and the concept of smart point clouds, it is essential to review different iconic or state-of-the-art models that have demonstrated promising results for indoor semantic segmentation on datasets like S3DIS and ScanNet, while also able to perform in (near) real-time. This review forms the basis for selecting an appropriate model for (near) real-time semantic segmentation of point clouds.

Pointnet

PointNet introduced by Qi et al. (2017a) was revolutionary in the field of 3D deep learning because it introduced a fundamentally different approach to processing point cloud data compared to existing models. This new approach led to an increase in success for tasks such as 3D object detection, segmentation, and other geometric tasks (Qi et al., 2017a). PointNet uses a novel neural network architecture that directly processes 3D point cloud data. The more traditional methods convert irregular point cloud data into regular structures such as 3D voxel grids or 2D image collections, which increases the data size and reduces performance due to artifacts and information loss (Maturana and Scherer, 2015). In contrast, PointNet directly analyzes raw point cloud data and respects its unordered nature, making it more efficient and versatile (Qi et al., 2017a).

The paper introduces a novel architecture that is suitable for consuming unordered point clouds in 3D. The network consists of three key modules: a max pooling operation, a local and global information structure, and a joint alignment network (T-Net).
First, using a max pooling strategy, Qi et al. (2017a) try to overcome the challenge that point clouds are unordered, which means that any change in the order of the points should not affect the outcome. The idea to overcome this challenge is simple: each point in the point cloud is passed through an MLP, which independently processes each point's features. After processing the points, the network aggregates the features across all the points using the max pooling operation. This is a symmetric function, meaning the order of the inputs does not matter, it picks the maximum or average value from each dimension of the features. This ensures that the network output is invariant in order of points (Qi et al., 2017a).
The features of each point are pooled into a global feature vector. This vector summarizes the entire point cloud and can be used for tasks such as object classification. However, for tasks like semantic segmentation, where each point in the point cloud needs its own label, it is necessary to combine both local features and global features. PointNet achieves this by concatenating the global feature vector back to the original per-point features after max pooling. In this way, the prediction of each point is influenced by both its local information (its original characteristics) and the global context (the overall shape of the point cloud) (Qi et al., 2017a).
The last module PointNet uses is a joint alignment network, which addresses the need for geometric invariance in point cloud data. Since point clouds can undergo various transformations (e.g., rotations, translations), the network needs to be able to handle these changes without affecting its predictions. The joint alignment network ensures that the semantic labels assigned to each point are consistent, regardless of the orientation or position of the point cloud (Qi et al., 2017a).

PointNet++

PointNet++ is an extension of PointNet, designed for deep hierarchical feature learning on point sets, in particular 3D point clouds, and introduced by Qi et al. (2017b). The primary objective behind PointNet++ is to address the limitations of the original PointNet, which was one of the first models that focused on deep learning on point clouds. The basic idea of PointNet is to capture a spatial representation for each point in a point cloud and then combine these individual feature points into a unified global representation of the point cloud (Qi et al., 2017a).
However, according to Qi et al. (2017b) the original PointNet lacks the ability to capture local structures induced by the spatial arrangement of points in a metric space. This makes PointNet unsuitable for recognizing fine-grained patterns or for handling complex scenes with varying point densities. Furthermore, PointNet uses a global max pooling operation to aggregate features from all points into a single global representation. Although this pooling operation ensures invariance in the order of points, it causes the network to miss local patterns that could be crucial to understanding the structure within a point cloud (Qi et al., 2017b).
PointNet++ tries to solve this problem by introducing hierarchical feature learning. First, PointNet++ uses hierarchical grouping of points by using three different key layers: sampling layer, grouping layer, and PointNet layer.
The sampling layer selects a subset of points from the input that serve as centroids for the local regions. The grouping layer then forms local regions by identifying 'neighboring' points around these centroids. The PointNet layer applies a mini-PointNet to encode the patterns within each local region into feature vectors. This allows PointNet++ to progressively abstract features from smaller to larger neighborhoods, which is crucial to deal with complex scenes (Qi et al., 2017b).
Furthermore, PointNet++ integrates multi-resolution grouping (MRG) which is an alternative approach to multiscale grouping (MSG). MSG tries to combine features from different regions on different scales, allowing the network to capture both fine details and a larger context based on the density of the input data. However, the MSG approach is computationally expensive because it applies local PointNet to large-scale neighborhoods for each centroid. This becomes especially time-consuming at the lowest level, where the number of centroid points is typically very high, leading to a substantial increase in processing time, which is sub-optimal when you try to segment in (near) real-time). MRG avoids such expensive computations by using two feature vector layers, one from the output of the set abstraction layer (which includes the three key layers) from the previous level (which captures local context at the lower resolution) and another vector from processing the raw points directly at the current level. Compared to using these two layer vector feature, this method is more computationally efficient, as it eliminates the need to extract features in large-scale neighborhoods at lower levels (Qi et al., 2017b).

In conclusion, PointNet++ shows promising results for tasks like indoor semantic segmentation. By addressing the limitations of the original PointNet with hierarchical feature learning and introducing the more computationally efficient method of MRG, PointNet++ is able to capture both fine-grained details and larger contextual information. However, there are some limitations when using PointNet++ for indoor semantic segmentation. For instance, the model performance can degrade when dealing with sparse point clouds or cluttered scenes, as the hierarchical grouping may not fully capture intricate relationships between points. Additionally, the computational efficiency of PointNet++ may still be too slow in very large-scale environments, which could affect the (near) real-time performance in certain scenarios.

Dynamic Graph CNN for Learning on Point Clouds

Modern deep neural networks are specifically designed to handle the irregular structure of point clouds by directly processing raw point cloud data, rather than converting it into a more structured or regular form (like a grid or a mesh) (Qi et al., 2017a, 2017b).
However, according to Wang et al. (2019), this has one drawback: it overlooks the geometric relationships between points, which poses a fundamental limitation in capturing local characteristics. To overcome this drawback, Wang et al. (2019) introduce EdgeConv, which captures local geometric structures while preserving permutation invariance. Rather than directly generating point features from their embeddings, EdgeConv produces edge features that represent the relationships between a point and its neighboring points, meaning the order in which neighboring points are processed does not affect the outcome.
In addition, unlike traditional graph-based approaches, DGCNN uses a dynamically updated graph. The neighborhood of each point, defined by its k-nearest neighbors, is recalculated after each layer of the network based on the evolving feature embeddings of the points. As a result, proximity in feature space may differ from proximity in the original input space, allowing for a non-local diffusion of information throughout the point cloud. This dynamic adjustment helps the model capture both local geometric structures and global relationships as it processes deeper layers of the network, ultimately improving its ability to learn from irregular point cloud data (Wang et a., 2019).
By introducing the EdgeConv operation, Wang et al. (2019) aim to capture local geometric structures while preserving permutation invariance. As a result, DGCNN is able to represent relationships between points more effectively. Its dynamic graph updating mechanism allows for a non-local diffusion of information, enabling the model to capture both local and global features as it processes deeper layers of the network.
These innovations enable DGCNN to outperform traditional point cloud processing models, making it suitable for tasks like 3D object classification and semantic segmentation. However, despite these strengths, DGCNN still faces challenges when it comes to computational complexity, particularly when it comes to handling large-scale datasets or (near) real-time applications, due to the increased processing time caused by dynamic graph updates.

Point-Voxel CNN

To process LiDAR 3D point cloud data in (near) real time, it is essential to use an efficient and fast 3D deep learning model. According to Liu et al. (2019), several challenges motivated their novel approach with Point-Voxel CNN (PVCNN). First, voxel-based models consume a significant amount of memory. During the voxelization process, multiple points are merged, which can cause information loss. Conversely, when working with high-resolution 3D point cloud data, the computational cost and memory requirements of this process increase drastically. Second, while point-based models are more memory-efficient compared to voxel-based methods, they suffer from poor memory locality and irregular memory access, which makes these models inefficient for feature extraction. These challenges collectively contribute to the inefficiencies of current voxel- and point-based 3D models. Liu et al. (2019) propose PVCNN as a solution, combining the memory efficiency of point-based methods with the computational locality of voxel-based methods to overcome these challenges.
The voxel-based feature aggregation begins with normalization of the 3D coordinates of the points to ensure consistency. The points are translated into the local coordinate system and scaled to fit within a unit sphere. Thereafter, the voxelization process begins, where the normalized point cloud is transformed into voxel grids. This is achieved by averaging the points that fall within the same voxel grid to create a single feature vector for each grid cell. The voxelization helps capture neighborhood information at a coarse level, reducing the memory footprint by using a lower resolution for the voxel grid. After vox-

elization, standard 3D volumetric convolutions are applied to the voxel grids to aggregate features from neighboring grids.
Once the voxel-based convolutions are completed, the features need to be mapped back to the original point cloud representation. PVCNN uses trilinear interpolation for this step, ensuring that the features mapped to each point remain distinct, preserving fine-grained information for further processing. In addition to voxel-based feature aggregation, PVCNN uses point-based feature transformations, as relying solely on low-resolution voxel-based features may not be sufficient for semantic segmentation. This point-based feature transformation is achieved using an MLP, which transforms the features of each point based on its own attributes without aggregating neighborhood information. This step generates fine-grained, high-resolution details specific to each point.
The final step involves fusing the features from the voxel-based and point-based branches to create a more comprehensive representation of the 3D data. The combination of these two branches allows PVCNN to achieve both high accuracy and efficiency (Liu et al., 2019).
By combining features from both the voxel-based and point-based branches, PVCNN creates a comprehensive representation of the 3D data. Furthermore, this combination enables PVCNN to achieve both high accuracy and efficiency. However, despite its advantages, PVCNN is not without limitations. The compromise between voxel and point-based methods may still result in the loss of finer details, and the complexity of combining two different feature extraction approaches introduces additional computational overhead. Finally, PVCNN may not always outperform task-specific models designed for particular applications, especially in scenarios that demand highly specialized representations.

PointNeXt

In recent years, many new and sophisticated models have outperformed older models such as PointNet and PointNet++. However, Qian et al. (2022) state that this improvement is mainly due to better training strategies and larger model sizes rather than groundbreaking architectural innovations. They researched this claim and tested improved training strategies on the existing PointNet++ model. Its performance increased significantly from 77.9% to 86.1% without any architectural changes.
To further modernize the classical architecture of PointNet++, Qian et al. (2022) introduced a new architecture called PointNeXt. This architecture incorporates features such as the inverted residual bottleneck design and separable MLPs, enabling the model to scale more effectively while maintaining high computational efficiency. PointNeXt builds on the existing architecture of PointNet and PointNet++ but incorporates improved training strategies that involve data augmentation, optimization techniques, and model scaling. This is achieved by implementing two key aspects: receptive field scaling and model scaling.
First, the authors revisited the strategy by experimenting with different radius values, finding that a smaller radius improved the overall performance of the model. Additionally, the model was enhanced by introducing relative position normalization. In PointNet++, the network learns the relative positions of neighboring points, but optimization becomes more challenging because the relative positions are small, leading to weak gradients. By normalizing the relative positions using the radius of the neighborhood, the network learns more efficiently, improving the overall performance of the model (Qian et al., 2022).
The second key aspect introduced in PointNeXt is model scaling. This involves modifying the architecture to make it deeper or wider, increasing the model's capacity to learn more complex features. The most notable modification is the introduction of inverted residual MLP blocks. These blocks allow the model to scale more efficiently through three key concepts: residual connections, separable MLPs, and an inverted bottleneck design. The residual connections alleviate the problem of vanishing gradients, especially in deeper networks. The separable MLPs reduce computational costs while enhancing the network's ability to extract point-wise features by separating MLPs into layers that process neighborhood features and point features independently. The inverted bottleneck design expands the number of channels before reducing them again, enriching the extractions of features without incurring heavy computational burdens (Qian et al., 2022).
In addition to the introduction of Inverted Residual MLP blocks, three key modifications were made to the macro architecture. First, the design of the PointNet++ encoder was standardized for both classification and segmentation by increasing the number of Set Abstraction blocks for classification from 2 to 4, while maintaining 4 blocks for segmentation at each stage. Second, a symmetric decoder was implemented, adjusting its channel size to match that of the encoder. Lastly, a stem MLP (an additional MLP layer) was added at the start of the architecture to project the data from the input point cloud data in a higher-dimensional space (Qian et al., 2022).

In conclusion, PointNeXt successfully revisits and modernizes the classical PointNet and PointNet++ architectures by integrating improved training strategies and efficient scaling mechanisms. By focusing on receptive field scaling, model scaling, and the introduction of inverted residual MLP blocks, PointNeXt demonstrates significant performance gains on multiple benchmarks without requiring complex architectural innovations.

OneFormer3D

Kolodiazhnyi et al. (2024) propose a new segmentation model called OneFormer3D, a unified approach for 3D point cloud segmentation that addresses semantic, instance, and panoptic segmentation tasks. Traditional approaches typically use separate models for each task, but OneFormer3D employs a multitask framework that addresses all these tasks simultaneously using a transformer-based decoder.
OneFormer3D begins by extracting features from 3D point clouds. Each input point in the point cloud is represented by its 3D coordinates and color (RGB). These point clouds are voxelized and processed using a U-Net structure composed of sparse 3D convolutions, resulting in point-wise features. Following this, flexible pooling is implemented, allowing the model to pool superpoints or voxel features. In the superpoint scenario, the features of each superpoint are averaged, reducing the computational burden while preserving the local structure. This flexible pooling strategy reduces millions of points into fewer superpoints or voxels, making subsequent processing more efficient.
The next step involves using the transformer-based decoder, which is central to the model's ability to perform multi-task learning. It accepts instance and semantic queries as input and, through cross-attention mechanisms, generates a set of masks corresponding to objects or regions within the 3D point cloud. The use of a transformer enables the model to capture global relationships between points in the cloud. Furthermore, this framework utilizes learnable kernels that generate masks for semantic and instance categories using a single model. These kernels are processed through a shared transformer decoder, allowing joint training for all segmentation tasks (Kolodiazhnyi et al., 2024).

Superpoint Transformer

Robert et al. (2023) introduce a new transformer-based model designed to process 3D point clouds by leveraging both local and global features in an efficient manner. The superpoint Transformer consists of two key components.
First, an efficient hierarchical Superpoint Partition is used. This technique divides the point cloud into a hierarchical structure of superpoints, where each superpoint represents a group of nearby points with similar geometric properties. By clustering points into superpoints, the model reduces the complexity of the input data, allowing it to capture local features more effectively while significantly reducing computational burden (Robert et al., 2023).
Thereafter, the Superpoint Transformer introduces a superpoint transformer-based attention mechanism, which looks similar to the popular U-net. However, instead of using grids, points, or graph sub-sampling, the approach uses different partition levels. The idea of this attention mechanism is to allow the model to dynamically learn the importance of each point relative to others in a local neighborhood. This is essential for capturing fine-grained spatial relationships in 3D space. One key component of this attention mechanism is encoding the adjacency between points which are based on spatial proximity. This component ensures that the relative position of the points is taken into account when learning the relationships between the points (Robert et al., 2023).
All in all, the implementation of an efficient hierarchical Super Point partition and the self-attention mechanism leads to a state-of-the-art performance on three benchmark datasets (S3DIS, KITTI-360 and DALES) that are often used for semantic segmentation. It's ability to capture both local and global features makes it highly effective for semantic segmentation, this in combination with its scalable design ensures it can handle large and complex data efficiently.

Point Transformer V3

The development of Point Transformer V3 (PTv3) is motivated by the recognition that previous 3D models, particularly those based on transformer-based architectures, have struggled with efficiency due to limitations in scaling (Wu et al., 2024). Unlike the rapid advancements seen in 2D vision and natural language processing, 3D point cloud processing has lagged, primarily because of the limited size and diversity of point cloud datasets, making it difficult to apply scaling principles effectively.

PTv3 is designed to overcome the trade-offs between accuracy and efficiency in previous models by scaling up while maintaining simplicity. This is done by implementing the following improvements. First, they introduce point cloud serialization, which transforms unstructured point cloud data into a structured format by organizing the points based on their spatial relationship. This structuring is done by using space-filling curves, which traverse every point within a 3D space while maintaining some degree of spatial proximity. Two key space-filling curves are used for this, the Z-order curve, which is valued for its simplicity and ease of computation, and the Hilbert curve, which is recognized for its superior locality-preserving capabilities compared to the Z-order curve. By adjusting the order of the traversal, Wu et al. (2024) introduce Trans Z-order and Trans Hilbert. These variants can provide alternative insights into relationships, potentially capturing unique local patterns that the original patterns may miss.
Secondly, they introduce serialized attention, which leverages the structured nature of serialized point clouds to improve the efficiency of attention mechanisms, which are often used in transformer architectures. Instead of using complex neighborhood construction strategies like K-Nearest Neighbor (KNN), the authors introduce patch attention, where points are grouped into patches along the serialized order, and attention is performed within these patches. Using patch grouping, the computational complexity compared to traditional attention mechanisms decreases drastically. Finally, Wu et al. (2024) focused on optimizing the architecture for scalability and efficiency by utilizing techniques like pre-norm blocks, conditional positional encoding, and grid pooling.
By introducing point cloud serialization, serialized attention and focusing on the architecture, PTv3 shows significant advancements in the field of 3D point cloud processing. By its improved overall efficiency, scalability, and memory efficiency, PTv3 appears well positioned for real-time indoor segmentation. However, further real-world testing in complex environments and with varying point cloud densities is necessary to fully determine its effectiveness for (near) real-time semantic segmentation.

Point-SAM

Point-SAM introduced by Zhou et al. (2024), is a transformer-based segmentation model specifically designed for 3D point clouds. It extends the functionality of the 2D Segment Anything Model (SAM) to the 3D domain. Point-SAM addresses several challenges inherent in 3D data, such as irregularity, scalability, and the need for diverse annotations. The model uses an innovative Voronoi-based tokenizer to efficiently convert raw point clouds into embeddings, preserving local structures while reducing computational complexity. This tokenizer divides the point cloud into patches based on the Voronoi diagrams, which not only balances efficiency with effectiveness but also enhances the model's scalability. The computational complexity includes reduced memory usage and improved processing speed compared to the more traditional methods, such as K-nearest neighbors (KNN) tokenization.
The model builds on the foundation established by SAM as developed by Kirillov et al. (2023), which introduced a promptable segmentation task and a corresponding model architecture capable of zero-shot generalization to new tasks. The architecture of Point-SAM consists of three key components: a point-cloud encoder, a prompt encoder, and a mask decoder. The point-cloud encoder utilizes a transformer-based approach, with the input point cloud first tokenized into different patches. These patches are embedded and processed through ViT to generate point-cloud embeddings. The prompt encoder processes point- and mask-based prompts, which allow the model to perform promptable segmentation tasks, such as interactive annotation or zero-shot segmentation.
A critical feature of Point-SAM is its use of self-attention and cross-attention mechanisms within the mask decoder, which are applied in both prompt-to-point-cloud and point-cloud-to-prompt directions. These attention mechanisms allow the model to effectively integrate prompt information, iteratively refining segmentation masks. The mask decoder generates precise segmentation masks by upsampling the point-cloud embedding, and a dynamic linear classifier is used to predict foreground probabilities for each point.
Point-SAM is trained on a mixture of different datasets, including PartNet, ScanNet, and ShapeNet, which provides both part-level and object-level annotations. To enhance label diversity and overcome the scarcity of annotated 3D data, Point-SAM uses a data engine that generates pseudo-labels through knowledge distillation from the 2D SAM model. This process helps to expand the diversity of masks, improving the model's zero-shot transferability to new tasks and datasets. The model achieves state-of-the-art performance in tasks such as zero-shot segmentation, interactive annotation, and instance proposal generation on benchmark datasets like S3DIS and KITTI360 (Zhou et al., 2024).
By combining the strengths of SAM's promptable architecture with specialized adaptations for 3D data, Zhou et al. (2024) provide a robust solution for 3D segmentation tasks. Its ability to handle diverse

point-cloud data, coupled with its efficient tokenization and segmentation capabilities, makes it a foundational model for real-time 3D segmentation and large-scale applications. Point-SAM can be used to support a variety of applications, including those that require real-time processing and large-scale data handling, offering significant improvements in efficiency and scalability compared to existing methods.

Summary

This review covered nine advanced models designed to address the challenges of indoor segmentation in 3D point clouds. Each model presents a unique approach to increase accuracy and efficiency, with innovative methods for processing and interpreting point cloud data. All of these models demonstrate significant advances in addressing issues such as computational efficiency, hierarchical feature extraction, local and global context integration, and real-time processing. In the results part of this research, a comparison of these models is conducted, evaluating their performance on benchmark datasets and their suitability for (near) real-time semantic segmentation. Based on this comparison, one model will be selected and integrated into the system to achieve (near) real-time segmentation.

Table 3.1: Overview of the different segmentation models

Model	Year	Type	License	Library	Pros	Cons
PointNet	2017	CNN-based	MIT	PyTorch	Efficient, respects unordered input	Misses local context, poor detail capture
PointNet++	2017	CNN-hierarchical	MIT	PyTorch	Captures local/global features	Degrades on sparse/cluttered data
DGCNN	2019	Graph-based	MIT	PyTorch	Captures local geometry via EdgeConv	Computationally heavy due to dynamic graphs
Point-Voxel CNN	2020	Hybrid (point + voxel)	MIT	PyTorch	Combines fine and coarse context	Overhead from dual-branch architecture
PointNeXt	2022	CNN-based	MIT	PyTorch	Efficient, scalable architecture	Gains rely on scaling, not novelty
OneFormer3D	2023	Transformer-based	CC BY 4.0	PyTorch	Multi-task, efficient superpoint pooling	Complex, preprocessing-heavy
Superpoint Transformer	2023	Transformer-based	MIT	PyTorch	Reduces input, captures local/global features	Loses fine details, transformer cost
Point Transformer V3	2023	Transformer-based	MIT	PyTorch	Efficient patch-based transformer	Real-world performance not yet proven
Point-SAM	2024	Transformer-based	MIT	PyTorch	Promptable, efficient tokenization	Heavy architecture, needs large training data

3.2 Decision Making in Emergency Situations

The second section of the related work explores the critical role of decision making in emergency situations, with a focus on improving SA to improve response effectiveness. First, it examines the concept of SA, its hierarchical levels as defined by Endsley, and its application in dynamic, high-pressure environments such as those in emergency response. This section also connects SA to (near) real-time segmentation, emphasizing how quickly and accurately processed data contribute to better decision making. The integration of SLAM technology as a tool for improving SA is also discussed, highlighting its importance in providing (near) real-time spatial data for decision support.

3.2.1 Situational Awareness

The concept of SA is a critical component in decision making, particularly in dynamic and high-pressure environments such as emergency response. It refers to the ability of individuals and teams to perceive, comprehend, and anticipate environmental changes. Thus, decision-makers must quickly assess and understand complex, dynamic, and uncertain environments (Kapucu and Garayev, 2011). Endsley (2015) defines situation awareness as "the perception of the elements of the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future". Therefore, SA plays a crucial role in the way decision makers perceive, comprehend, and project information in a complex and constantly changing environment.

Levels of Situational Awareness

Endsley's framework (2015) introduces three hierarchical levels of SA, where each stage is necessary but not sufficient on its own to reach the next. It explains the cognitive processes and mechanisms people use to assess situations and develop SA, along with the task and environmental factors that influence their ability to achieve it. The model follows a sequence of information processing: starting with perception, moving to interpretation and then finally to prediction (Stanton et al., 2001).

Level 1: Perception

Perception is the most basic level of SA. At this level, the individual perceives raw information from their environment. Hereby, no interpretation or integration of the data occurs; it is merely the initial receipt of sensory inputs. For example, in aviation, a pilot might perceive information from aircraft instruments, the behavior of the aircraft, or external cues such as other planes, terrain, or air traffic control, but does not yet process or understand the significance of these data (Stanton et al., 2001).

The perception of relevant information can come from various senses (visual, auditory, or tactile). For example, a flight controller relies on visual displays showing aircraft near an airfield, while a surgeon may depend on tactile feedback during an operation. Verbal and non-verbal communication also play a role in the collection of data (Smit, 2020).

However, the reliability of these information sources is critical at this stage. The operator's trust in the data, shaped by their experience with the sources and the quality of the data, impacts their confidence in using it. Incomplete or unreliable data can hinder the development of SA. As a result, even reaching the first level of SA can be challenging in environments where collecting accurate data is difficult (Endsley, 2015). Achieving the first level of SA involves perceiving the status, attributes, and dynamics of the relevant elements, which lay the foundation for the higher levels of SA.

Level 2: Comprehension

At the second level of SA, raw data from the first level is transformed into meaningful information. This level can only be reached if the elements of perception are available and properly perceived. The key process at the second level of SA is integrating and synthesizing the separate pieces of data to form new insights (Smit, 2020). For example, within the context of this research, combining segmented point clouds with visualization and mapping can lead to a more coherent understanding of the situation.

According to Endsley (2015), the experience an operator has is crucial at this stage, as those with well-developed mental models are better able to synthesize information, integrating new data into their understanding of the situation.

The second level of SA goes beyond simple data perception; it involves understanding the significance of the perceived elements in relation to the operator's tasks and objectives. For example, a pilot integrates information about fuel levels, time, distance, and tactical threats to understand how these factors affect their mission. The ability to comprehend these connections allows the operator to judge whether their

actions are producing the intended outcomes. More experienced individuals are often able to achieve higher levels of comprehension than their less skilled counterparts, even if they both have access to the same level 1 data (Stanton et al., 2001). The second level of SA involves the synthesizing and interpreting of various pieces of information, providing a deeper understanding of the situation and its implications.

Level 3: Projection

The third and highest level of SA involves the ability to project the future status of elements in the environment based on their current state and dynamics, which allows for timely and effective decision making (Endsley, 2015). Achieving the third level of SA builds on the comprehensive understanding developed in the previous levels, requiring not only perception and comprehension, but also the ability to anticipate how the situation will evolve.

However, this projection highly depends on a well-developed mental model and the accurate integration of temporal data, allowing the operator to predict future events and potential outcomes. For example, a pilot might predict potential aircraft conflicts based on current flight paths, giving him time to adjust and avoid issues. The ability to foresee future states is critical for timely and effective decision making in dynamic environments, as it allows operators to plan actions ahead of time to meet their goals (Stanton et al., 2001).

The accuracy of this projection depends on the reliability of the information gathered at the earlier levels and requires significant cognitive resources to achieve.

Situational Awareness Model

The SA model introduced by Endsley (2015), presented in Figure 3.5, illustrates how the three levels of SA work together to influence decision making and action performance. The model shows that Level 1: Perception forms the foundation, where raw data from the environment are perceived. This data is then processed and integrated into meaningful information at Level 2: Comprehension, which allows the operator to understand the current situation in relation to their goals. Finally, Level 3: Projection, enables the operator to anticipate future states based on the understanding of the current situation.

Furthermore, the figure also shows the feedback loop between the state of the environment and SA, emphasizing how continuous information flow and feedback are critical for updating SA. Task/system factors, such as workload, stress, system design, and complexity, influence the development and maintenance of SA, while individual factors, such as goals, experience, training, and preconceptions, further affect the operator's ability to build and maintain SA.

Ultimately, the decision-making process is based on accurate and timely SA, leading to the performance of actions that can be monitored and adjusted based on further changes in the environment. The ongoing dynamic process of gathering, interpreting, and projecting information is crucial to ensuring effective decision-making and action in complex, dynamic environments.

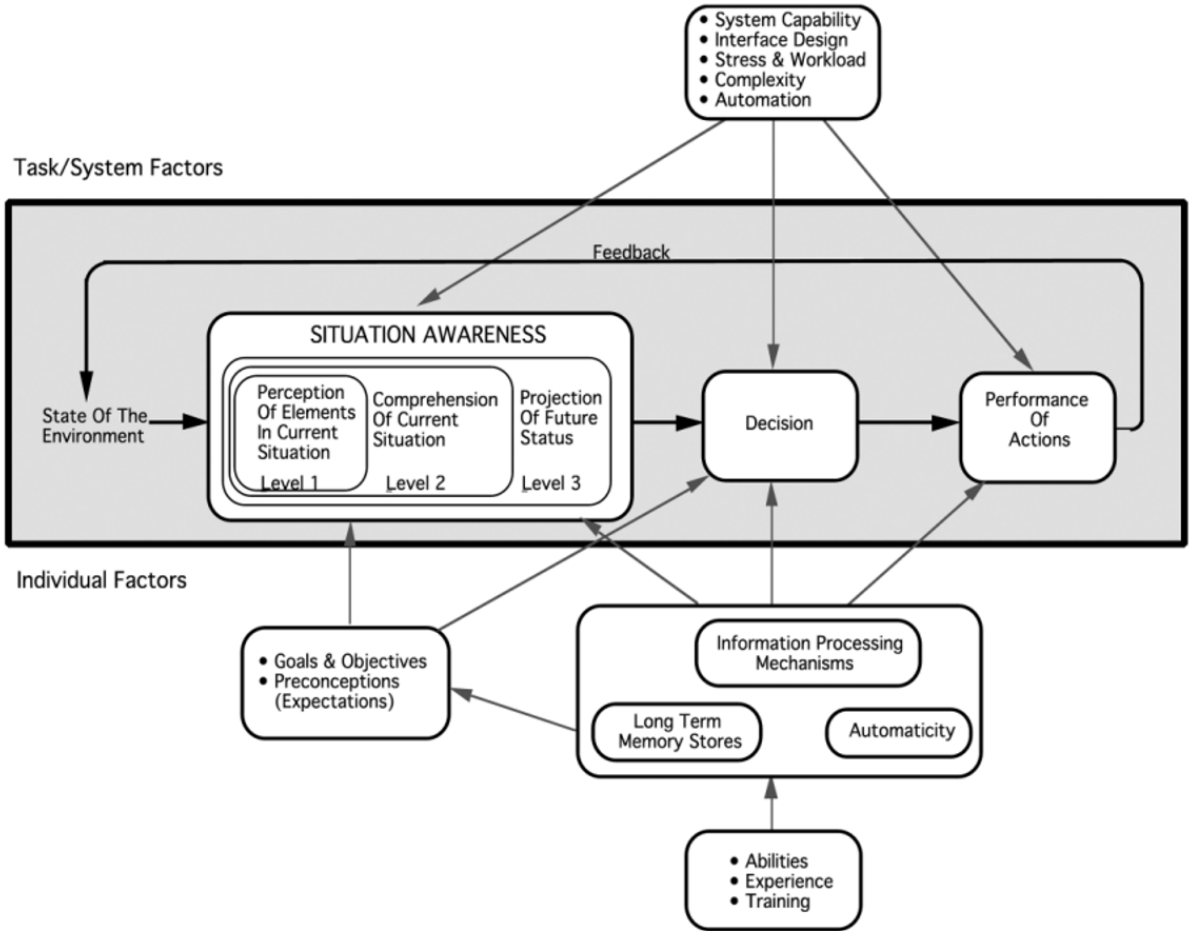


Figure 3.5: Model of SA in dynamic decision making (Endsley, 2015).

Situation Awareness in (Near) Real-Time Segmentation

After introducing the different levels of SA, the connection to (near) real-time segmentation must be made. This is because (near) real-time segmentation involves rapidly processing and analyzing incoming data to make quick decisions. In this context, the ability of users and systems to perceive, comprehend, and project changes in the environment is essential.
At the most basic level of SA, the focus is on identifying key environmental elements of raw sensory input (Endsley, 2015). For (near) real-time segmentation, this corresponds to the system's ability to perceive and recognize different features or objects within the 3D point cloud. This (near) real-time data from the LiDAR scanner is collected and processed to capture the current state of the environment. It is important to achieve accurate and timely data at this stage because it sets the foundation for higher levels of SA.
After achieving the first level of SA, the next step is to comprehend the segmented data. At this level of SA, the raw data is processed and integrated into meaningful information. In the context of this research, comprehension involves interpreting segmented 3D point clouds to identify, for example, objects, surfaces, or other points of interest.
The final and highest level of SA, projection, involves anticipating future states of the environment based on the current segmentation data. In (near) real-time segmentation, this refers to the ability of both the system and the users to predict how the environment might evolve and what actions may be required. For example, in emergency response scenarios, the ability to predict potential hazards or changes in the environment based on segmented data could enable proactive decision making and improve outcomes.

Testing for Situation Awareness

This research explores the extent to which a (near) real-time 3D point cloud segmentation model can help remote observers make informed spatial decisions. A key aspect involves evaluating how the segmen-

tation data and its visualization contribute to the user's ability to perceive, comprehend, and anticipate the environment, which are the key elements of SA. To investigate the effectiveness of visualization, it is important to assess how well SA is supported by the visualization of the segmented point cloud.
However, measuring SA is not an easy task. Many of the critiques of the 1995 model by Endsley (1995) focus on the difficulties related to accurately measuring SA. In response to these challenges, Endsley conducted a systematic review that describes several methods for evaluating the extent to which SA is fostered by various systems (Endsley, 2021). The review by Endsley (2021) can be subdivided into two parts: process testing and direct testing.

Process Testing
The first method to measure SA is through process testing. This kind of test involves observing how individuals or teams interact with stimuli and measuring their reactions to assess how SA develops in real-time. This method, often referred to as process indices, focuses on analyzing participants' behavior and cognitive processes during task performance. Various tools are used, including eye tracking, verbal protocol analysis, physiological measurements, and communication tracking. Each method provides unique insights into how individuals allocate attention and process information relevant to a specific task. These methods contribute to understanding the underlying cognitive mechanisms of SA by correlating certain behaviors. However, as noted by Endsley (2021), interpreting these behaviors as direct measures of SA is challenging. Individual differences, previous experiences, and stress levels in high-stakes environments can negatively affect reactions, making it difficult to generalize the results of one participant to another.

According to Salmon et al. (2009), in team settings, this process becomes even more complex. Individual SA may not accurately reflect team SA, which encompasses shared understanding and coordination between team members. Observers often find it challenging to assess whether SA is effectively distributed among team members based on individual behaviors alone. Commonly used tools, such as the Situation Awareness Global Assessment Technique (SAGAT) and the Situation Awareness Rating Technique (SART), struggle to address these nuances in team dynamics (Endsley, 2021).

In conclusion, process testing provides valuable but limited insight. Although it reveals attention allocation and behavioral responses, it lacks the comprehensive scope needed to accurately measure SA in all its dimensions. As Endsley (2021) suggests, isolated process measurements do not provide a complete picture of SA, underscoring the importance of combining multiple measures and contextual awareness to develop a holistic understanding of SA in complex environments. Furthermore, in 1995, Endsley already recognized the limitations of using process testing to determine performance as the only indicator of SA effectiveness. She, therefore, argued for a more direct measure of SA to test the effectiveness. (Endsley, 1995).

Direct Testing
In contrast to process testing, direct testing aims to assess SA as a state of knowledge. This method involves measuring the level of SA directly, either through subjective self-assessments or objective questions about the environment (Endsley, 2021). Endsley describes direct testing as one of the most effective ways to capture a snapshot of an individual's SA, given that it focuses on the information they actually possess about their surroundings rather than the processes they used to acquire it.

Subjective Measurements
Subjective measurements of SA offer a straightforward way to assess an individual's perceived awareness by simply asking participants about their mental model of the situation. However, these assessments also come with limitations. Participants may be unaware of what they do not know, resulting in potential gaps between perceived and actual awareness (Hamilton et al., 2017). Studies have shown that subjective SA ratings often reflect the confidence levels of the participants rather than their true situational understanding. Although observer ratings of SA attempt to bypass self-assessment biases, they still face challenges, as observers lack access to the participant's mental representation of the situation and may rely on perceived performance rather than actual SA (Endsley, 2021).

Objective Measurements
Objective measurements of SA assess an operator's knowledge by asking questions about the environment which are scored as correct or incorrect. This approach eliminates biases associated with subjective

measures by focusing on factual knowledge rather than perceived awareness (Endsley, 2021). Objective measurements categorize SA into three different levels (perception, comprehension, and projection). These levels allow researchers to assess how well participants perceive relevant elements, understand their meaning, and predict their future. Using these types of measurement, it provides a structured way of comparing the SA of different participants, giving a more direct measure of the SA without relying on subjective judgments. An example of these questions can be seen in Figure 3.6.

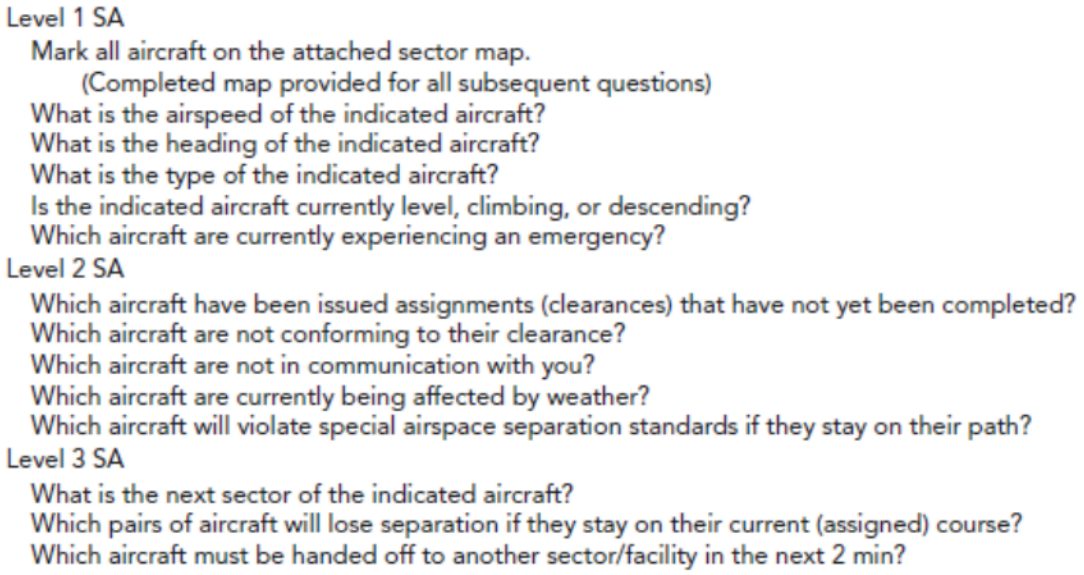


Figure 3.6: Example of questions used in a direct objective testing method (Endsley, 2021).

SLAM

To increase SA for first responders, a constant flow of information is needed, this information can then be utilized to increase the responders decision making process and thereby the first responders performance. For this, Smit 2020 utilized Simultaneous Localization and Mapping (SLAM), this technology is crucial in the world of indoor navigation and environment mapping. It enables systems for a wide range of mobile devices, such as robots, augmented reality (AR) devices, or even iPhones equipped with LiDAR scanners, to construct a map of an unknown environment while simultaneously keeping track of their own location within the map. SLAM is therefore crucial for operations in uncertain environments, where more traditional positioning systems, such as Global Navigation Satellite Systems (GNSS), fail due to signal interference from structures such as walls and ceilings (Rantakokko et al., 2011). This limitation is especially evident in indoor spaces, where the need for accurate (near) real-time spatial data is critical for decision making.

In the context of first responder operations, SA plays a crucial role in ensuring timely and effective responses to emergencies. The more accurate and comprehensive the SA, the better the quality of the decisions made by the responders, ultimately leading to improved performance during emergencies (Endsley, 2015). However, indoor environments present significant challenges to first responders due to the lack of reliable and up-to-date spatial data, such as floor plans or real-time tracking of personnel (Tashakkori et al., 2015). This is where SLAM technology proves to be valuable.
Using SLAM in indoor operations, first responders can generate (near) real-time dynamic maps of the environment, which can then be transmitted to remote command centers or displayed to first responders on-site. These maps not only provide a clear view of the current environment, but could also track the movements of first responders within, offering critical insights into blocked pathways, open spaces, and the positions of the first responders. These data contribute to building a Common Operational Picture (COP), a centralized (near) real-time visualization that improves perception and comprehension of the operational environment (Li et al., 2014).

SLAM improves SA by providing the first level of SA (perception) through the collection of accurate and (near) real-time data (Endsley, 2015). SLAM then enables the second level (comprehension) by combining these raw data to create meaningful visualizations that allow first responders to understand

the implications of the current situation. These visualizations support dynamic decision making, allowing for a better assessment of risk, the identification of hazards, and the allocation of resources in (near) real-time. In addition, continuously updating the map as the environment changes allows for the projection of future conditions, which is the third level of SA. This capability helps to anticipate future risks and plan interventions before they become critical (Endsley, 2015).

In conclusion, SLAM not only addresses the technical challenges of indoor navigation but also directly improves the SA of first responders by providing a more accurate, comprehensive, and dynamic understanding of the environment. Through its ability to generate real-time maps and track movement, SLAM enables better informed decision-making, ultimately improving operational performance during emergency situations.

3.3 Visualization in Decision Making

In emergency response and other high-pressure scenarios, effective visualization is critical to improve SA and support timely and informed decision making. By transforming complex data such as point clouds into accessible and actionable insights, visualization tools enable users to better understand spatial relationships, identify patterns, and make crucial decisions in dynamic environments. The last part of the related work section explores the principles and advancements in cartography and geovisualization, focusing on their application to indoor environments. Examines the evolution of mapping techniques, the theoretical foundations of effective visual design, and the unique challenges of indoor cartography. Through the combination of existing methods such as the New Map Communication Model, the Map Use Cube, and considerations of cognitive load, this section highlights how visualization can be optimized to meet the diverse needs of users in critical situations.

3.3.1 Introduction to Visualization in Decision Making

In complex and high-risk environments, such as emergency response, effective and informed decision making is crucial to success. However, this relies heavily on the ability to interpret large amounts of information quickly. Therefore, effective visualization of these data is essential to transform raw data into actionable insights by presenting them in formats that improve human cognition and SA (Smit, 2020). Kraak and Ormeling (2020) emphasize that maps are not just tools for navigation, but also serve as powerful forms of communication. This highlights the importance of not only visualizing the data, but also interpreting and conveying it effectively in decision making processes. Good visualization and the ability to interpret these data are particularly critical for first responders, especially during situations such as building fires and rescue operations. Timely decisions can mean the difference between success and failure, making the role of clear visualization even more vital (van der Meer, 2018).

The theoretical basis of visualization in decision making is based on its ability to support SA. This concept is vital for first responders navigating uncertain and dangerous environments. Visual tools can provide them with immediate and comprehensible information on building layouts, fire spread, and even location of personnel in the building, all of which are essential to formulate and execute effective tactics (Smit, 2020). An important aspect of geovisualization in emergency response decision making is the integration of 2D and 3D mapping techniques. In his master thesis, van der Meer (2018) explored the use of ToggleMaps, which allow users to switch between 2D and 3D visualizations of indoor environments, offering different perspectives on spatial information. This flexibility is important because different users, such as commanders, operators, and operational firefighters, have different information needs. Although commanders might prefer a high-level overview of the situation using 2D maps, first responders located near the emergency might benefit more from detailed 3D representations of their immediate surroundings.

Bavle et al. (2023) further demonstrate the importance of real-time mapping technologies by introducing SLAM to provide up-to-date indoor spatial information. SLAM technologies allow responders to build real-time maps of complex environments, helping both on-site responders and remote command centers maintain SA. By dynamically tracking personnel and environmental changes, SLAM-based visualization systems provide critical data for ongoing operations, enabling responders to adapt their strategies as conditions change.

Finally, the theoretical concept of cognitive load is also relevant when designing visual tools for decision-making in emergency situations. Visualizations that are too detailed or complex can overwhelm users and hinder rather than support decision making. Therefore, the design of visualization tools must strike a balance between providing enough information to support decisions and avoiding information overload. This requires consideration of cartographic principles such as symbolization, color use, and map simplification (Fang et al., 2020).

3.3.2 Theoretical Foundations of Data Visualization

To increase SA for first responders, it is crucial to explore the fundamental principles of effective data visualization. By understanding these principles and integrating different models and techniques, it is possible to create more intuitive and informative visual representation. This can significantly improve decision making in dynamic, high-pressure environments, ensuring that first responders have the neces-

sary context and information to act data-driven and effectively.

The ability to create and use geographic visualizations, especially in the form of cartographic maps, is one of the most fundamental methods of human communication that dates back to the early forms of language (MacEachren, 1995). Geographic visualizations have played a significant role throughout history, particularly in earth sciences and navigation, long before the use of computer-based visualizations. Some of the first examples of geographic visualizations are map-like wall paintings from the Stone Age, showing the surroundings of early humans (Nöllenburg, 2007). Since then, cartography has continually evolved. Modern, computer-based geographic visualizations build upon this long tradition of cartographic knowledge. Traditional examples of static visualizations include thematic maps, which illustrate spatial patterns of topics such as climate or population density. In recent decades, the use of modern visualization technologies has opened up new possibilities to explore, understand, and communicate spatial phenomena. Maps, as a key tool for visualizing geospatial data, enable users to understand spatial relationships, retrieve information on distances, directions, and areas, reveal patterns, and quantify relationships. The momentum in the handling of digital geospatial data, which began in the 1980s, has further propelled the development of geographic visualizations, making them an indispensable tool in various fields (Kraak and Ormeling, 2020). In the next section, three different models are combined to develop a new framework that combines their insights, offering a comprehensive approach to geographic visualization in decision making.

New Map Communication Model

The new map communication model, introduced by Kraak and Ormeling in 2003 and expanded by Kent (2018), shown in Figure 3.7, reflects the evolution of cartography from a static unidirectional process to a dynamic user-influenced interaction. Traditional cartographic communication focused on optimizing the transfer of information from the map-maker to the user, assuming that the user's interpretation would closely align with the map's intended message. However, this process is prone to errors due to incomplete data, misrepresentation, or misinterpretation by the map maker and the map user (Kraak and Ormeling, 2020).
In this new model, user feedback plays a crucial role; it creates a loop between map makers and map users. Modern tools and technologies have allowed users to receive real-time feedback. Maps are no longer static artifacts, but living documents that can be refined, revised, and reshaped based on user input. This shift in map design emphasizes the importance of communication accuracy while recognizing that maps are subject to individual interpretation and should continually evolve to meet the needs of their audiences (Kent, 2018).

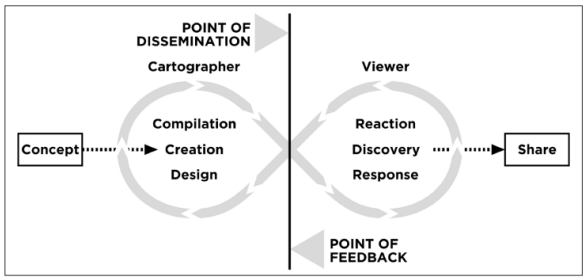


Figure 3.7: New map communication model (Kent, 2018).

Map Use Cube

The Map Use Cube, introduced by MacEachren (1995) and later expanded by Kraak and Ormeling (2020), provides a conceptual framework to understand how maps are used in different contexts based on three key dimensions. This cube, shown in Figure 3.8, positions various tasks of map use along axes that consider the level of interaction, the nature of the task, and the audience. The cube helps categorize how users interact with maps and visualize geospatial information.
The cube consists of three axes. First, there are the users. This axis reflects for whom the map is intended. Maps designed for private users are specified for individual users who interact with the data to generate insights and hypotheses. However, maps designed for public use are intended for wide dissemination and

aim to communicate information clearly and effectively to a broader audience.
The second axis describes the purpose of the map. Maps are used for exploration, where users interact with data to discover unknown patterns, relationships, and insights. However, maps present established knowledge to the audience in a straightforward manner, communicating findings rather than encouraging discovery.
The last axis describes the interaction of the map. The level of interactivity varies from low, where users passively receive information from static maps, to high, where users actively engage with the map interface.
The four visualization goals, exploration, analysis, synthesis, and presentation, are often placed along a diagonal within the Map Use Cube. At one end of the diagonal lies exploration, characterized by private, highly interactive tasks where users engage with maps to generate hypotheses and insights. At the other end, there is presentation, where knowledge is shared with a broad audience using low-interaction, explanatory visualizations. Analysis and synthesis tasks fall somewhere in between, with varying levels of user interaction and public reach.

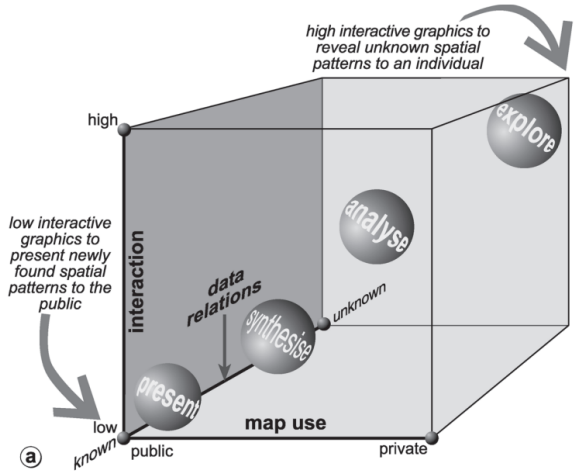


Figure 3.8: The Map Use Cube (Kraak and Ormeling, 2020; MacEachren, 1995).

Cognitive Load

Cognitive load is a key concept in cartography, referring to the total cognitive resources required to process information during tasks (Fang et al., 2020). In this context, cognitive load can be understood as the mental effort needed to acquire and interpret specific information. The cognitive load has two sides: A high cognitive load can make understanding a map more challenging, time-consuming, and less effective. On the other hand, a map with a low cognitive load tends to be easier to understand (Bunch and Lloyd, 2006).
The cognitive load theory consists of three types of loads. First, there is the intrinsic load, which refers to the complexity of the knowledge being acquired, regardless of the method used to acquire it (Sweller, 2011). In cartography, an example of intrinsic cognitive load is the complexity involved in understanding maps. Interpreting elevation contours, slopes, and landscape features requires a foundational grasp of geographical concepts and spatial relationships. This complexity exists regardless of how the information is presented, as the underlying knowledge itself is inherently challenging.
The second type is the extrinsic load, which involves cognitive demands that are not related to learning tasks (Furtado et al., 2018). In cartography, an example of extrinsic cognitive load is the impact of design elements on a map, such as excessive color use, complex symbols, or cluttered legends. These elements add unnecessary mental effort, making it harder to interpret essential information such as spatial relationships or specific landmarks. Although these design features do not directly contribute to understanding the content of the map, they can distract from the primary learning goals and increase cognitive strain.
The last type is the germane load, which supports the processing of the intrinsic load. In cartography, an example of germane cognitive load is the use of well-designed legends, color gradients, or clear labeling to highlight key information on a map. For example, a thoughtfully organized legend helps users quickly interpret symbols, while consistent color gradients can make it easier to distinguish elevation or land cover types.

In conclusion, the cognitive load theory plays a crucial role in cartography. A high cognitive load can hinder the user's understanding of maps, while a low cognitive load can indicate insufficient detail. An effective map design should aim to minimize extraneous cognitive load while enhancing germane load, thereby facilitating the interpretation of complex spatial information (Furtado et al., 2018).

Combining the New Map Communication Model, the Map Use Cube, and the Concept of Cognitive Load

When combining the new map communication model with the Map Use Cube and the cognitive load theory, a comprehensive framework emerges that enhances our understanding of map effectiveness in user interactions. The new map communication model emphasizes a dynamic relationship between map makers and map users, where real-time feedback allows for continuous refinement of maps to meet user needs (Kent, 2018). This interaction is crucial in managing cognitive load, as it recognizes that effective communication is not only about transmitting information, but also involves user interpretation and engagement.
The Map Use Cube further enriches this framework by categorizing maps according to their intended audience, purpose, and level of interaction (Kraak and Ormeling, 2020; MacEachren, 1995). This categorization helps identify how different users interact with maps, which, in turn, influences their cognitive load. For instance, maps designed for exploration encourage high interaction and are likely to foster a germane load, enhancing the user's ability to process and synthesize information. In contrast, maps intended for presentation may impose a higher extrinsic load if they lack interactivity, potentially overwhelming users with information without providing the opportunity for engagement (Fang et al., 2020).
By integrating these different theories and models, we can develop strategies to optimize map design that align with the cognitive load principles. Effective maps should not only communicate information clearly, but also facilitate user engagement and interaction, thus minimizing extraneous cognitive demands. This holistic approach ensures that maps serve their fundamental purpose of enhancing understanding and supporting decision-making processes, especially in critical scenarios where quick and accurate comprehension is essential.

3.3.3 Basic Principles of Indoor Map Design

Following the explanation of the New Map Communication Model, the Map Use Cube, and the concept of Cognitive Load, the next step is to explore the fundamental principles of map design to effectively visualize indoor environments as a cartographer. In the broader field of cartography, effective map design is essential for communicating spatial information. A cartographer must be clear about the purpose, topic, and intended audience of the map. Drawing from communication theory, they should ask: How do I convey specific information to a particular audience and is it effective? (Tyner, 2014).

To achieve effective communication, cartographers use various techniques in their designs. The following sections introduce these key techniques and explain how they help structure information, making complex spatial data more accessible to users.

To refine visual organization in 3D maps, cartographers employ hierarchical organization, which can be divided into intellectual and visual hierarchies. The intellectual hierarchy determines the order of importance of map elements according to the map's purpose. Based on this intellectual hierarchy, a visual hierarchy can be created, where the most important elements stand out (Krygier and Wood, 2024). Achieving this distinction often involves enhancing figure-ground relationships, where objects that need to stand out are clearly distinguished from the background. Various techniques can be used to achieve this figure-ground contrast, as can be seen in figure 3.9 by Krygier and Wood, 2024. Working with segmented point clouds further simplifies this process. By assigning distinct colors to specific segments of interest, cartographers can emphasize key figures more efficiently and dynamically. This capability makes it easier to highlight essential features within a complex indoor space relevant for emergency situations, such as entrances, staircases, or emergency exits, by adjusting their color to stand out against the rest of the environment. This approach not only supports clear figure-ground relationships, but also improves user navigation and SA in indoor settings (Nikoohemat et al., 2020).
In the context of 3D point clouds, visual hierarchy plays a crucial role in structuring spatial information effectively. Unlike traditional 2D maps, where elements such as size, shape, and texture are used to establish prominence (Bertin, 1983). 3D point clouds rely on additional depth-related factors such as perspective, occlusion, and layering to guide user attention; applying visual hierarchy in 3D point cloud environments is essential to improve SA and ensure critical elements are easily interpretable, especially in time-sensitive applications such as emergency response(Krygier and Wood, 2024).

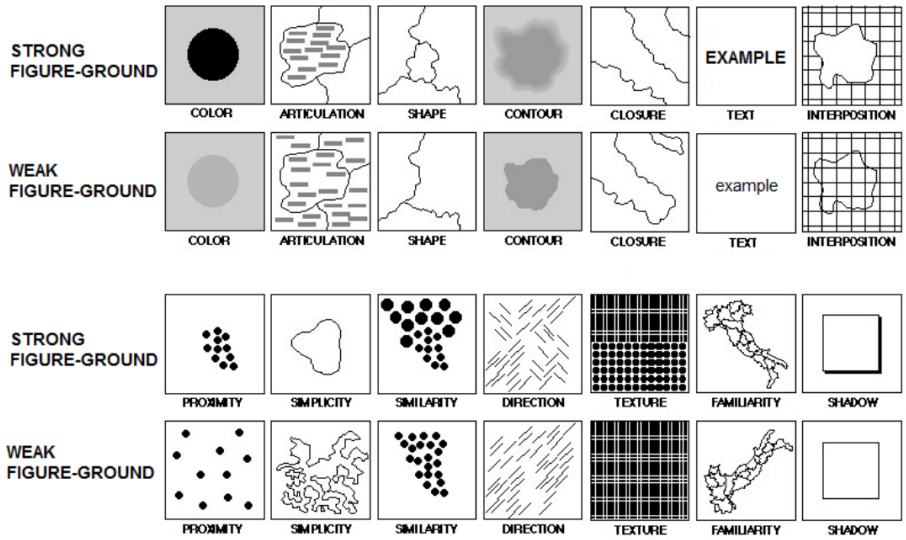


Figure 3.9: Visual Hierarchy: Techniques for establishing strong figure-ground (Krygier and Wood, 2024).

In contrast to traditional 2D visualizations, modern geovisualization techniques take advantage of technologies such as 3D displays and virtual environments. These technologies are especially relevant for indoor cartography, where buildings often have multiple floors, and these complex layouts cannot be easily represented using a single 2D top-down view. As humans naturally perceive the world in three dimensions, 3D visualizations align more closely with our cognitive processes, making them useful for tasks such as indoor navigation (Gupta et al., 2017). As a result, the cognitive approach to the use of maps differs significantly between indoor and outdoor contexts. Although outdoor maps utilize absolute coordinates based on cardinal directions (east, south, west, and north), indoor maps rely on relative coordinates that define positions in terms of "front," "back," "left," and "right." This fundamental difference shows that the principles of outdoor maps are not fully applicable to indoor maps and that specific principles for indoor visualization must be introduced.

The application of Bertin's graphic variables in 3D cartography plays a significant role in simplifying the complexity of indoor environments. As Nossum (2011) demonstrated in his IndoorTube visualization, topological relationships between floors can be emphasized over geometric accuracy, reducing clutter, and improving user navigation. IndoorTube focuses on showing only essential features, such as corridors, rooms, elevators, and stairs, while using different colors to distinguish between floors. This approach, though abstract, ensures that users can navigate complex multilevel buildings without being overwhelmed by unnecessary detail (Nossum, 2013). This method reflects the ongoing adaptation of cartographic principles to meet the needs of indoor environments, where users often prioritize ease of navigation and decision-making over geometric precision.

In addition, cartographic symbols and color schemes play an important role in making indoor visualizations more readable and functional. Research by Lorenz et al. (2013) shows that 3D representations with well-placed landmarks significantly improve user satisfaction and spatial understanding. Landmarks such as large open spaces, staircases, or elevators provide reference points that help users navigate a complex indoor environment. In such cases, simplicity in symbol design is crucial to avoid overwhelming users. Teixeira et al. (2021) emphasize that cartographic design in indoor spaces requires specific approaches that consider the unique dynamics of indoor environments, which differ from outdoor mapping in terms of scale, detail, and interaction.

In addition, advances in interactive cartography allow users to manipulate visualizations in real time, enhancing their understanding of indoor spaces. Nossum (2011) described the different levels of interactivity in indoor maps, ranging from static visualization to fully interactive maps, where users can rotate, zoom and manipulate the viewer to explore the indoor environment. This is especially important for indoor maps, where users may need to visualize multiple floors or retrieve detailed information on demand. Interactivity, combined with clear symbolization and landmark placement, ensures that users can navigate complex spaces more effectively.

Furthermore, when it comes to mapping indoor environments, traditional methods may not always be applicable. Unlike traditional outdoor spaces, indoor environments have unique characteristics, including smaller scales, higher volumes of information, clear segmentation, and enclosed spaces. Furthermore,

indoor visualizations face challenges such as occlusion (where objects block each other depending on the viewpoint) and the difficulty of estimating depth, which require careful design to ensure usability (Lorenz et al., 2013). However, as noted by Teixeira et al. (2021), indoor cartography is still in its early stages and many challenges remain. For example, there is no consensus yet on what kind of information should be prioritized for indoor navigation or how to best visualize multiple levels of a building. Although traditional cartographic principles such as abstraction and simplification are essential, they must be adapted to the unique challenges of indoor environments, where users need to perceive depth and movement differently than in outdoor spaces (Fang et al., 2020).

In conclusion, modern cartographic methods and geovisualization techniques continue to evolve, especially in the context of indoor environments. Although Bertin's original graphic variables remain fundamental, they have been extended and adapted for 3D and digital applications, offering new possibilities for improving decision-making in indoor spaces. As technology advances and user needs become more complex, the development of innovative visualization techniques, such as IndoorTube and interactive 3D mapping, will be essential to ensure that indoor cartography continues to support effective navigation and spatial understanding.

Chapter 4

Workflow Design

The workflow of this research can be divided into three distinct phases, each with specific objectives, data, and tools.

- **Phase 1:** focuses on the selection and integration of an existing semantic segmentation model for indoor (near) real-time segmentation.
- **Phase 2:** explores the user interaction with different types of visualizations, focusing on the different information needs between different users.
- **Phase 3:** involves evaluating the overall effectiveness of the visualizations based on interactions with the different users. This phase reflects on the quality of the segmented point cloud and the feedback provided by the users on the different visualizations of the segmented point cloud.

4.1 Phase 1: Model Selection and Integration

4.1.1 Sub-question 1: Model selection

To answer the first sub-question, a literature review is conducted on computer vision, semantic segmentation, various segmentation techniques, and nine state-of-the-art semantic segmentation models that have proven to be effective in indoor environments and are capable of processing point cloud data efficiently in (near) real-time.

After the introduction to computer vision, semantic segmentation and the different segmentation models in the theoretical framework, one of the discussed models is selected based on its ability to segment in (near) real time and its performance on datasets like S3DIS and ScanNet. The S3DIS dataset (Standford large-scale 3D Indoor Spaces dataset) is a comprehensive collection of 3D point cloud data used for semantic segmentation and object detection indoor spaces. The dataset was introduced by Armeni et al., (2016) and developed using 3D scans captured by RGB-D sensors that record both color and depth information. The dataset covers large-scale indoor environments, including office spaces, educational facilities, but also building elements such as walls, floors, and ceilings, and furniture such as chairs, tables, and sofas. In total, the dataset covers more than 6000, square meters of scanned indoor space, containing more than 215 million points (Armeni et al., 2016).
The ScanNet V2 dataset is a RGB-D video dataset designed for scene understanding and object recognition in indoor environments. The dataset was introduced by Dai et al., (2017), and is built from 2D and 3D data, capturing detailed indoor scenes. ScanNet consists of over 2.5 million RGB-D frames captured in 1513 scans of 707 unique real-world spaces.
The S3DIS and ScanNet datasets are often used to benchmark the performance of deep learning models for 3D object recognition and segmentation in indoor environments.
The S3DIS dataset is benchmarked in two ways: Area 5 mIoU and 6-fold cross-validation mIoU. For testing the performance of Area 5 mIoU, the dataset is divided into six areas, each corresponding to a separate part of the indoor environment. The models are trained in five of these areas and tested in Area 5, which is treated as the unseen test area. For the performance of the 6-fold cross-validation mIoU, the model is trained and tested iteratively, using one area as the test set while the remaining five are used for training, cycling through all six areas.

The ScanNet dataset is benchmarked in: Val mIoU and Test mIoU. For measuring Val mIoU performance, the dataset is split into training and validation sets. Performance is reported on the validation set, which is used during model development to assess how well the model is learning and generalizing before testing on unseen data. For the performance of the Test mIoU the test set contains unseen data used to evaluate the final model's performance. The Test mIoU is considered the definitive benchmark for comparing different models, as it reflects the model's real-world generalization capabilities.

The mIoU stands for mean Intersection over Union, which is a common evaluation metric used in segmentation tasks. The mIoU measures the accuracy of a model by comparing the predicted segmentation with the ground truth. First, the Intersection over Union (IoU) is calculated per class, and the formula is as follows:

$$\text{IoU} = \frac{\text{Intersection}}{\text{Union}}$$

Hereby, the intersection is the area where the predicted segmentation and the ground truth overlap and the union is the total area covered by either the predicted segmentation or the ground truth. Here, the output of the IoU measures how well the predicted segmentation aligns with the actual ground truth for each class.

The mean IoU (mIoU) is calculated by averaging the IoU values across all classes. In the formula, N represents the total number of classes, and the mIoU is calculated as follows:

$$\text{mIoU} = \frac{1}{N} \sum_{i=1}^{N} \text{IoU}_i$$

Based on the theoretical framework and the results on the S3DIS and ScanNet datasets, the first subquestion "Which existing deep learning segmentation model is best suited for (near) real-time point cloud segmentation in indoor environments?" is answered.

4.1.2 Sub-question 2: Reproducibility

After the selection of a suitable model for (near) real-time point cloud segmentation, the next step is to integrate the chosen model on a personal device.

The integration process starts by implementing the PointTransformer V3 + PPT (PVT3) model on a personal device, this is done by cloning the model's repository from GitHub. With the repository cloned, the model needs to be trained. However, an assessment of the capabilities provided by the CGI laptop (specifications in Figure 4.1) revealed that completing the training on this device would take an estimated 1800 hours, a time frame that is not feasible for this research. In order to overcome this, a pre-trained version of the PVT3 model is used.

Using a pre-trained model brings both advantages and limitations. The primary benefit is the significant time saved in the model deployment, allowing immediate testing and validation. Additionally, pre-trained models often leverage extensive training datasets, which can enhance model performance in generalized tasks. In the case of PTV3, the model was trained in Areas 1, 2, 3, 4, and 6 of the S3DIS dataset and tested in Area 5, which is widely regarded as the most challenging area within the dataset and is often used to test the performance of a model.
However, a limitation of using a pre-trained model is that the model classes are predefined, restricting the classification categories to the following 13 classes: ceiling, floor, wall, beam, column, window, door, table, chair, sofa, bookcase, board, and clutter. This means that if you want to classify specific objects related to improving SA, such as fire extinguishers and emergency exits, the model would need to be retrained with additional labeled data that include these objects. However, this is not feasible due to the significant resources required for retraining. As a result, while pre-trained models like PTV3 are valuable for general tasks, their limitations in accommodating domain-specific requirements, such as objects critical for SA, must be acknowledged.

After setting up the pre-trained model, the next step is to validate and test it using the S3DIS dataset. However, initial tests on the CGI-provided laptop revealed a hardware compatibility issue: the laptop's GPU has a compute capability of 7.5, while the PTV3 model requires 8.0 or higher. To resolve this, the model is implemented and tested on a CGI desktop computer with the necessary specifications (specifications in Table 4.1).

Device	Processor	RAM	GPU	OS
Precision 7560	Intel i7-10850H CPU @ 2.70GHz	32 GB DDR4	NVIDIA Quadro T1000 4 GB DDR6	**Windows 11 Pro**
Desktop PC	Intel® Core i7-14700K, 3.4 GHz	32 GB DDR4	ASUS TUF Gaming NVIDIA GeForce RTX 4080 Super OC Edition	**Windows 11 Pro**

Table 4.1: Specifications of the devices

After this implementation of the model on the CGI desktop computer, the pre-trained model is tested on the S3DIS dataset to get the same mIoU as the authors of the PVT3 model. This is done using the test.py script provided by the authors in the Pointcept repository (Appendix 1).
However, when the test.py script is run, the accuracy is lower than the accuracy reported by the authors. After some tests, the conclusion can be drawn that this was due to the use of a newer version of the code. After implementing an older version of the code (as suggested on GitHub), this problem is solved.

To address the second sub-question, "To what extent is it possible to integrate the chosen segmentation model on a personal device and reproduce the accuracy results reported by the authors on the training data?" precise testing is essential. The testing process begins by subsampling a dense point cloud into a sequence of voxelized point clouds, ensuring complete coverage of all points. Each subsampled segment is predicted independently, and the resulting predictions are aggregated to form a complete prediction of the entire point cloud. This approach provides a more accurate evaluation than a simple mapping or interpolation of the predictions. Furthermore, the testing framework incorporates test time enhancement (TTA), which improves the stability and robustness of evaluation performance by applying multiple augmentations during inference. This comprehensive testing method ensures a reliable comparison of the accuracy of the model on the personal device, offering insight into the extent to which the model's accuracy can be reproduced compared to the reported results.

4.1.3 Sub-question 3: Segmenting Self-Acquired Data

After successfully testing the model on the S3DIS dataset, the next step is to assess its performance on self-acquired datasets. For this purpose, the model was tested on a scan of the author's house, featuring two distinct types of scans. The first scan was conducted thoroughly, ensuring high-quality data capture by allowing sufficient time to scan each detail and object in the environment. In contrast, the second scan simulated the conditions faced by first responders, where time was limited, and the scan was carried out at a faster pace. This second type of scan represented a lower quality dataset, reflecting the challenges of rapid data acquisition in fast-paced emergency situations. This distinction in scan quality serves multiple purposes. First, it allowed the model to be tested on self-acquired data, providing insight into how the model performed in different scenarios. Additionally, it addressed a critical question for CGI's use case: could the model reliably segment lower-quality data in a way that remained useful for first responders or similar applications? By exploring these aspects, the evaluation aimed to determine whether the model was adaptable and robust under varying field conditions.

The scans are acquired using an iPhone 11 Pro, using the RTAB-Map 3D LiDAR Scanner app, which allows the precise capture of LiDAR data. The data gathering process begins with a thorough scan (shown in Appendix 2) of the author's home, covering three floors and seven rooms. This scan results in a point cloud with a total of 6.208.869 points. The second, faster-paced scan (shown in Appendix 3) covers the same floors and rooms, producing a point cloud with a total of 2.103.268 captured points. In addition to these scans, a third scan was conducted on the eighth floor of the CGI office to evaluate the proof-of-concept in a different environment. This point cloud, which contains 14.074.311 points, serves as a test data set to assess the effectiveness of different visualization techniques outside the home environment (shown in Appendix 6).

To ensure that the self-acquired datasets are compatible with the pre-trained PointTransformer model, a preprocessing pipeline has to be developed. The script, preprocessing_normals.py (shown in Appendix 4), transformed the raw point cloud data into a format consistent with the S3DIS dataset, facilitating seamless integration into the existing workflow. The preprocessing process starts by loading the raw point cloud and extracting essential features such as 3D coordinates and RGB colors. As LiDAR scans often lack normals, they are calculated directly using Open3D's hybrid search method, which ensures accurate estimation of surface orientation based on local geometric relationships.
To further refine the data, the script applies a bounding box to crop the point cloud, removing extraneous points outside the scanned area. This step focuses the dataset on the relevant indoor features while maintaining the integrity of the scanned environment. Once processed, the data are saved as.npy files following the S3DIS format, with separate files for coordinates, colors, and normals. Since self-acquired datasets lack ground-truth annotations, placeholders are used for semantic and instance labels.

The inference process is carried out using a dedicated script designed to process the pre-processed datasets with the PointTransformer model. This script, inference.py (shown in Appendix 5) ensures that the data are correctly transformed to meet the model input requirements, applying grid sampling and normalization to standardize spatial scales and orientations. Once prepared, the scans were passed through the segmentation model, where each point in the cloud was classified into one of the predefined semantic classes. A color map was applied to the segmentation output, assigning a unique color to each class for visualization purposes. Finally, the results, including coordinates, RGB colors, normals, and segmentation labels, were saved and integrated into the PVT3 model repository for further analysis and refinement.

Based on the outcomes of this process, the third sub-question: "How can the selected segmentation model be integrated and tested on self-acquired data?" is answered.

4.2 Phase 2: Visualization and Interaction

After testing the model on self-acquired data, the focus of the research is on user interaction and map design. In particular, on how to design clear and simple visualizations that can increase SA for first responders by visualizing the segmented point cloud data.

4.2.1 Sub-question 4: Defining User Needs

Effectively communicating the segmented data to different user groups is a critical step in this research, as it ensures that relevant information reaches those who need it most in a clear and actionable manner. To create user-centered and impactful visualizations, it is essential to first explore the different needs of each user group. Thus, the focus is on understanding the needs of three primary user groups; control room personnel, local responders, and commanders in the Command Place Incident (CoPI). Each group has different roles and responsibilities during emergency incidents, requiring customized visualizations for their specific tasks.

Control room personnel serve as the first point of contact for incident reports and are responsible for coordinating alarms, dispatching emergency services, and monitoring the ongoing situation. Their visualizations must prioritize rapid and accurate dissemination of information to support effective communication and coordination between field units and command structures. Higher-level roles within the control room, such as duty officers or scenario planners, require more detailed and dynamic visualizations that integrate layered data, enabling them to predict risks and coordinate multiagency response during high GRIP-level incidents.
Local responders, including firefighters, police officers, and paramedics, operate directly on site and address emergencies in dynamic and high-stress situations. Visualizations must provide actionable insights tailored to their specific tasks, such as building layouts, escape routes, and hazardous areas. Simplified and intuitive visual tools are essential to ensure that responders can quickly assess risks and perform their tasks effectively. (Near) real-time updates further enhance their ability to adapt to evolving conditions.
CoPI commanders operate at a strategic level, overseeing multi-agency coordination and managing resource allocation during incidents. Their visualizations must provide a comprehensive overview of the situation, consolidating (near) real-time updates from multiple sources, including responder locations, incident progression, and potential risks. These visual tools should enable collaboration between agencies and support the efficient distribution of tasks and resources.

Before proceeding with the development of visualizations, it is crucial to first gain a comprehensive understanding of the specific needs of each user group. This understanding is achieved through a combination of interviews with professionals from the safety regions of Rotterdam and Amsterdam, as well as with fire brigade and police experts, and Rotterdam control room personnel. In addition, an examination of the relevant scientific literature and the specific literature on the different user groups is required. The insights gathered from these sources provide a detailed overview of the requirements and expectations of the user groups. These findings are presented in the results chapter, where the specific needs of control room personnel, local responders, and CoPI commanders are analyzed in greater depth. These findings will serve as the basis for the design and implementation of user-centered visualizations aimed at increasing SA during emergency incidents.

Based on these findings, the fourth sub-question is answered: What are the key information needs during emergencies to support decision making and situational awareness for various stakeholders?

4.2.2 Sub-question 5: Effective Visualization

To address the fifth sub-question the methods for developing effective visualizations of the segmented 3D point cloud data are guided by three core theories that were introduced in the related work section: the map use cube, cognitive load theory, and the new map communication Model. These frameworks are applied systematically to understand the needs of different user groups and to design visualizations that improve SA and decision-making during emergency responses.

First, the map use cube is used to systematically categorize the intended user groups along three dimensions: user type, purpose of use, and level of interaction. This framework is applied to define the specific visualization requirements for each user group involved in emergency responses. First, control room personnel are identified as analytical users that required moderately interactive visualizations that allowed toggling between different data layers, such as structural segmentation. Their visualizations needed to accommodate dynamic workflows and the synthesis of large volumes of data. Second, local responders were classified as exploratory users with low levels of interaction, necessitating simplified visualizations that could be quickly interpreted in time-sensitive situations. Finally, CoPI commanders were placed as strategic users that required high-interaction visualizations to integrate and synthesize multiple data layers for comprehensive situational overviews.

Second, the cognitive load theory informed the design principles for creating visualizations that balance between complexity and usability.
To manage intrinsic cognitive load, visualizations must be designed to present essential structural and navigational elements, such as walls, doors, and hazard zones, in a clear and simple way. The principle of progressive disclosure was applied, enabling users to start with a simplified overview of the data and gradually access more detailed layers as needed.
The extraneous cognitive load, which arises from poorly designed visual elements, must be minimized by streamlined visualizations. This includes eliminating redundant features, reducing the excessive use of colors, and designing intuitive legends that dynamically display only the symbols and colors relevant to the active data layers. In addition, care was taken to ensure that the visual layout and symbology were simple yet effective, allowing users to focus on their tasks without unnecessary distractions.
To enhance the germane cognitive load, visual emphasis techniques such as color gradients, contrast, and interactive elements are used. These design features aimed to support users in synthesizing and applying information, particularly when identifying critical features such as evacuation routes or hazardous zones. These principles ensured that visualizations not only communicated segmented point cloud data effectively, but also supported the cognitive processes required for decision making in emergency scenarios.

At last, the new map communication model is used to facilitate an iterative design process, focusing on active interaction between mapmakers and users. This process began with the creation of initial visualization prototypes, which were tested through user-interaction sessions, with CGI colleagues that work within the emergency response sector, and explored the segmented 3D point cloud data. During these meetings, feedback mechanisms were implemented to capture user input on usability, clarity, and relevance of the visualizations. Participants are encouraged to provide information on how effectively the visualizations supported their decision-making processes and to suggest refinements to improve the final visualization design.

Based on the use of the map use cube, the cognitive load theory and the new map communication model, a cohesive understanding of user interaction and map usability has been developed to address the fifth sub-question: How can scientific principles and cartographic methods be applied to effectively communicate segmented 3D point cloud data to diverse user groups?

4.2.3 Sub-question 6: Proof of Concept

After researching the different methods and defining the different user needs, the next step is to develop a proof of concept to evaluate how the segmented data could be effectively communicated to the different groups of users. This process is structured into different steps that involved creating a design, segmenting a self-gathered point cloud, and developing various levels of visualization to address the diverse needs of the users.

The first step involved creating a conceptual design for visualizing the segmented point cloud. This is done using Krita, an open source digital illustration software used for graphic design and digital art. Krita supports a wide range of tools, such as customizable brushes and vector layers, making it suitable for creating detailed 2D visualizations.
To guide the design process, a screenshot of the self-gathered point cloud is imported into Krita. This screenshot serves as the base layer, allowing for accurate alignment of the conceptual design with the actual spatial configuration of the environment. The distinct categories, such as walls, ceilings, floors, furniture, doors, and other components, are then overlaid onto the point cloud image and visually represented using unique colors and symbols. This approach ensured clarity and ease of understanding while maintaining consistency with real-world data.
Multiple design concepts were created and tested with relevant users to gather feedback. Based on these insights, a final visualization approach is developed for the point clouds of both the house and the CGI office.

Due to limitations with the PointTransformer V3+PTT model and its inability to process the self-gathered point cloud effectively, manual segmentation of the self-gathered point cloud is carried out using CloudCompare. CloudCompare is a point cloud processing software that allows users to interactively classify and segment point clouds based on visual and spatial characteristics. The self-gathered point cloud is segmented into predefined categories by manually assigning labels to the points.
This segmentation process involves identifying and isolating every object in different rooms, including objects such as beds, heating elements, lights, and even decorative items like a globe. These segmented objects are then grouped into broader classes, which are further refined into progressively detailed levels. In the end, three levels of visualizations are created to test their impact on SA. For each visualization level, different proof-of-concept designs are developed using various color schemes to determine which colors were most effective in conveying the segmented data.

- **Basic Visualization:** This level consists of seven fundamental classes: walls, ceilings, floors, doors, stairs, windows, and objects. This basic categorization provides an overview of the structural and functional components of the environment.
- **Obstacle Differentiation:** In the second level, a distinction is introduced between different types of obstacles, categorizing them as electrical devices or furniture. This aims to enhance the understanding of potential hindrances in an emergency scenario
- **Detailed Object Classification:** The final level includes more granular distinctions within the objects and furniture categories, breaking them into subcategories. This more detailed visualization aims to explore whether increased specificity in data representations improved SA.

These progressively detailed visualizations form the basis for testing whether varying levels of detail and categorization in segmented data could improve SA for end users. The iterative process of segmentation and visualization design allows for the creation of a comprehensive set of proof-of-concept visualizations, tailored to address different levels of user needs.

For each of these different visualizations levels, three distinct color schemes are implemented to investigate their effect on SA. The choice of color schemes is critical. as it influences how quickly and accurately users can interpret the segmented data. The color schemes are defined as follows:

- **Functional Color Scheme:** This scheme emphasizes clarity and practical usability by assigning colors according to their functional significance. For example, critical navigational elements, such as floors and doors, are colored green and yellow to denote safe pathways and access points, while hazardous components, such as electrical devices, are colored red. Walls and obstacles are depicted in dark gray to denote barriers. This approach is particularly beneficial for first responders who require rapid and clear information under high-stress conditions.

- **Contrast Based Color Scheme:** Focusing on maximum differentiation, this scheme employs highly saturated and opposing colors to ensure immediate visual distinction between categories. For example, doors may appear in magenta, windows in cyan, and obstacles in bright orange.
- **Realistic Color Scheme:** In an effort to achieve intuitive and naturalistic representation, this scheme mirrors the true colors of objects found in real-world settings. Elements such as doors and wooden furniture are shown in various shades of brown, metallic components are shown in gray, and screens are rendered in black.

For the proof of concept, scans of the authors' house and a CGI office are processed; for each of the three visualization types, all three color schemes, functional, contrast based, and realistic are applied, resulting in nine visualizations per location and a total of 18 visualizations overall.

By following the steps outlined above the sixth sub question "How can a proof-of-concept visualization be developed to demonstrate the effective communication of segmented point cloud data, tailored to different user needs?" will be answered.

4.3 Phase 3: Reflection and Evaluation

The final phase of this research evaluates the effectiveness of the visualizations developed in the second phase. This phase focuses on analyzing user feedback to assess how well the 3D visualizations created in Cloud Compare support SA, user interaction, and decision making in (near) real-time scenarios.
Through structured evaluations, user responses are examined to determine the strengths and limitations of the different visualization approaches. This process provides insights into potential improvements, ensuring that the visualizations align with practical applications and effectively convey critical spatial information. The findings of this phase contribute to the refinement of the design of future visualization methods, making them more suited to real-world operational needs.

4.3.1 Sub-question 7: Reflection and Evaluation

To assess the effectiveness of the 3D visualizations of the segmented point cloud, a structured evaluation is carried out through expert interviews. The primary objective of this evaluation is to determine how well these visualizations improve SA, facilitate decision making, and support user interaction in (near) real-time emergency scenarios.

The evaluation involves qualitative interviews with professionals from multiple domains who work in the emergency response sector. Participants include experts from Rotterdam and Amsterdam safety regions, fire brigade personnel, Rotterdam control room personnel, and CGI colleagues with expertise in geospatial technology and visualization. By involving professionals in these different fields, this research ensures a comprehensive assessment of how different stakeholders interpret, interact with, and use visualizations in operational contexts.
During the evaluation, participants are presented with the proof-of-concept visualizations developed in the second phase. These visualizations include three levels of segmentation, basic, obstacle differentiation, and detailed object classification, each rendered in three distinct color schemes: functional, contrast based, and realistic. The interviews aim to capture user perspectives on the clarity, usability, and overall effectiveness of these visualization methods in supporting SA.
The interviews follow a semi-structured format, allowing for open-ended responses while ensuring key topics are covered. Participants are asked about the general effectiveness of the visualizations in conveying spatial information, with a focus on whether different levels of segmentation improve or hinder their ability to interpret the indoor environment. Specific questions address the role of color schemes in object recognition and differentiation, examining whether functional, contrast-based, or realistic color assignments are most effective for their decision-making processes.

Through these interviews and tests, the last sub-question is answered "How effective are the proof-of-concept visualizations in increasing situational awareness through the communication of segmented point cloud data?"

Chapter 5

Results

5.1 Phase 1: Model Selection and Integration

5.1.1 Sub-question 1: Model Selection

This first section evaluates nine state-of-the-art deep learning models for indoor semantic segmentation of point clouds, focusing on two primary criteria: segmentation accuracy and their applicability for (near) real-time performance. The evaluation is based on quantitative results across two benchmark datasets, including S3DIS and ScanNet. The aim is to identify the best-suited model for tasks requiring accuracy and efficiency in indoor environments.

Segmentation Accuracy

Figure 5.1 summarizes the segmentation accuracy of the various models on the S3DIS (Area 5 and 6-fold cross-validation) and ScanNet datasets reported by the different authors. The results are measured in terms of mIoU, which is the key metric to evaluate the segmentation accuracy.

Table 5.1: Model comparison on S3DIS and ScanNet benchmarks

Model	S3DIS mIoU Area 5	S3DIS 6-fold	ScanNet val mIoU	ScanNet test mIoU
PointNet	41.1	47.6	X	X
PointNet++	53.5	54.5	53.5	55.7
DGCNN	47.9	56.1	X	X
Point-Voxel CNN	X	59.0	X	X
PointNext	70.5	74.9	71.5	71.2
Oneformer3d	72.4	75.0	76.6	X
Superpoint Transformer	68.9	76.0	X	X
Point Transformer V3	73.4	77.7	77.5	77.9
Point Transformer V3 + PTT	74.7	80.8	78.6	79.4
Point-SAM	X	86.2	X	X

The results clearly demonstrate the progression in segmentation performance across the different segmentation models, reflecting the theoretical advances discussed in the related work section. Early models like PointNet and PointNet++ introduced foundational techniques for processing point cloud data, particularly the use of MLPs and symmetric pooling functions to take into account the unordered nature of point clouds. However, these models struggled with capturing complex spatial relationships, which is critical to understanding complicated indoor environments. Specifically, PointNet, with its simplistic approach, achieved a mIoU of 41.1 on the S3DIS Area 5 dataset, reflecting its limited ability to handle local and global spatial features (Qi et al., 2017a). The introduction of hierarchical grouping in PointNet++ marked a step forward, resulting in a mIoU of 53.5. However, as mentioned in the related work section, the global max-pooling operation used in these models often led to the loss of fine-grained details, affecting their performance in cluttered or large-scale scenes (Qi et al., 2017b).
The transition to more advanced architectures, such as OneFormer3D and PointNeXt, represents a significant increase in performance by addressing the shortcomings of previous models. OneFormer3D,

leveraging a multitask framework, aligns with the framework's emphasis on integrating semantic, instance, and panoptic segmentation tasks. This architecture demonstrated its ability to handle various tasks effectively, achieving an mIoU of 76.6 on ScanNet validation data (Kolodiazhnyi et al., 2024). Similarly, PointNeXt introduced modernized training strategies, including relative position normalization and improved model scaling, which allowed it to achieve a score of 74.9 mIoU on the S3DIS six-fold validation. These innovations reflect the theoretical focus on enhanced feature extraction and hierarchical learning, which enables the model to perform well even in complex indoor environments (Qian et al., 2022).
As seen in 5.1, transformer-based models emerge as the most effective models, consistent with the discussion in the related work section. Point Transformer V3, for example, leverages self-attention mechanisms to balance local and global feature integration, achieving 77.7 mIoU on S3DIS 6-fold validation and 77.9 mIoU on ScanNet test data. The enhanced version, Point Transformer V3 + PPT introduces pre-trained tokenizers to further improve accuracy, reaching 80.8 mIoU on S3DIS 6-fold and 79.4 mIoU on ScanNet test data. These results underscore the effectiveness of attention-based architecture in handling complex large-scale indoor scenes (Wu et al., 2024).
The standout performance of Point-SAM, achieving the highest accuracy of 86.2 mIoU on S3DIS 6-fold validation, aligns with the theoretical framework's emphasis on the potential of promptable architectures for generalization. The Voronoi-based tokenizer, central to Point-SAM's design, efficiently segments point clouds by preserving local structures while maintaining computational efficiency. In addition, its zero-shot segmentation capability positions it as a versatile model for various indoor applications, particularly where adaptability is crucial (Zhou et al., 2024).
When analyzed in the context of the related work section, these findings validate the progression from foundational methodologies to state-of-the-art advances in semantic segmentation. The integration of long-range dependencies, hierarchical feature learning, and multitask optimization has driven the development of models capable of addressing the complexities of indoor environments. The theoretical principles outlined underpin the practical success of models such as Point Transformer V3 + PTT and Point-SAM, establishing them as leading models for achieving both accuracy and efficiency in point cloud segmentation tasks.

(Near) Real-Time Applicability

While accuracy is a critical factor in evaluating segmentation models, their applicability in (near) real-time scenarios is equally important, especially for time-sensitive applications such as disaster response and emergency management. As emphasized in the related work, processing speed and SA are crucial for first responders who rely on rapid and accurate data interpretation to make informed decisions in dynamic environments. Models must therefore strike a balance between segmentation accuracy and computational efficiency to meet these requirements effectively.
The early segmentation models, such as PointNet and PointNet++, laid the groundwork for processing 3D point cloud data (Qi et al., 2017a, 2017b). However, their architectures were not designed with real-time performance in mind. PointNet, for example, uses a global max-pooling operation to aggregate features, which limits its ability to process local spatial relationships efficiently. Although PointNet++ introduced hierarchical feature extraction, which slightly improved both accuracy and computational efficiency, its reliance on a global feature representation still made it unsuitable for (near) real-time tasks. The computational inefficiencies and lack of scalability of these models become apparent when applied to large-scale indoor environments, where processing delays can hinder time-critical decision-making.
With the introduction of more advanced architectures, such as OneFormer3D and PointNeXt, a significant improvement was marked in segmentation accuracy and computational performance. OneFormer3D with its multi-task framework enables simultaneous semantic, instance and panoptic segmentation, making it highly versatile. However, the computational overhead introduced by its complex architecture limits its scalability for real-time deployment (Kolodiazhnyi et al., 2024). Similarly, PointNeXt, while more efficient due to modernized training strategies and its hierarchical design, still requires substantial computational resources, resulting in challenges for resource-constrained environments such as edge devices or real-time cloud-based systems (Qian et al., 2022).
Transformed-based models such as Point Transformer V3 and its enhanced version, Point Transformer V3 + PPT, offer significant advancements in scalability and computational efficiency. These models leverage serialized attention and patch-based grouping techniques, which reduce the computational complexity associated with processing large-scale point clouds. Serialized attention enables the models to focus computational resources on relevant spatial regions, thereby improving processing speed without compromising accuracy. Additionally, the use of pre-trained point tokenizers further improves scalabil-

ity by enabling the models to reuse learned embeddings, reducing the need for extensive computational resource during inference. These advancements make transformer-based models promising candidates for (near) real-time segmentation, particularly in scenarios where rapid and accurate processing of large datasets is required (Wu et al., 2024).
Point-SAM, while achieving state-of-the-art accuracy, faces inherent challenges in real-time deployment due to the computational demands of its Voronoi-based tokenizer. This tokenizer is designed to reduce memory usage and enhance scalability by segmenting the point cloud into local regions based on geometric proximity. Although effective in improving accuracy and preserving local structures, the computational cost associated with generating and processing these tokenized regions is significant. This increase in computational cost can introduce delays that prevent Point-SAM from achieving seamless real-time interaction, in particular, in large-scale indoor environments where dense point clouds must be processed rapidly (Zhou et al., 2024).
However, despite the theoretical advances and promising performance of these models, achieving true (near) real-time segmentation remains a challenge with current hardware and processing capabilities. Models like Point Transformer V3 + PPT and Point-SAM demonstrate the potential to bridge this gap through architectural innovations and scalability-focused strategies. However, their practical implementation in real-time systems requires further optimization. Techniques such as hardware acceleration, algorithmic refinement, and parallel processing can potentially address these challenges. Additionally, as noted in the related work section, the integration of adaptive techniques such as progressive inference (where lower-resolution segmentation results are incrementally refined) could further improve real-time applicability.

5.1.2 Sub-question 2: Reproducibility

The integration of the selected model PTv3, on a personal device was carried out after resolving the hardware limitations related to the GPU of the laptop provided by CGI. The PTv3 model, which had been pre-trained on the S3DIS dataset, was implemented on a desktop computer with the necessary GPU specifications, ensuring compatibility with the model's requirements. The integration process involved cloning the model repository from GitHub, setting up the environment, and loading the pre-trained weights.

The first step in the validation process was to test the model's performance on the S3DIS dataset, specifically focusing on whether the model could reproduce the accuracy results reported by the authors. As detailed in Table **??**, the authors reported an mIoU of 80.8 on the 6-fold S3DIS dataset and mIoU 74.7 on area 5. To verify the reproducibility of these results, the model was evaluated on the same dataset using the pre-trained weights.

The following result was obtained:

- Semantic Segmentation (Area 5): The pretrained PTv3 + PTT model achieved a mIoU of 0.7483, which closely matched the reported value of 74.7%. This confirms that the model's performance on this area is consistent with the results presented in the original paper.

The test script was unable to complete the evaluation in all rooms, successfully processing only the first 28 of the 68 rooms in the dataset. This limitation is most likely attributed to hardware constraints, as the authors of the model conducted their tests using an RTX 4090 GPU, while my setup utilized an RTX 4080.
Despite this difference, all other aspects of the implementation were reproduced exactly as described in the original study. This result suggests that, when integrated into personal hardware, the PointTransformer V3 model can achieve performance similar to that reported by the authors on the training data. Furthermore, the specific hardware configuration of the CGI desktop, particularly the GPU with compute capability 8.0, played a critical role in ensuring the model's stable execution without performance-related errors.

To visualize the segmented S3DIS dataset, a visualization part was added to the inference script, as shown in Appendix 5. In this script, each class within the dataset is assigned a distinct color, which facilitates the differentiation of the segmented objects within the point cloud. The color map is predefined for various semantic classes such as 'ceiling', 'floor', 'wall', 'beam', 'column', and others, with each class represented by a specific RGB color. Following the model's inference, the output point cloud is processed by mapping the segment labels to their corresponding class colors. The resulting point cloud (shown in figure 5.1), enriched with class-specific color annotations, is then saved in PLY format for further visualizations. Additionally, the segmented point cloud is displayed using Open3D, providing an interactive 3D visualization of the segmented environment. This approach ensures that the segmented point cloud is both visually accessible and easy to interpret for analysis and evaluation.

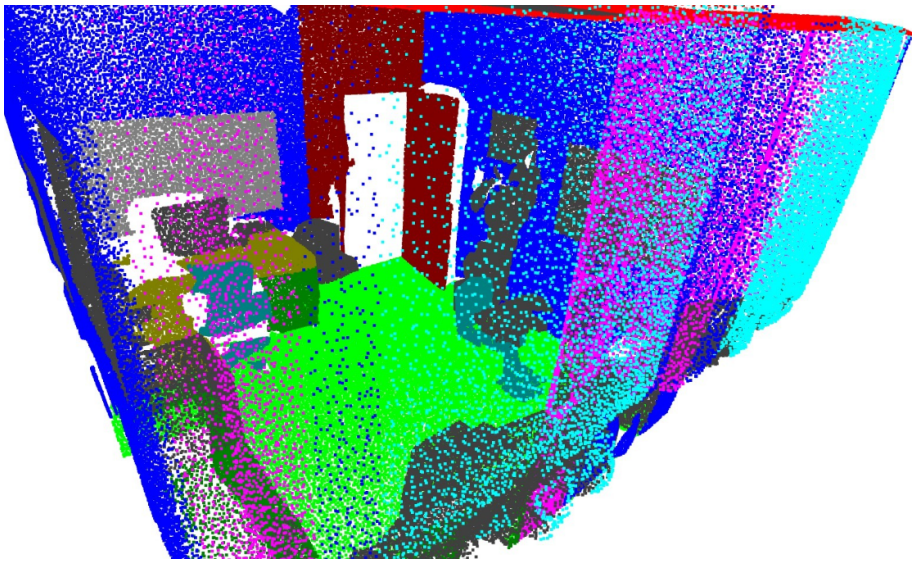


Figure 5.1: Visualization of Segmented Point Cloud of S3DIS dataset

In conclusion, the integration of the PTv3 model onto a personal device was successfully accomplished, with no significant technical issues arising once the appropriate GPU specifications were met. The model, when evaluated using the pre-trained weights on the S3DIS dataset, demonstrated near-identical performance to the results reported by the authors. These findings highlight the robustness of the model and its ability to maintain accuracy when deployed on personal hardware, provided the device meets the necessary GPU specifications. Ultimately, the results support the feasibility of integrating the chosen segmentation model in personal devices without compromising its performance.

5.1.3 Sub-question 3: Segmenting Self-Acquired Data

After validating the reproducibility of PointTransformer V3 + PTT on the S3DIS dataset, the next step was to test the performance of the model on self-acquired data. This sub-question aimed to evaluate the model's ability to segment beyond the structured and noise-free S3DIS dataset to more complex, real-world datasets. The self-acquired data, collected using an iPhone LiDAR scanner, represent diverse and less controlled scanning conditions, simulating scenarios more relatable to first responders.
First responders often operate in dynamic and unpredictable environments, where the ability to capture and interpret spatial data quickly and under suboptimal conditions is critical. Unlike the highly controlled and meticulously labeled S3DIS dataset, the self-acquired data reflect the challenges faced in the field, such as lower point densities, uneven coverage, and increased noise due to time constraints and equipment limitations. Using an iPhone LiDAR scanner, the dataset closely mimics the type of scanning technology that could be deployed in real-world emergency scenarios, where compact, mobile, and readily available devices are favored. Testing the model under these conditions provides valuable insight into its practical applicability and robustness in scenarios where lives and safety depend on timely and accurate segmentation outputs.
Despite using a robust preprocessing pipeline and achieving successful inference on S3DIS, the results on self-acquired data revealed significant challenges. These limitations ultimately led to a decision to shift focus to the next stage of the research, which focuses on visualization and practical recommendations.

To prepare the self-acquired datasets for segmentation, the raw point cloud was preprocessed using a custom Python script (Appendix 4). This preprocessing step ensured compatibility with the model's input requirements, transforming the data into the S3DIS format. The script extracted features such as 3D coordinates, RGB colors, and computed normals using Open3D, addressing the lack of normals in the raw LiDAR scans. Additionally, a bounding box was applied to crop the point cloud isolating relevant indoor features and removing extraneous data outside the scanned area. This careful preprocessing allowed the data to be structured and standardized for inference.
The inference script (Appendix 5) used the preprocessed datasets to segment the point clouds. The script applied transformations such as grid sampling and normalization, aligning the data with the model's trained parameters. After processing, the data were passed through the PointTransformer model. which predicted semantic labels for each point in the point cloud, and which predicted semantic labels for each point in the point cloud. The results were then visualized using a predefined color map that assigned distinct colors to each semantic class.
However, all segmentation attempts on the self-acquired LiDAR failed completely, and the model erroneously labeled every point as "clutter." After careful analysis, I have determined that the root cause of this failure was mistakes in my implementation of the model pipeline, stemming from my own limited programming experience. In essence, the model was not applied correctly to the new data, leading to a total breakdown in its ability to distinguish object classes. It is true that the self-collected point clouds differed from the S3DIS training data in several ways, for example, they contained more noise, had a lower point density, and featured different indoor layouts and object geometries. These discrepancies would naturally make generalization more difficult for any model and may have further hindered its performance on the new scans. However, these factors are ultimately secondary. The primary issue was the flawed implementation itself; my lack of expertise in fine-tuning and debugging the pipeline for this novel dataset meant that critical errors went unfixed, ultimately causing the segmentation to fail on the self-acquired data.

Given the persistent challenges and limited progress despite considerable effort, a strategic decision was made to move onto the next phase of this research. This phase focuses on developing robust visualization techniques and deriving practical insights from the findings rather than continuing to refine the segmentation process on self-acquired datasets. This decision reflects the pragmatic prioritization of time and resources, ensuring that the project remains aligned with its broader objectives.

In conclusion, while the preprocessing and inference pipelines were validated on the S3DIS dataset, the self-acquired datasets highlighted significant limitations in the model's ability to generalize to noisy, real-world data. Despite considerable effort to address these challenges, the resolution of the issues proved to be beyond the current scope of this research and my capabilities. These findings underscore the importance of training models on more diverse datasets and tailoring methodologies to address the complexities of field conditions, although such improvements would require resources and expertise beyond what was feasible for this project.

5.2 Phase 2: Visualization and Interaction

5.2.1 Sub-question 4: Defining User Needs

This section analyzes the distinct needs and requirements of the three main user groups involved in the emergency response: control room personnel, local responders, and commanders of the Command Post Incident (CoPI). The insights presented here are primarily based on interviews conducted with professionals working within the emergency response sector, including personnel from control rooms and field units. These interviews were complemented by the literature on emergency management practices and the theses of Bart-Peter Smit (2020) and Tom van der Meer (2018). Understanding the specific data requirements of each user group is fundamental to developing effective tools that support decision making during emergencies.

Control Room Personnel

The control room functions as the central coordination center during emergency incidents, facilitating the management of distress calls, assessing the severity of situations, dispatching the appropriate responders, and maintaining communication between field units and higher-level command structures. The role of the control room is crucial to ensure the efficient and effective deployment of emergency services, where personnel interact with visual data to support decision making at different stages of the emergency.
The personnel in the control room have specific roles that vary depending on their position within the organization and the stage of the emergency. These roles include intakers, dispatchers, and higher-level personnel such as officers of service (OvD) and calamity coordinators (CaCO). Each role has different responsibilities, and each role has specific visual data needs that support their tasks, especially when it comes to 3D segmented point clouds, which provide detailed and spatially accurate representations of incidents that support decision making.

Intakers

Intakers are the first point of contact in the emergency response process. Their primary responsibility is to respond to incoming emergency calls as quickly as possible while ensuring that crucial information is gathered efficiently and accurately. This includes identifying the nature and severity of the incident, verifying the location, and collecting essential details from the caller to facilitate an appropriate response. Given the high-pressure environment in which they can operate, intakers must quickly assess incomplete or chaotic information while maintaining clear communication with distressed callers.
Currently, intakers are mainly relying on audio communication and textual input systems to record incident details. Their role is focused on information intake rather than operational coordination, which means that they do not actively engage with visual data for decision making. Since their primary task is to extract verbal descriptions from callers and relay this information to dispatchers, access to 3D point cloud data would not significantly improve their workflow. Instead, their effectiveness depends on structured questioning techniques and efficient data entry into dispatch systems.

Dispatchers

The role of dispatchers in the control room is multifaceted and critical to effective emergency response coordination. Once the intakers collect and verify the essential details of an emergency, the dispatchers take over, ensuring that the appropriate emergency units are assigned to the incident. Their tasks include assessing the priority and classification of incidents, adhering to established procedures, and providing actionable guidance to emergency responders. Dispatchers ensure that vital information is communicated to emergency units and relevant stakeholders, including police, fire services, ambulance teams, and other crisis management entities (NIPV, 2020). Their SA is built on information from the intaker and on geographic data, as confirmed during an interview with the control room. Dispatchers rely on 2D maps with real-time data layers, street view imagery, and live camera feeds to establish a comprehensive understanding of the situation. These tools allow them to determine the best response strategies and ensure that the units arrive at the scene well informed.

During interviews with control room personnel, a key takeaway was that “less is more” when it comes to visual information. Dispatchers work in high-pressure environments where clear and concise data is critical to making split second decisions. Any additional visualization must improve decision making without introducing unnecessary complexity or cognitive overload.
Although dispatchers currently work effectively with 2D maps, the integration of (near) real-time 3D

point clouds, if carefully designed, could provide valuable additional context. RGB point clouds could serve as a supplementary tool by offering a spatially accurate representation of the environment in the real world. For example, in complex incidents such as fires in unfamiliar buildings or hazardous material spills, an RGB point cloud could provide a more intuitive spatial overview, helping dispatchers understand obstacles, exits, or hazards that may not be immediately apparent in 2D representations
Although segmented point clouds provide advanced analytical capabilities by distinguishing objects and structures within a scene, they may not always be necessary for dispatchers. Instead, an RGB point cloud, which visually replicates the real-world scene, could be sufficient for most cases. However, in large-scale or high-risk incidents, segmented point clouds could offer deeper insights, such as identifying specific hazardous areas or structural weaknesses, which may improve coordination between dispatchers and field units.
Thus, while control room personnel emphasize the importance of keeping visualizations simple, extra layers of geographical data can contribute to better decision making if integrated in a user-friendly way. A well-balanced approach that prioritizes clarity, relevance, and ease of use will ensure that dispatchers benefit from these visual tools without being overwhelmed.

Higher-Level Personnel

Higher-level personnel in the control room, such as OvDs and CaCOs, play a crucial role in overseeing emergency operations and ensuring effective coordination between field units and command structures. Unlike dispatchers, who focus on immediate incident response, these personnel engage in strategic decision making, resource allocation, and high-level communication to manage complex incidents. Their responsibilities require maintaining a comprehensive understanding of the evolving situation, ensuring that on-site teams receive timely and accurate information to perform their tasks effectively.
To achieve this, higher-level personnel use geographical data to increase SA, using 2D maps, live camera feeds, and other real-time data layers. Although these tools provide a broad overview of the incident, additional layers of information, such as (near) real-time 3D point clouds, can significantly increase their decision making capabilities. Segmented point clouds allow them to visualize the interior structure of a building, the spread of hazards, and the movement of responders, helping them anticipate challenges and optimize resource deployment. For example, by analyzing a 3D model of a burning building, an OvD can advise on the safest entry points for responders or pinpoint high-risk areas requiring immediate attention.
Interviews with control room personnel again emphasized that "less is more" when it comes to visual data, as overly complex interfaces can overload users and slow down decision making. However, when carefully integrated, extra visual information can be extremely valuable, leading to a safer operational environment. Compared to dispatchers, higher-level personnel often have more time to analyze the scans in more detail. By providing these additional data, for example, in the form of electrical devices or non-electrical devices, movable or non-movable objects, higher-level personnel can make better informed decisions on how many responders are needed, how to allocate resources efficiently, and the safest approach to entering a building. This improved SA not only improves the safety of emergency responders, but also optimizes resource management, potentially saving lives and reducing unnecessary risks.
Furthermore, segmented 3D point cloud data can serve as a valuable tool beyond real-time operations. After an incident, these data can be used for debriefing, analysis, and training purposes, allowing emergency services to review response strategies, improve future operations, and improve preparedness for similar scenarios. By incorporating these insights into training programs, first responders can better understand spatial environments and refine their tactical approaches, ultimately contributing to more effective and safer emergency responses in the future.

Local Responders

The local responders, comprising of firefighters, police, rescue personnel, and medical teams, represent the front line in emergency incidents. They are directly involved in the unfolding crisis and are tasked with executing high-stakes time-sensitive operations such as firefighting, rescuing trapped individuals, ensuring public safety, and providing medical aid.
Local responders operate under extreme pressure in environments that are dynamic, hazardous, and often difficult to navigate. Their primary needs are to obtain actionable, real-time information to support rapid decision-making and effective task execution. Key data requirements for these responders include building floor plans, evacuation routes, identification of risk zones (such as areas prone to collapse or containing toxic materials), and real-time awareness of the positions of team members. Traditional tools such as 2D maps and live camera feeds are essential for fulfilling these needs, but the introduction of (near) real-time 3D point cloud visualizations, both standard RGB and segmented, offers additional ca-

pabilities that are particularly valuable in complex or high-risk scenarios.

Standard RGB point clouds, which provide (near) real-time visualizations of the environment, are especially useful for supporting local responders. These point clouds capture spatially accurate color-coded representations of the scene, allowing responders to visualize their surroundings, locate hazards, and navigate efficiently. For example, during a fire in a multistory building, an RGB point cloud can provide responders with real-time insights into the layout, such as open passages or blocked routes, while also highlighting dynamic elements such as fire spread. The (near) real-time nature of RGB point clouds ensures that responders have access to up-to-date information, which is critical for making fast, informed decisions. Furthermore, 3D segmented point clouds, while not always essential, serve as an additional layer of analysis that can provide deeper insights into the environment. Segmentation categorizes elements within the scene, such as identifying structural features, hazards, or specific objects, which can be particularly useful in complex or large-scale incidents. For example, segmented data can pinpoint structural weaknesses, identify high-risk zones, or highlight areas where victims may be trapped. This added level of detail enables responders to prioritize their actions more effectively and address critical areas with precision.

However, the interviews suggest that standard RGB point clouds are often sufficient for the majority of emergency scenarios. Their ability to deliver (near) real-time updates and intuitive visual representations meets the immediate needs of local responders without adding unnecessary complexity. Segmented point clouds are more beneficial in situations where additional layers of information are needed, such as hazardous material incidents or scenarios involving large, unfamiliar, or partially collapsed structures. In such cases, segmented point clouds can complement RGB visualizations by providing strategic insights to responders and their command teams. Furthermore, CoPI commanders stationed outside the incident site can communicate actionable insights derived from these data sets. This approach minimizes the cognitive load on the responders, allowing them to focus on their immediate tasks while still benefitting from the improved SA provided by segmented data.
In conclusion, standard RGB point clouds are a highly effective tool for supporting SA and operational efficiency of local responders, providing real-time visual data that meet their immediate needs. Segmented point clouds, while not always required, offer a valuable supplementary layer of information for complex or high-risk scenarios. Together, these technologies improve the ability of local responders to navigate dynamic conditions, avoid hazards, and execute their missions with greater precision and safety. By incorporating RGB and segmented point clouds into emergency response workflows, responders can be equipped with versatile tools that adapt to the demands of different situations, ensuring the best possible outcomes during critical moments.

CoPI Commanders

CoPI commanders are responsible for overseeing the strategic coordination of the emergency response. The CoPI is tasked with making decisions about resource allocation, priority response, and ensuring that all emergency services operate in a coordinated manner. As described in the Kwalificatiedossier Leider CoPI (NIPV, 2009), the role of the CoPI includes leading operational efforts, managing information flows, and ensuring effective crisis communication. Commanders must operate under the guiding principles of the GRIP framework and adapt to the complex, high-pressure nature of crisis scenarios, which are characterized by urgency, uncertainty, and conflicting interests (NIPV, 2009).

Commanders in the CoPI primarily function at the "inrichten" and "richten" levels of emergency management, focusing on organizing, directing, and monitoring the execution of the crisis response. They are tasked with leading a multidisciplinary team and ensuring the proper functioning of all processes within the CoPI. This role involves assessing incoming information from multiple sources, deciding on escalation levels (e.g., GRIP), and maintaining continuous SA to optimize the response. In addition, they are responsible for ensuring the accurate dissemination of information and effective collaboration between operational units.
For CoPI commanders, 3D segmented point clouds can play an important role in making informed strategic decisions. These visualizations allow commanders to get a (near) real-time accurate representation of the building's layout, key structural elements, hazard zones, and the locations of responders. With the ability to view the environment in three dimensions, commanders can effectively assess the scope of the incident, prioritize resources, and coordinate between different emergency services.

The use of 3D segmented point clouds is particularly important in complex or large-scale incidents, where traditional 2D maps may not provide sufficient detail. These 3D visualizations offer detailed insights into areas that are difficult to navigate or assess, such as the interiors of buildings, hidden hazards, and evolving risks such as fire spread or toxic material release. By integrating 3D data, CoPI commanders can better manage the situation, allocate resources more efficiently, and ensure the safety of responders and the public.
The role of 3D segmented point clouds in improving commander SA is especially critical in scenarios that require high levels of coordination between multiple emergency services. For example, these visualizations can help commanders determine which areas are accessible for specific services, such as the fire department or medical teams, and identify potential evacuation routes for civilians. Additionally, they allow for the tracking of responder movements in real-time, helping commanders to monitor progress and adjust strategies as the situation develops.
In summary, CoPI commanders rely on 2D maps and 3D visualizations to make critical decisions about resource coordination and emergency management. The integration of tools like 3D segmented point clouds aligns with their responsibility for operational leadership and situational analysis, providing a deeper understanding of the incident's context. This allows commanders to lead more effectively, allocate resources in a timely manner, and ensure that all emergency services are working in alignment to mitigate risks and resolve the situation efficiently.

The following table summarizes the visual data needs for different groups of users during emergency response.

User Group	**Primary Needs**	**Role of RGB Point Cloud**	**Role of Segmented Point Cloud**
Intakers	Efficient intake of emergency calls; accurate information collection; rapid data entry.	Not applicable	Not applicable
Dispatchers	Clear, concise information; 2D maps with additional layers (e.g., street view, camera feeds); real-time updates.	Provides real-time visual representation of environments; helps identify hazards, obstacles, and layouts.	Improves understanding in complex incidents by providing detailed structural insights.
Higher-Level Control Room Personnel	Strategic decision-making; resource allocation; coordination across services; situational awareness using 2D maps and 3D visualizations.	Offers intuitive visualizations for real-time monitoring of responders, hazards, and key infrastructure. Supports rapid decision-making.	Provides detailed analysis of interiors and structural elements for resource optimization and safe route planning.
Local Responders	Actionable real-time information; building layouts; evacuation routes; hazard identification.	Supports navigation, hazard identification, and visualization of building layouts in real-time.	Identifies critical structural elements and hazards for precise action prioritization in high-risk scenarios.
CoPI Commanders	Strategic decision-making; resource allocation; coordination across services; situational awareness using 2D maps and 3D visualizations.	Provides intuitive, real-time visualization for situational analysis and broad coordination.	Delivers granular details for strategic coordination, including tracking responders and assessing structural risks.

Table 5.2: Overview of Different User Needs

In conclusion, when addressing the diverse needs of different user groups, it becomes clear that a one-size-fits-all visualization without flexibility would not be effective. However, instead of creating entirely separate visualizations for control room personnel, local responders, and CoPI commanders, the solution lies in a single unified visualization system with customization options. This system allows users to switch between different layers, such as RGB data and segmented parts, based on their specific tasks and needs. For example, dispatchers in the control room may rely on RGB data for an overview but can enable segmented layers for more detailed information about structural elements or hazard zones during complex incidents. Local responders, such as firefighters or police officers, could use the same system to visualize building layouts or navigate high-risk areas with segmented elements, such as doors, stairs, or hazardous zones. CoPI commanders could leverage the comprehensive visualization to access both broad overviews and granular details, ensuring that they maintain SA and strategic oversight.
This approach ensures that all users are working within a single, cohesive environment, reducing complexity while still allowing access to the precise level of information needed. By integrating dynamic toggling and multiple layers, the system prevents information overload while improving SA and decision-making across all levels of emergency response.
Based on the interviews, it became clear that objects such as doors, stairs, elevators, and windows are

critical for navigation and SA, especially in dynamic and hazardous conditions. In addition, structural elements such as walls, floors, ceilings, and columns were identified as essential to assess the stability of the building and identify safe access routes. These objects form the foundation for creating meaningful and actionable visualizations that cater to the specific needs of all emergency responders.
Furthermore, while different emergency services may benefit from additional objects tailored to their operations (e.g. fire hydrants for firefighters or barriers for police), this research focuses on structural elements and navigational objects that are universally relevant. These elements are essential to ensure safe and efficient operations for all stakeholders involved in emergency response scenarios.
This unified and flexible visualization system ensures that all user groups, from dispatchers to responders and commanders, can perform their roles effectively, contributing to a more synchronized and efficient emergency management system. In the end, providing additional information improves decision making, ultimately creating a safer environment for first responders.

5.2.2 Sub-question 5: Effective Visualization

After identifying the specific information needs of the various emergency response personnel, this subsection presents the results of applying scientific principles and cartographic methods to effectively communicate the segmented 3D point cloud data to emergency response stakeholders.

Map Use Cube

First the Map Use Cube theory is applied to position the different user groups within the three dimensions, user type, purpose of use, and level of interaction, with a focus on how segmented point cloud data meets their needs.

Control Room Personnel

Control room personnel, including dispatchers and calamity coordinators, serve as a private audience within the map use cube. Their primary role is to synthesize large volumes of dynamic information to allocate resources, monitor incidents, and maintain communication with field responders. Unlike public maps designed for widespread comprehension, their visualizations are made specifically for operational workflows and require detailed, context-specific data.

- **Position within the Map Use Cube:** Lower-level Control room personnel do not interact with segmented 3D point clouds or use it for analysis with the aim of identifying patterns, assessing structural risks and supporting decision-making. However, the higher-level control room personnel uses the segmented 3D point clouds for deeper analysis and scenario planning. Their level of interaction is moderate, involving toggling between data layers (e.g., hazard zones, structural elements) to create a comprehensive operational overview.
- **Challenges for Cartographic Design:** The complexity of segmented data must be visualized in a way that minimizes cognitive load. For example, control room personnel may need to identify compromised areas within a building and relay this information to field responders. The visualization system should allow them to toggle relevant layers, such as structural segmentation or hazard zones, while maintaining clarity and avoiding data overload.

Local Responders

Local responders, such as firefighters, police officers, and medical teams, are positioned as a private audience that uses real-time data in unfamiliar high-stakes environments. They rely on the same 3D point cloud system, but require a simplified view for quick SA and navigation.

- **Position within the Map Use Cube:** Local responders primarily use the visualization for exploration, engaging with visualizations to understand spatial layouts, identify hazards and navigate effectively. Their level of interaction is low, as time-sensitive tasks limit their ability to manipulate complex interfaces.
- **Challenges for Cartographic Design:** The system must facilitate exploration with minimal interaction. For example, a firefighter entering a smoke-filled building must quickly interpret data to locate safe routes, blocked passages, or areas that require rescue efforts. Simplified, intuitive layers, such as RGB maps with hazard markers, can address this need. While the same system is used, it should default to a streamlined view designed for rapid interpretation.

CoPI Commanders

CoPI commanders represent a private audience with a strategic role in overseeing resource allocation and multi-agency coordination. They require access to the full spectrum of data layers for synthesizing information and modeling outcomes.

- **Position within the Map Use Cube:** CoPI commanders use the visualization system for synthesis, integrating multiple layers of information to gain a strategic overview. Their level of interaction is moderate to high, dynamically toggling data layers (e.g. structural risks, responder locations) and analyzing evolving risks.
- **Challenges for Cartographic Design:** The system must enable seamless switching between broad overviews and detailed views. For example, commanders who coordinate a high-rise fire can analyze floor-level segmentation while simultaneously monitoring the movement of responder teams. The same visualization should support these tasks through dynamic layer toggling and customizable presets specified to their strategic needs.

Cognitive Load

After defining the position of the different users in the Map Use Cube, the next step is to keep the cognitive load of the visualization low. Cognitive load refers to the mental effort required to process and interpret information, and understanding its implications is crucial to ensure that visualizations are functional and user-friendly. In this research, cognitive load theory is applied to address the challenge of presenting complex spatial data to diverse user groups.

One of the key challenges in visualizing segmented 3D point cloud data lies in managing the intrinsic cognitive load. This type of load is inherent to the complexity of the information itself, such as understanding the spatial relationships in a multi-story building or interpreting structural segmentation. For example, recognizing doors, walls, stairs, and hazard zones requires users to synthesize detailed spatial information. To address this, visualizations have to be designed to present only the most essential structural elements and navigational aids while allowing users to access additional layers of detail as needed. This approach, known as progressive disclosure, ensures that users can begin with a high-level overview (such as RGB point clouds for navigation) and gradually delve into more detailed segmentation layers when required. By structuring the data in this way, the complexity of the information becomes more manageable, allowing users to focus on their immediate tasks.
The second load, the extraneous cognitive load, which refers to unnecessary mental effort caused by poorly designed visual elements, is another critical consideration. Elements such as excessive color use, cluttered legends, or overly complex interfaces can distract users and make it harder to interpret key information. To mitigate this, visualizations have to be designed with simplicity and clarity in mind. For example, redundant elements such as clutter have to be removed and intuitive symbology has to be used to represent critical features such as hazard zones and navigable pathways. The use of dynamic layering can further reduce extraneous load by allowing users to toggle between layers, such as structural segmentation and responder locations, without being overwhelmed by irrelevant details. Furthermore, legends have to be optimized to dynamically display only the symbols and colors relevant to the active layers, ensuring that users could quickly interpret the visualization.
At last, the visualizations should be designed to enhance the germane cognitive load, ensuring that mental effort is directed toward understanding and applying new information. In the context of segmented 3D point clouds, this involves designing visualizations that actively support users in synthesizing and interpreting data. Techniques such as visual emphasis have to be used to highlight critical features, such as structural risks or evacuation routes, using color gradients, contrast, and animation.

New Map Communication Model

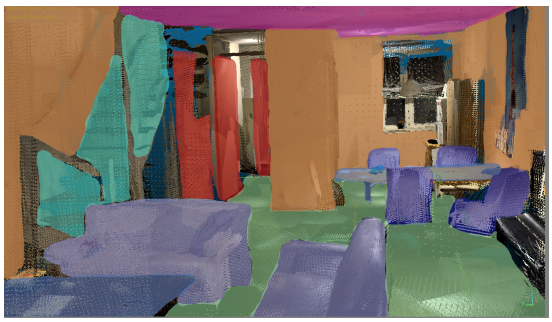
The last step is to assess the dynamic interaction between the map maker and the map user, especially in how map users provide feedback on the proof of concept and use the different visualizations to make more informed decisions.
In order to start the communication between the map maker and the map user, a user interaction test or a proof of concept has to be made. Hereby, map users and the map maker explore the segmented data. It is important to implement feedback mechanisms during these tests to capture user input on usability. During this interaction session, it is important that the map maker records the feedback, focusing on how users interpret the segmentation result, identify potential classifications, and even suggests refinements. After these feedback sessions, it is important to analyze the feedback for patterns that reflect the strengths or limitations of the model in delivering accurate information. By identifying these strengths and limitations, the visualization can be adjusted.

To conclude, the combination of the map use cube, cognitive load theory new map communication model, demonstrates that applying cognitive load theory enables the creation of visualizations that balance complexity and usability. By managing intrinsic load, minimizing extraneous load, and enhancing germane load, the visualizations support users in quickly interpreting and acting upon segmented 3D point cloud data. This approach ensures that all user groups,control room personnel, local responders, and CoPI commanders can effectively perform their roles, ultimately contributing to better decision making and coordination during emergency responses.

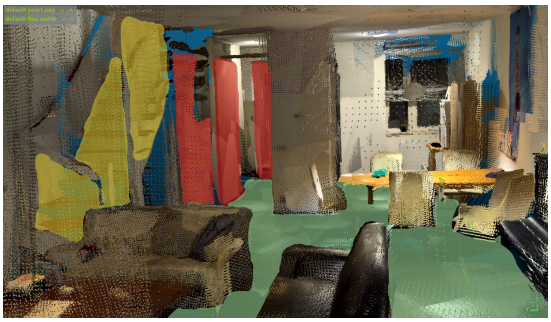
5.2.3 Sub-question 6: Proof of Concept

After defining the different user needs and applying scientific principles and cartographic methods to effectively communicate a segmented 3D point cloud to a diverse user group, it is time to develop different proof of concepts with the needs of the different users, scientific principles, and cartographic methods in mind.

The initial artist impressions created in Krita (as can be seen in figure 5.2) provided a structured approach to visualizing the segmented point cloud data. By overlaying distinct categories that could improve SA, such as walls, ceilings, floors, and furniture, onto a reference point cloud image, a clear and intuitive representation of the spatial environment was achieved.
During the testing phase, different users highlighted that color-coded segmentation was beneficial to distinguish between different object categories. Simpler representations were found to be easier to interpret, potentially leading to improved SA and decision making. Additionally, maintaining color consistency across different visualizations was crucial for user understanding.
This feedback played a key role in refining the visualizations created in Krita, which were later applied to the proof-of-concept for the point clouds of both the authors' house and the CGI office.



(a) Visualization of author's house in Krita.



(b) Visualization of author's house in Krita.

Figure 5.2: Comparison of two visualizations of the author's house created in Krita.

Due to the limitations of the PointTransformer V3+PTT model in processing self-gathered datasets, manual segmentation is performed in CloudCompare. This process classified the point cloud into seven primary categories: walls, ceilings, floors, doors, stairs, windows, and objects. These categories provide a structured framework for the basic visualization, which served as the basis for assessing the impact of color choices and visualization styles on user interpretation. For the obstacle differentiation and detailed object classification, further distinctions are made within the object category.
Following segmentation, three different visualization strategies are applied to assess their effectiveness in conveying spatial information. Each strategy uses a different color scheme, functional, contrast, and realistic, to determine how different color choices influence user perception and SA. First, the RGB point cloud of the ground floor of the author's home, shown in Figure 5.3, served as the first reference for these visualizations. Subsequently, the RGB point cloud of the CGI office shown in Figure 7.3, served as the second reference for these visualizations.

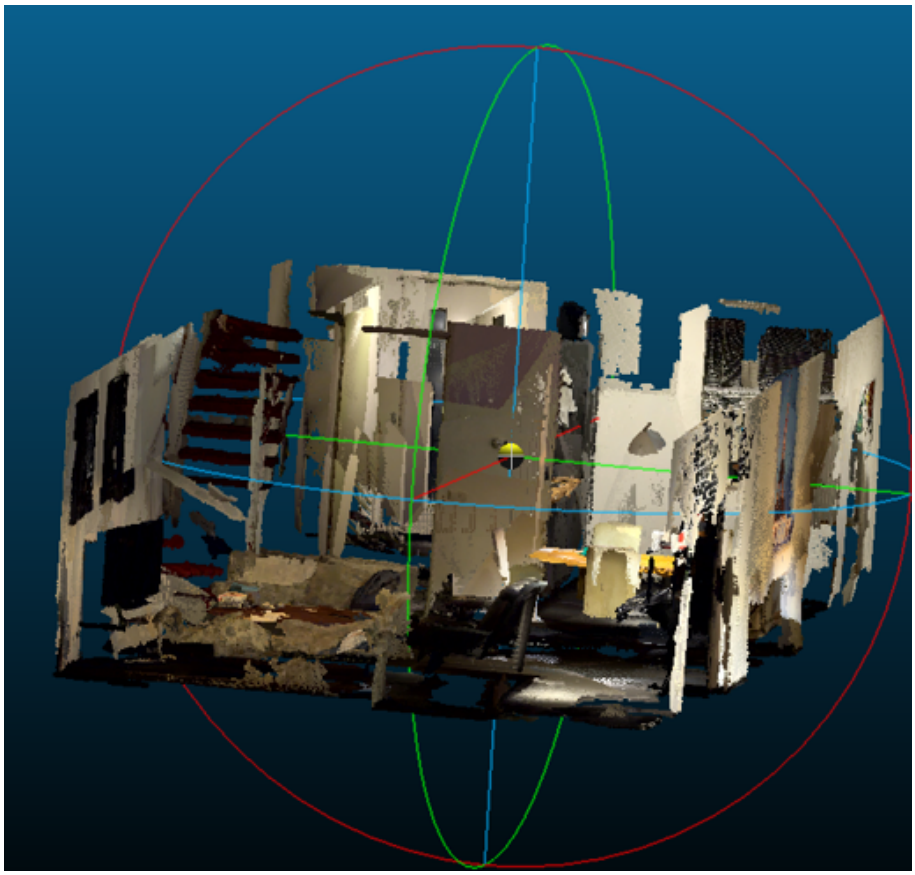


Figure 5.3: RGB point cloud of the ground floor of the author's home

Visualizations of the Author's Home

Basic Visualization

The basic visualization aims to provide a foundational understanding of the spatial environment, allowing users to interpret the general structure of the segmented point cloud. Hence, focusing on simplicity to keep the cognitive load at a minimum. To evaluate its effectiveness, three different color schemes are applied.

Functional Color Scheme

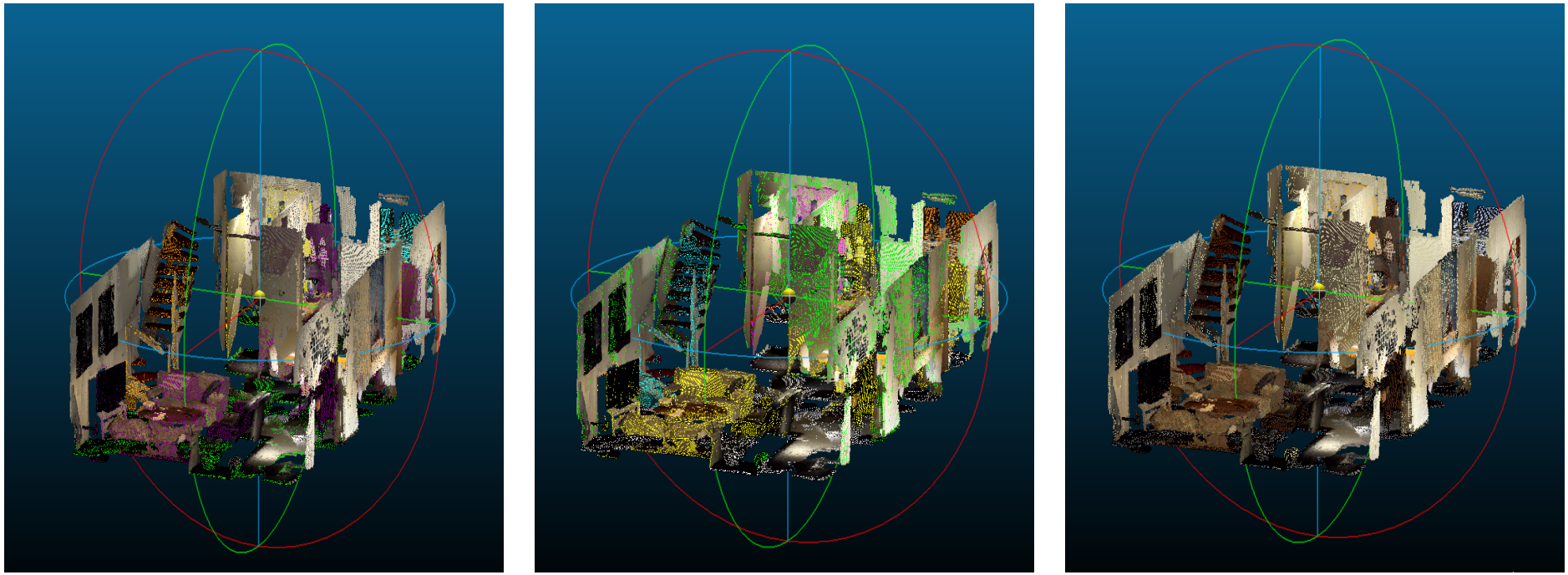
First, the functional color scheme is designed to improve clarity and usability by assigning colors based on their practical significance in an emergency response scenario. This scheme ensures that users could quickly identify key structural elements and potential hazards within the environment.
In this visualization, floors, doors, and stairs are highlighted in green, yellow, and orange, representing safe pathways and access points. The walls and obstacles are colored dark gray and purple, indicating barriers that could impede movement. This approach is particularly beneficial for first responders, such as firefighters and police officers, who need to rapidly assess their surroundings in high-stress situations to gain a clear understanding of the building.

Contrast Based Color Scheme

The contrast-based color scheme is developed to maximize differentiation between objects by using highly saturated and opposing colors. The goal of this scheme is to ensure that each category is immediately distinguishable.
In this visualization, pink is used for doors, orange for windows, green for walls, cyan for stairs, and yellow for obstacles, to ensure that all elements stand out distinctly from each other. This scheme is optimized for scenarios where the rapid identification of different objects is essential.

Realistic Color Scheme

The realistic color scheme aims to create a natural and intuitive representation of the environment by mirroring real-world object colors. This approach is particularly valuable for providing a clearer yet familiar alternative to the RGB view, leading to faster recognition.
In this scheme, doors and wooden furniture were displayed in various shades of brown, metallic elements such as appliances and lifts appeared in gray, and screens and electronic devices were shown in black.

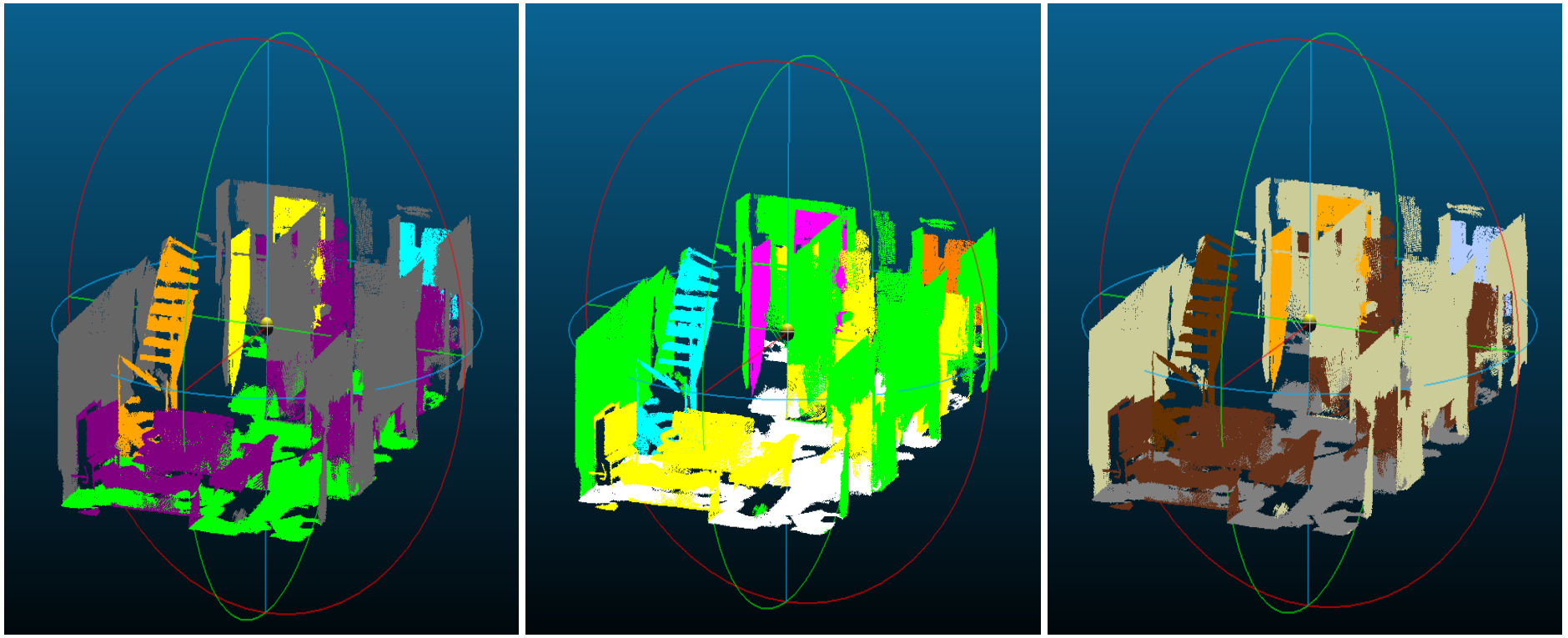


(a) Basic visualization in the functional color scheme (RGB + Segment).

(b) Basic visualization in the contrast color scheme (RGB + Segment).

(c) Basic visualization in the realistic color scheme (RGB + Segment).

Figure 5.4: Comparison of different visualization color schemes of the author's house (RGB + Segment).



(a) Basic visualization in the functional color scheme (Segment).

(b) Basic visualization in the contrast color scheme (Segment).

(c) Basic visualization in the realistic color scheme (Segment).

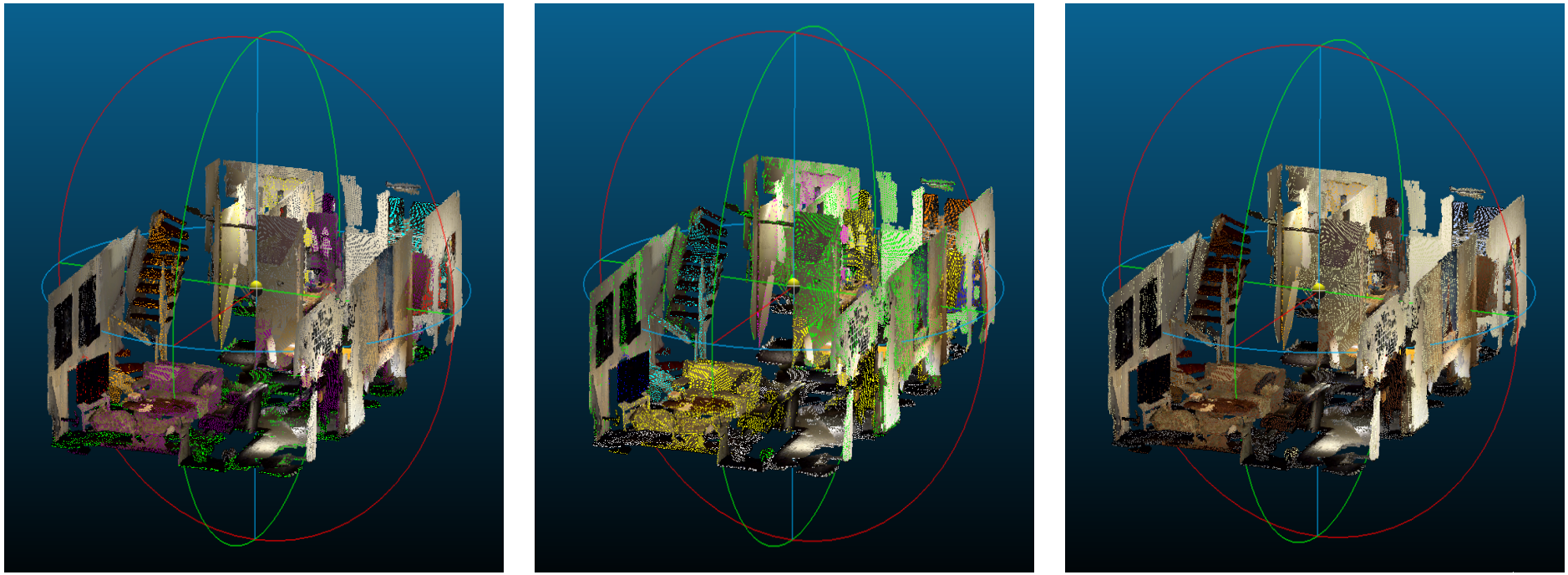
Figure 5.5: Comparison of different visualization color schemes of the author's house.

Obstacle Differentiation
To address the limitations of basic visualizations, the obstacle differentiation level introduced a further classification by distinguishing between general obstacles and electrical devices. This refinement aims to provide emergency responders with a clearer understanding of potential hazards within an indoor environment. The distinction between these two categories ensures that physical obstructions and electrical hazards could be quickly identified and assessed.

Functional Color Scheme
The functional color scheme now highlights electrical devices in bright red, immediately showcasing potential hazards, while all other color classifications remain unchanged. This ensures consistency in the different visualizations while improving critical risks for quick identification by first responders.

Contrast Based Color Scheme
The contrast based color scheme again emphasizes maximum visual distinction between objects by applying highly saturated and opposing colors. In this scheme, obstacles were assigned bright yellow, while electrical devices were displayed in dark blue, to ensure that both categories stood out clearly against the environment.

Realistic Color Scheme
The realistic color scheme is still designed to maintain the appearances of natural objects but now also provides differentiation between obstacles and electrical devices. In this scheme, obstacles are displayed in their real-world colors, such as dark brown for furniture, and electrical devices are assigned gray tones, reflecting the common appearance of screens, appliances, and control panels. However, the realistic color scheme makes it difficult to quickly distinguish between obstacles and electrical devices, as their natural colors often blend together. This can reduce the effectiveness of rapid hazard identification, potentially slowing decision making in critical situations.

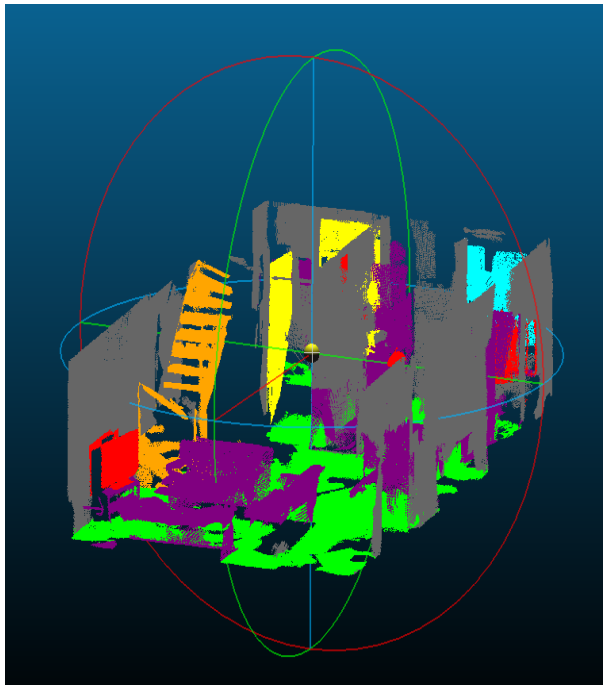


(a) Obstacle differentiation visualization in the functional color scheme (RGB + Segment).

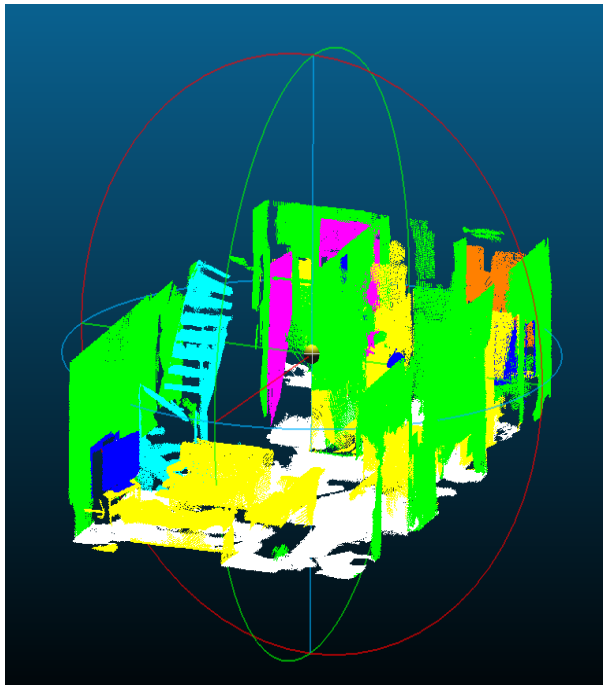
(b) Obstacle differentiation visualization in the contrast color scheme (RGB + Segment).

(c) Obstacle differentiation visualization in the realistic color scheme (RGB + Segment).

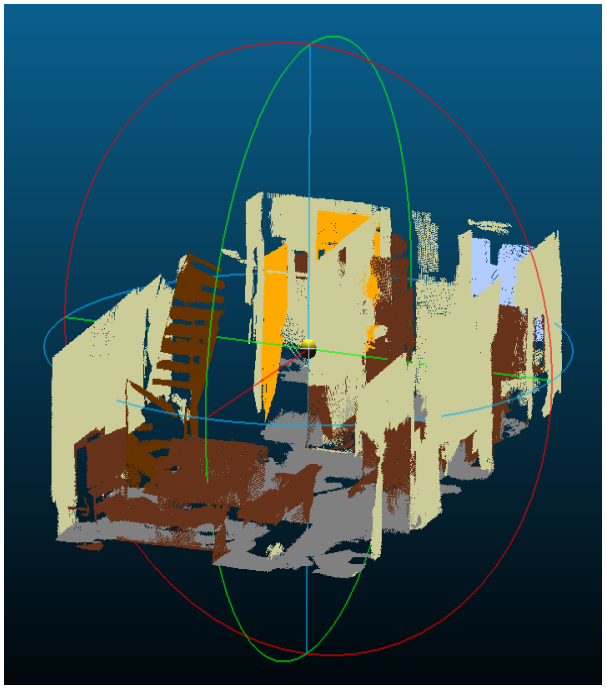
Figure 5.6: Comparison of obstacle differentiation visualization color schemes of the author's house (RGB + Segment).



(a) Obstacle differentiation visualization in the functional color scheme (Segment).



(b) Obstacle differentiation visualization in the contrast color scheme (Segment).



(c) Obstacle differentiation visualization in the realistic color scheme (Segment).

Figure 5.7: Comparison of obstacle differentiation visualization color schemes of the author's house (Segment).

Detailed Object Classification

The detailed object classification level introduces a more granular breakdown of the segmented objects, distinguishing individual furniture items such as beds, chairs, tables, and storage units, and further refining electrical devices into subcategories like screens, lighting fixtures, and appliances. This approach aims to provide a highly detailed visualization that enables users to recognize specific objects within the environment.

Although this level of detail provides a rich representation of indoor spaces, it also introduces significant complexity, making it more challenging to extract critical information quickly. The impact of this complexity is again analyzed across the three different color schemes: functional, contrast-Based, and Realistic.

Functional Color Scheme

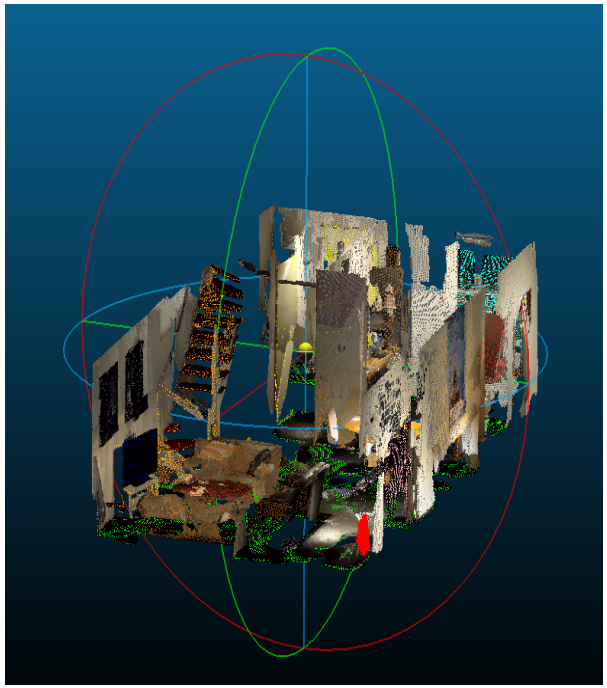
In the functional color scheme, further differentiation between objects does not improve the clarity. Instead, this scheme prioritizes safe pathways, access points, and potential hazards. By adding more object categories, the visualization can become too overwhelming, making it harder to quickly assess these critical elements, potentially reducing its effectiveness in emergency situations.

Contrast Based Color Scheme

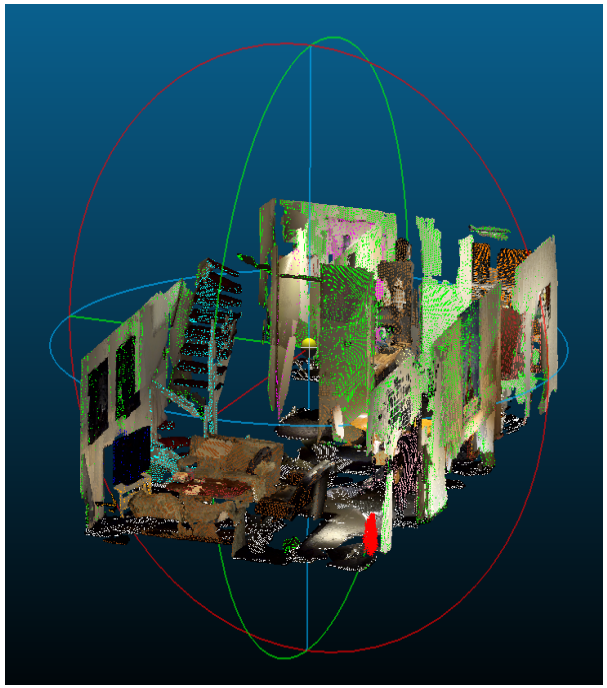
The contrast based color scheme again applies highly saturated and distinct colors to each object class to ensure maximum differentiation. However, similar to the functional color scheme, excessive differentiation can make it difficult to quickly assess the situation. Although distinct colors improve object separation, they can also create visual clutter, making it more challenging to focus on key elements.

Realistic Color Scheme

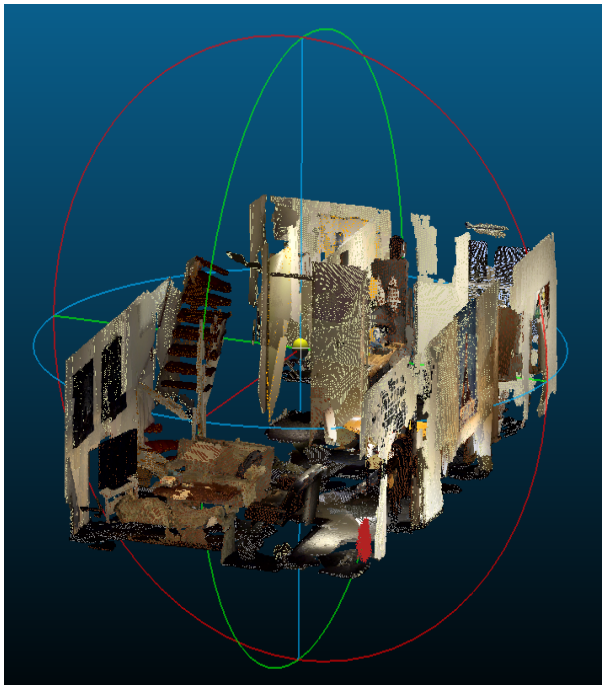
The realistic color scheme tried to mirror the real world appearances as closely as possible. The furniture remained brown, the metal appliances grayed and the screens appeared black. This approach is designed to provide an intuitive representation of the environment, making it easier to recognize objects without relying on artificial color coding. In contrast to the functional and contrast-based visualizations, the additional differentiation in the realistic color scheme could enhance SA by providing more information about the environment, allowing users to interpret the scene more naturally.



(a) Detailed object classification visualization in the functional color scheme (RGB + Segment).

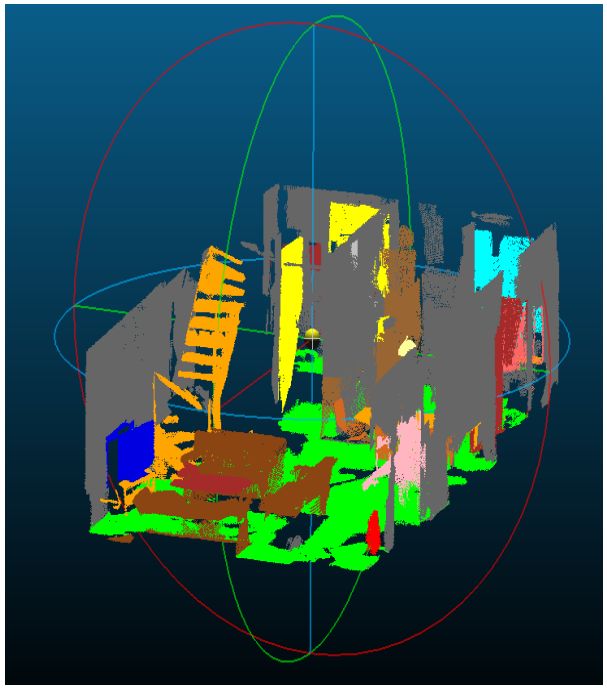


(b) Detailed object classification visualization in the contrast color scheme (RGB + Segment).

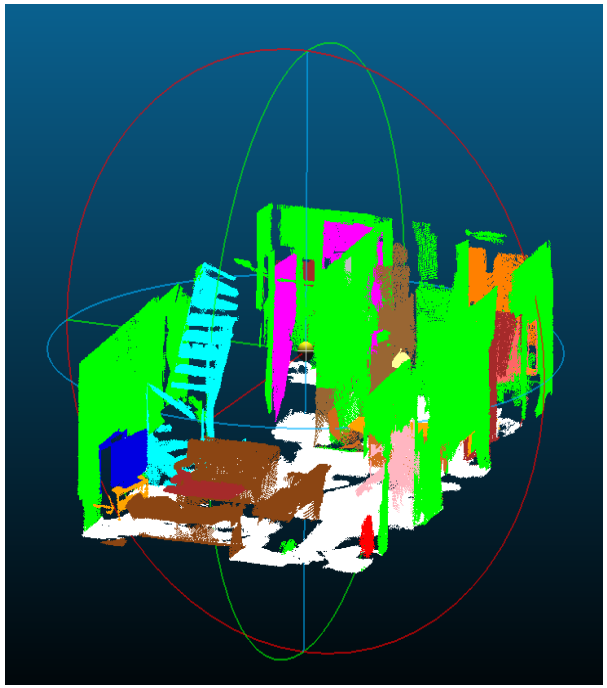


(c) Detailed object classification visualization in the realistic color scheme (RGB + Segment).

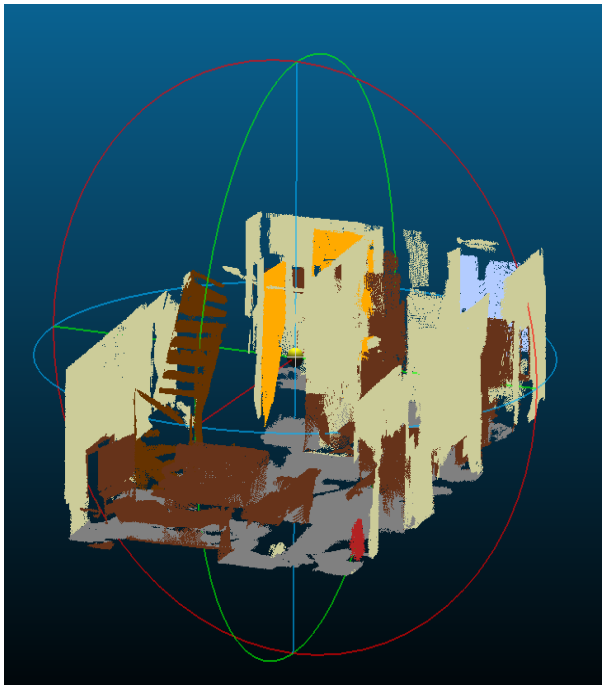
Figure 5.8: Comparison of detailed object classification visualization color schemes of the author's house (RGB + Segment).



(a) Detailed object classification visualization in the functional color scheme (Segment).



(b) Detailed object classification visualization in the contrast color scheme (Segment).



(c) Detailed object classification visualization in the realistic color scheme (Segment).

Figure 5.9: Comparison of detailed object classification visualization color schemes of the author's house (Segment).

Visualizations of the CGI Office

After visualizing the ground floor of the author's home, the next step is to create visualizations in a different environment, an office building. Unlike a residential space, an office presents a more structured layout with standardized furniture arrangements, workstations, and open areas, making it an ideal setting for evaluating how segmentation and color schemes perform in a professional workspace.

The CGI Office scan is used as the basis for these visualizations, allowing for a direct comparison between home and office environments. By applying the previously tested segmentation levels and color schemes to this new setting, the goal was to determine whether the same visual approaches remain effective or if adjustments are needed to account for the differences in spatial complexity and object categorization within an office.

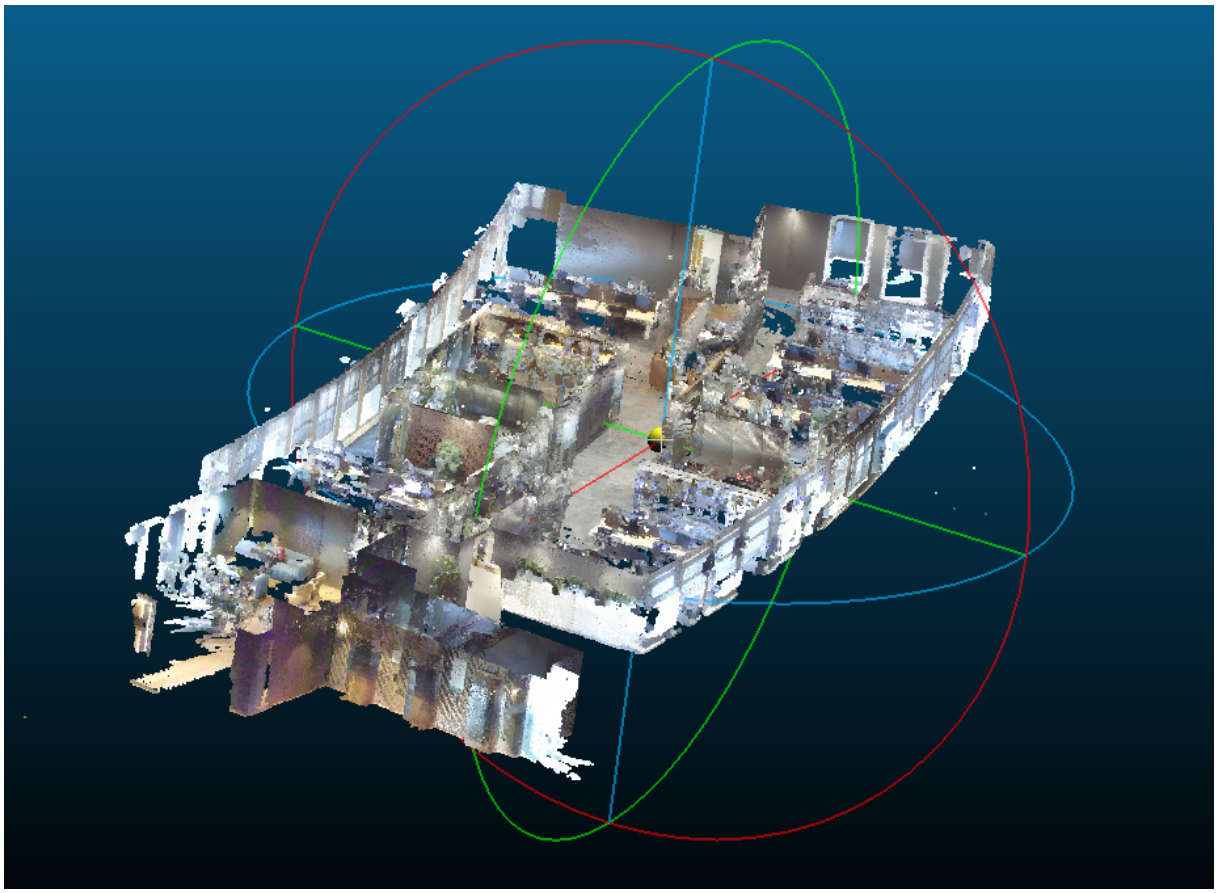


Figure 5.10: RGB point cloud of the CGI office.

Basic Visualization

Again, basic visualization aims to provide a basic understanding of the office environment, ensuring that structural components such as walls, floors, and doors are clearly distinguishable. To evaluate the effectiveness of this segmentation, three different color schemes are applied (the same as for the author's home).

Unlike the home environment, which contains diverse furniture arrangements and varied object placement, the office scan presents a more structured and repetitive layout, consisting primarily of workstations, partitions, and meeting rooms. This uniformity influences the effectiveness of segmentation, as it becomes more critical to differentiate between furniture and structural components due to the high density of objects.

Functional Color Scheme

The functional color scheme assigns colors based on the practical usability of different elements in the office space. The floors and doors are colored green and yellow, indicating safe access routes and access points. The walls and partitions are neutral gray, representing the structural barriers that defined the layout of the office. The furniture and workstations are assigned purple tones, making them clearly visible without overwhelming the visualization. The office scan also includes a new category, people, and they are marked in red to highlight their significance within the environment.

Contrast Based Color Scheme

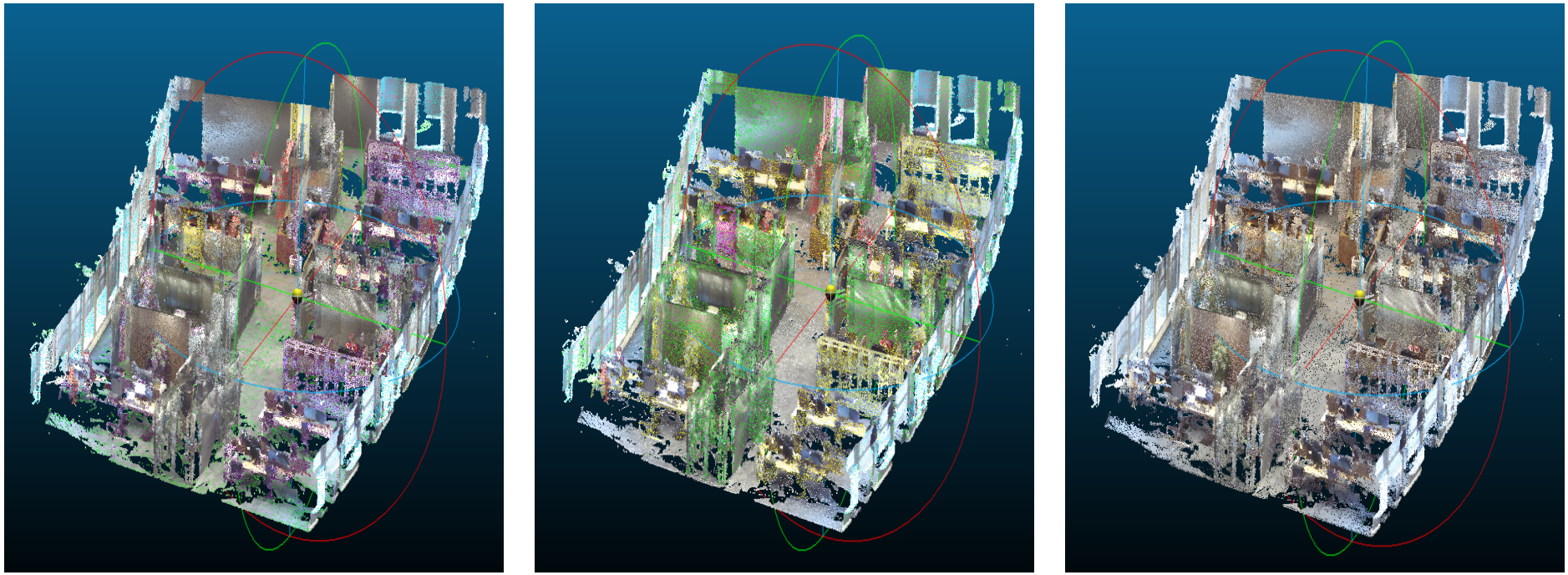
The contrast based color scheme is designed to maximize differentiation between objects by using highly saturated and opposing colors. This approach ensures that each category is clearly distinguishable.

In this visualization, the floors were assigned white, providing a neutral background that enhanced the contrast with other elements. The furniture and workstations are colored yellow, which makes them easily distinguishable from the structural components. The windows are displayed in bright blue, ensuring they stood out as transparent or open areas in the visualization. Doors were highlighted in magenta, differentiating them from walls and partitions while marking access points. People are represented in red, reinforcing their importance within the environment and ensuring immediate visibility.

Realistic Color Scheme

The realistic color scheme aims to mirror the appearances of objects in the real world, providing an intuitive and natural representation of the office environment. This approach was particularly useful for training simulations, architectural visualization, and emergency response planning, where familiarity with real-world colors helped users quickly recognize objects.

In this visualization, the floors retain their natural gray or wood textures, while maintaining a familiar appearance. Furniture, such as desks and chairs, is displayed in brown , reflecting common office materials. The windows are shown in light blue, mimicking real-world glass reflections. People are colored in realistic skin tones, making them harder to see compared to the red in the color schemes.

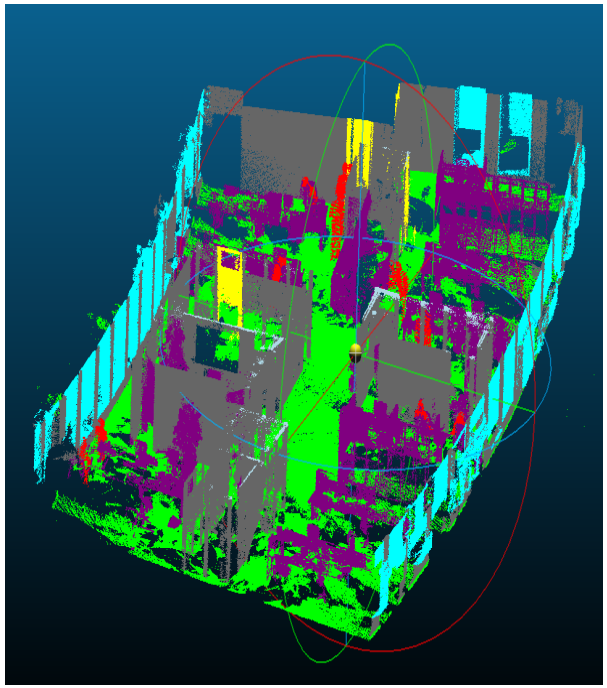


(a) Basic visualization of the CGI office in the functional color scheme (RGB + Segment).

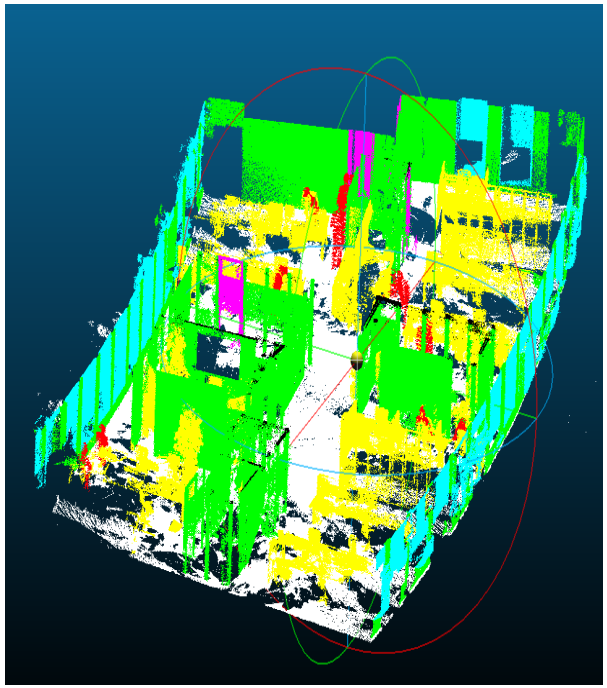
(b) Basic visualization of the CGI office in the contrast color scheme (RGB + Segment).

(c) Basic visualization of the CGI office in the realistic color scheme (RGB + Segment).

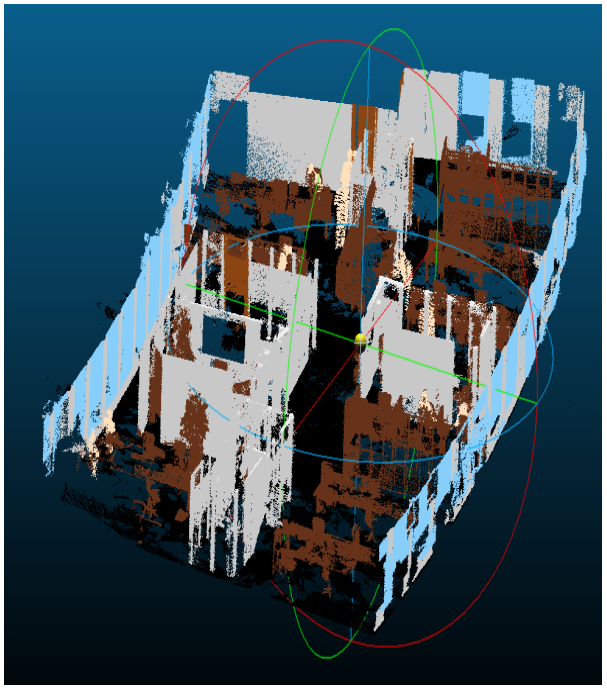
Figure 5.11: Comparison of basic visualization color schemes of the CGI office (RGB + Segment).



(a) Basic visualization of the CGI office in the functional color scheme (Segment).



(b) Basic visualization of the CGI office in the contrast color scheme (Segment).



(c) Obstacle differentiation visualization of the CGI office in the realistic color scheme (Segment).

Figure 5.12: Comparison of basic visualization color schemes of the CGI office (Segment).

Obstacle Differentiation

Again, to further expand the basic visualization, the obstacle differentiation level introduced a further classification by distinguishing between general obstacles and electrical devices. This refinement aims to provide emergency responders with a clearer understanding of potential hazards, ensuring that physical obstructions and electrical risks could be quickly identified and assessed.

Functional Color Scheme

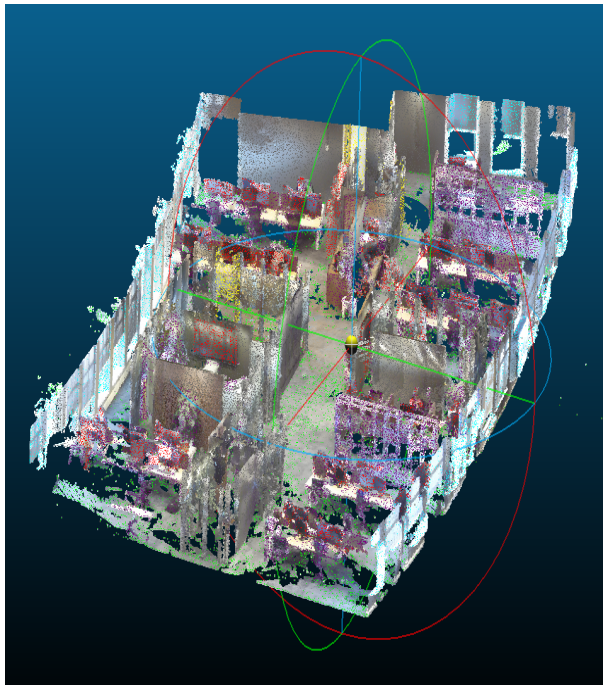
The functional color scheme assigned colors based on the practical usability of objects, ensuring that emergency responders could quickly recognize obstacles and electrical hazards. In this visualization, general obstacles such as desks, chairs, and partitions are displayed in purple, representing potential barriers in the office space. In addition, for people, electrical devices are also highlighted in red, drawing attention to possible fire hazards or power-related risks. The floors and movement pathways are colored green and yellow, maintaining visibility for safe navigation.

Contrast Based Color Scheme

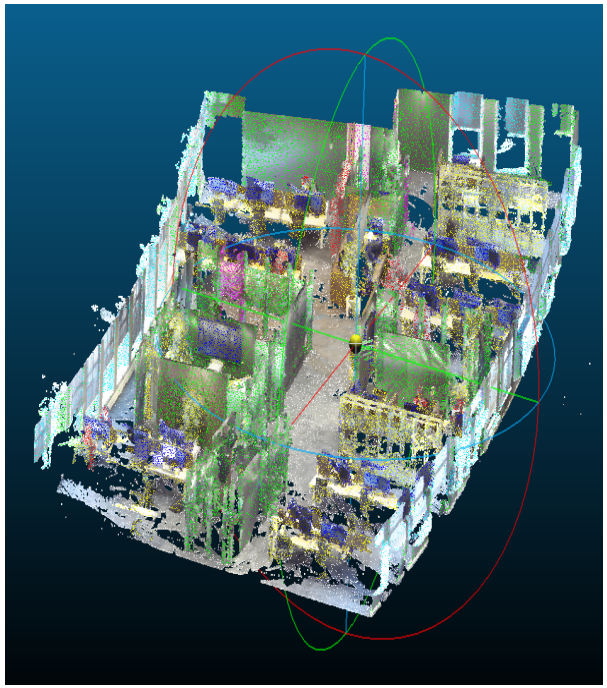
The contrast based color scheme maximizes visual differentiation between obstacles and electrical devices by using highly distinct colors. In this visualization, general obstacles are colored bright yellow, so that furniture is immediately visible. Electrical devices were highlighted in dark blue, marking potential hazards in the environment. The floors remain white, providing a neutral background that improves contrast. Doors are pink and windows are cyan, maintaining a clear separation between access points and structural components.

Realistic Color Scheme

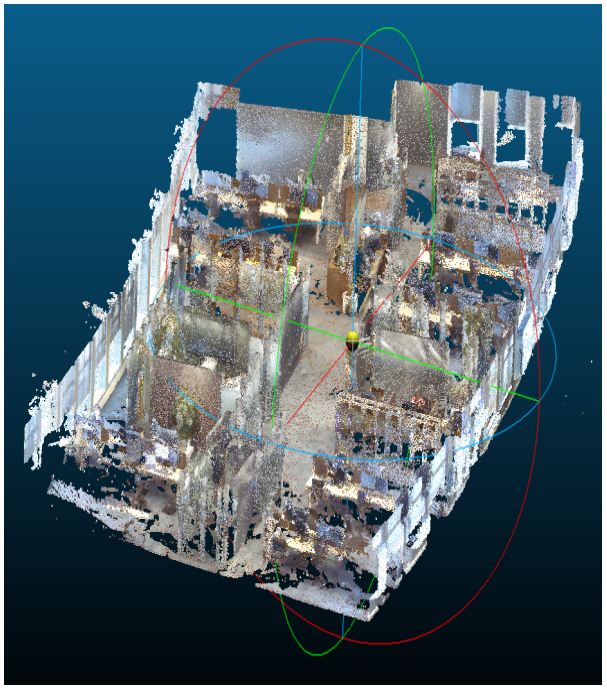
The realistic color scheme aims to maintain the appearances of natural objects while ensuring that electrical devices remain distinguishable from other objects. In this visualization, desks, chairs, and partitions are displayed in their natural brown or gray tones, reflecting standard office furniture materials. Electrical devices are assigned a grayish color, which helps differentiate them while still maintaining a realistic aesthetic. The floors retained their original textures, such as carpet or polished surfaces, while the windows remained light blue and the doors matched the variations in real-world color.



(a) Obstacle differentiation visualization of the CGI office in the functional color scheme (RGB + Segment).

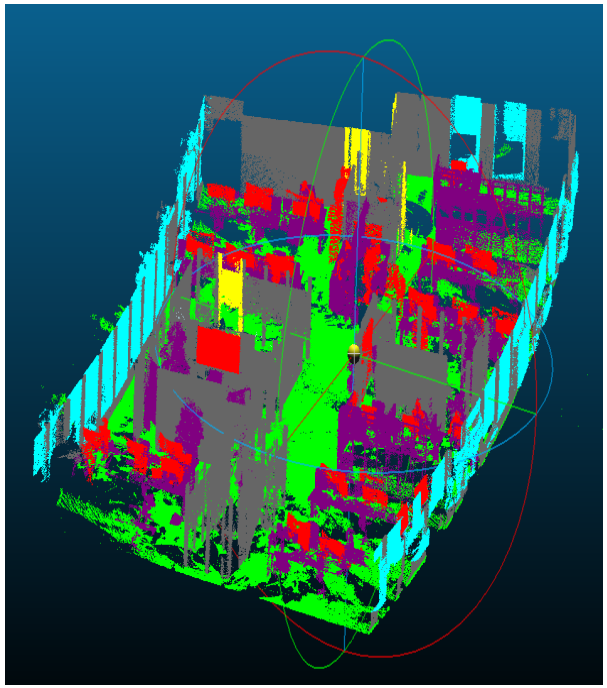


(b) Obstacle differentiation visualization of the CGI office in the contrast color scheme (RGB + Segment).

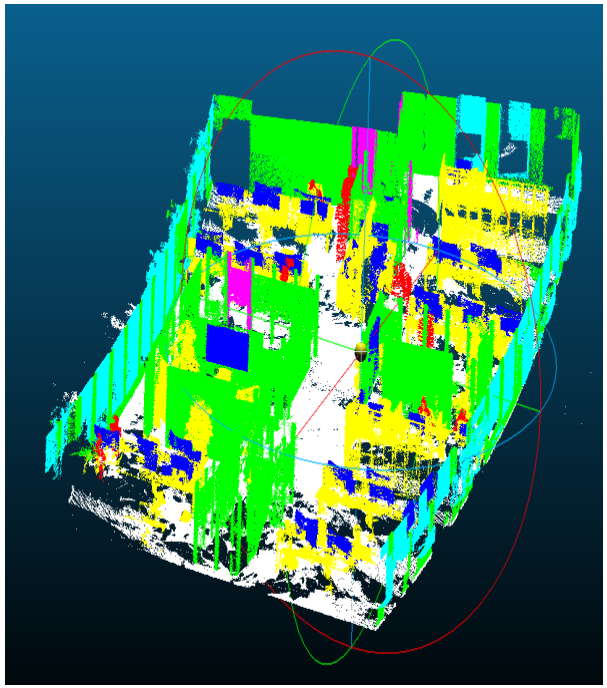


(c) Obstacle differentiation visualization of the CGI office in the realistic color scheme (RGB + Segment).

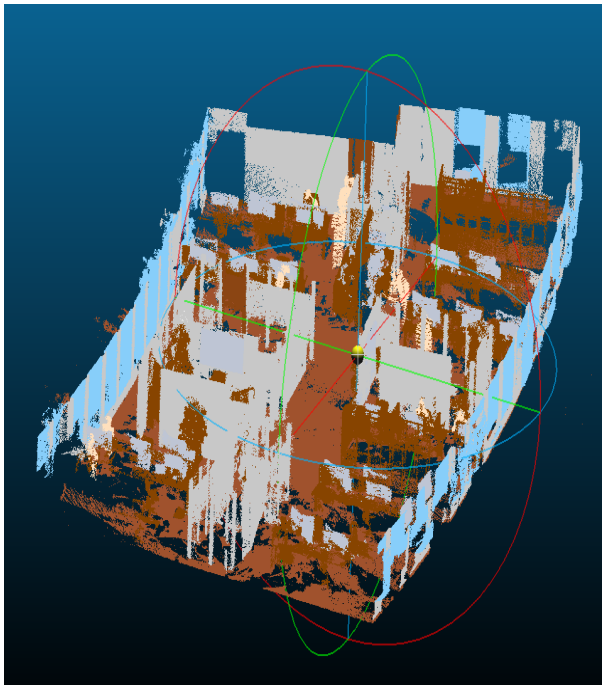
Figure 5.13: Comparison of obstacle differentiation visualization color schemes of the CGI office (RGB + Segment).



(a) Obstacle differentiation visualization of the CGI office in the functional color scheme (Segment).



(b) Obstacle differentiation visualization of the CGI office in the contrast color scheme (Segment).



(c) Obstacle differentiation visualization of the CGI office in the realistic color scheme (Segment).

Figure 5.14: Comparison of obstacle differentiation visualization color schemes of the CGI office (Segment).

Detailed Object Classification

The detailed object classification level introduces a more granular breakdown of the segmented objects, distinguishing individual furniture items such as desks, chairs, tables, and partitions, and refining electrical devices into subcategories like monitors, printers, and lighting fixtures.
Although this level of detail provides a highly descriptive representation of the office environment, it also introduces significant complexity, making it more challenging to extract critical information quickly.

Functional Color Scheme

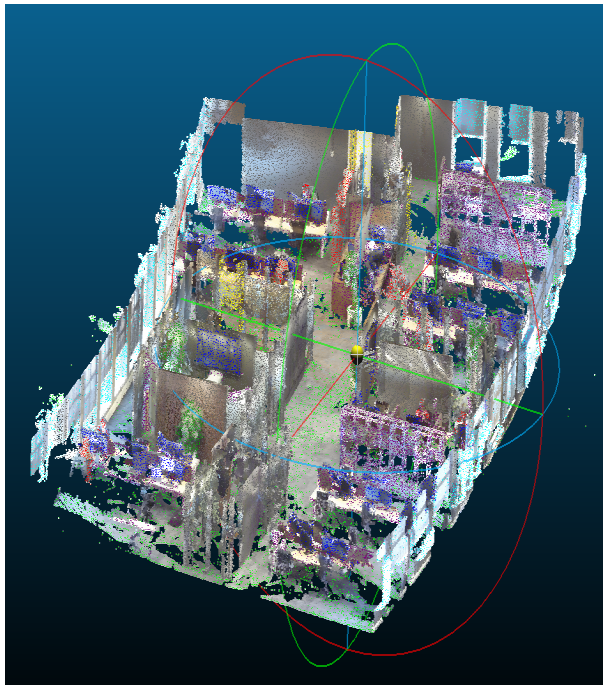
In the visualization of the functional color scheme, the workstations, desks, and chairs are colored purple, allowing them to be visually distinguishable without overwhelming the scene. Partitions and walls are colored dark gray. Electrical devices, such as monitors and printers, are marked in dark blue, ensuring that they are positioned as potential hazards. The floors and doors are kept green and yellow, maintaining a clear visual pathway for navigation. Lastly, the people remain red, highlighting their presence as both dynamic elements and potential obstacles within the environment.

Contrast Based Color Scheme

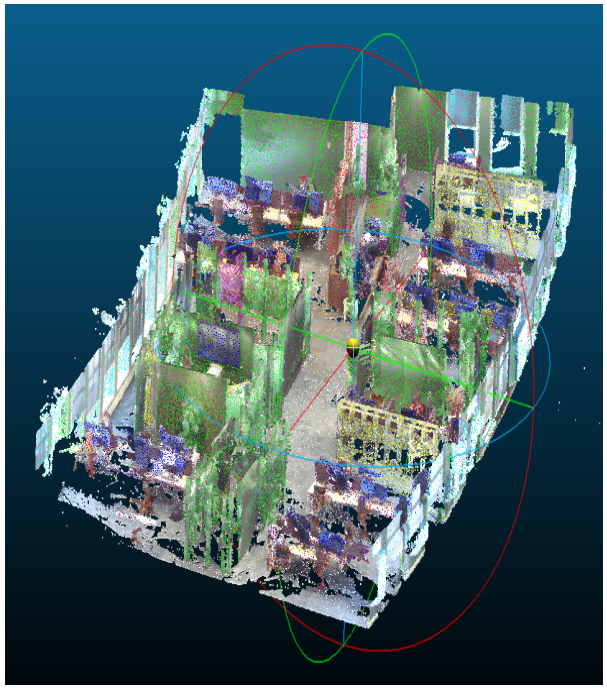
The contrast based color scheme used highly saturated colors to differentiate every object class, ensuring maximum object separation. In this visualization, the desks are colored purple, the chairs orange, and the partitions green, allowing each category to be easily distinguishable. Electrical devices, such as monitors and printers, are colored blue, ensuring that potential hazards remained visible. The floors are kept white, providing a neutral background that improves the contrast with the various categories of objects.

Realistic Color Scheme

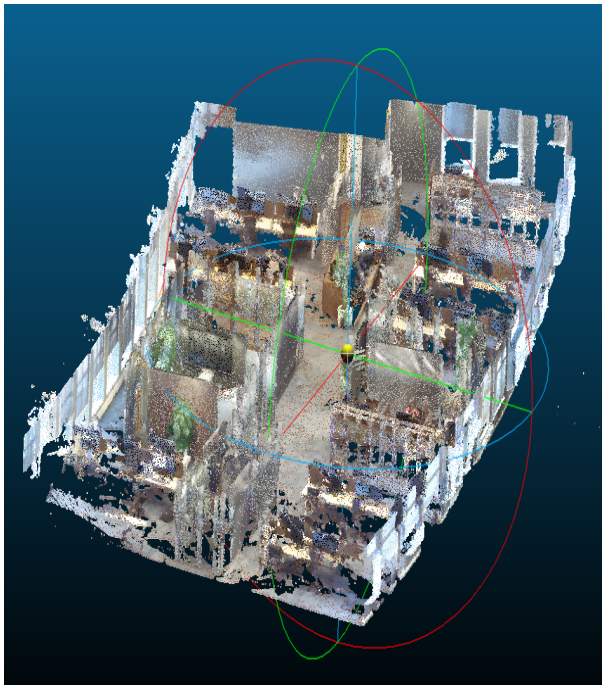
The realistic color scheme aims to preserve object familiarity by mirroring real-world appearances, making the environment more intuitive to navigate. In this visualization, wooden desks and chairs retained brown hues, metallic elements such as partitions and storage cabinets are displayed in gray, and electrical devices, including monitors and appliances, are black. The floors retained their natural textures, such as carpet or polished surfaces, and the windows remained light blue, mimicking the reflections of real-world glass.



(a) Detailed object classification visualization of the CGI office in the functional color scheme (RGB + Segment).

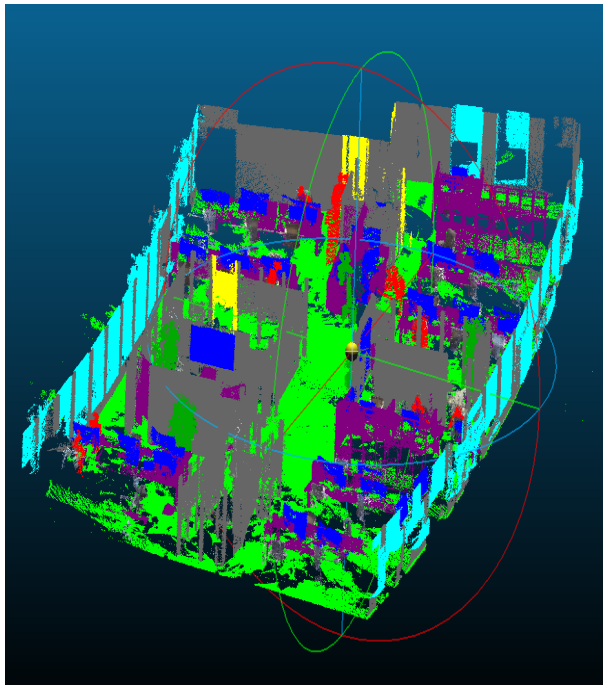


(b) Detailed object classification visualization of the CGI office in the contrast color scheme (RGB + Segment).

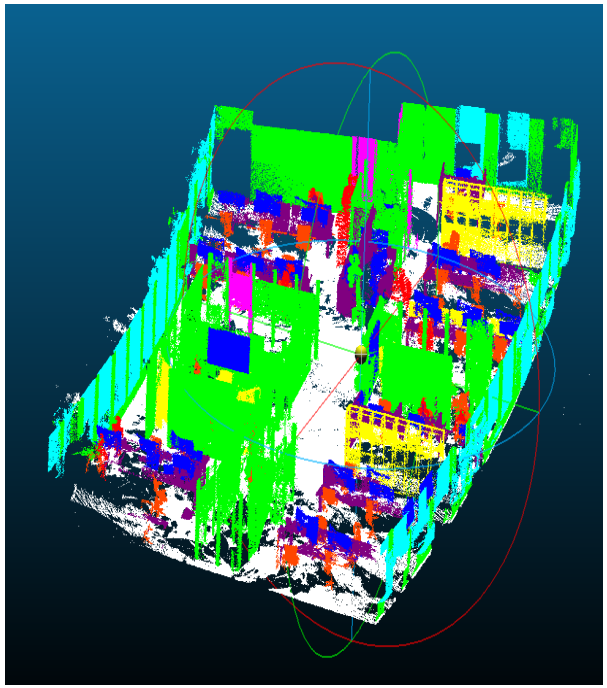


(c) Detailed object classification visualization of the CGI office in the realistic color scheme (RGB + Segment).

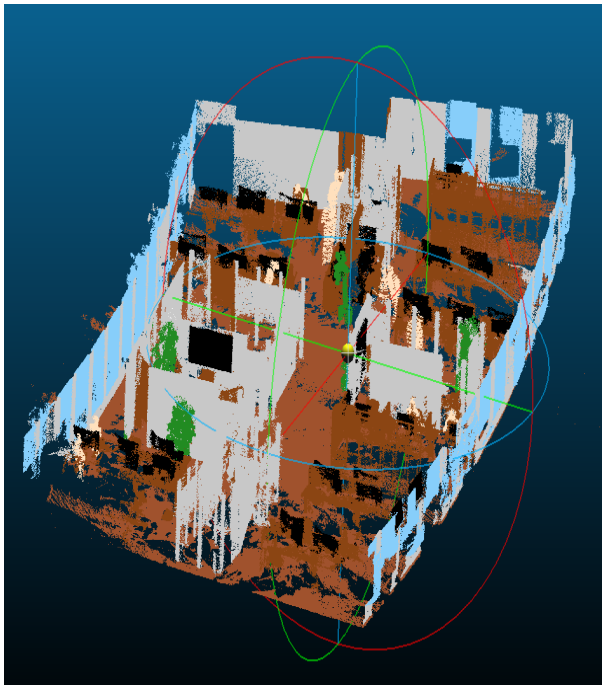
Figure 5.15: Comparison of detailed object classification visualization color schemes of the CGI office (RGB + Segment).



(a) Detailed object classification visualization of the CGI office in the functional color scheme (Segment).



(b) Detailed object classification visualization of the CGI office in the contrast color scheme (Segment).



(c) Detailed object classification visualization of the CGI office in the realistic color scheme (Segment).

Figure 5.16: Comparison of detailed object classification visualization color schemes of the CGI office (Segment).

5.3 Phase 3: Reflection and Evaluation

5.3.1 Sub-question 7

After creating the proof of concept in Cloud Compare, the final stage of this research is to test the effectiveness of the visualizations. The primary objective is to assess whether these visualizations improve SA and support decision making in (near) real-time emergency scenarios. The evaluation is carried out through expert interviews with emergency response professionals, safety region representatives, fire brigade personnel, Rotterdam control room personnel, and CGI colleagues.
During the feedback sessions, all the different visualizations were presented in CloudCompare, allowing the users to explore and assess them interactively. Rather than passively showing screen dumps of the different visualizations. During these feedback sessions, the author systematically guided them through each visualization, facilitating a structured discussion on their effectiveness. Throughout these sessions, continuous questioning was used to prompt reflections on usability, clarity, and potential improvements. This iterative process ensured that participants had enough opportunity to express their preferences and provide detailed feedback based on their operational expertise. The findings of sub-question 4 suggested that different user groups, control room personnel, CoPI commanders, and local responders, would have distinct visualization requirements based on their respective roles in emergency response operations. Given the diversity in responsibilities and decision making processes, it was expected that preferences for visualization types and color schemes would vary significantly between these groups. However, contrary to initial expectations, the results revealed a strong consensus among all participants about the most effective visualization method. Despite their differing perspectives, all user groups identified the same visualization approach as the most suitable to support SA and real-time decision making.

Effectiveness of Different Visualization Types

Basic Visualization
The proof of concepts with the basic visualization were designed to provide a general spatial overview using different color schemes to highlight structural elements. This visualization type was the first result presented to all user groups during the evaluation process. As this was their initial introduction to the potential of point cloud visualization for emergency response, many participants expressed surprise and enthusiasm for the possibilities this technology offers. Across all user groups, there was a strong positive first impression, with respondents noting that the visualization provided a clear and structured overview of the environment, making it easier to interpret building layout, see spatial relationships and make more informed decisions based on the segmented point cloud.

Respondents from the safety regions were particularly impressed by the ability to visualize a building's interior in such detail in such a short time. Highlighting the potential for improved SA when coordinating emergency responses. Furthermore, CoPI commanders recognized the value of this approach, emphasizing that it could significantly improve their ability to assess the locations of incidents before physically arriving at the scene. Finally, local responders were enthusiastic about the potential to navigate complex environments more efficiently using 3D point cloud visualizations.
Although the basic visualization was well received, several respondents suggested that a further classification of objects could improve its effectiveness. The current implementation effectively distinguishes structural components such as walls, floors, and ceilings, but is not a distinction in, for example, objects that could further improve its usability. The respondents noted that identifying moving versus fixed objects would be particularly useful for planning entry and evacuation routes. In addition, he also emphasized the importance of clearly highlighting navigational features, as these elements are crucial during high-pressure emergency situations.

Despite these suggestions for improvement, all user groups agreed that the basic visualization already provided substantial value. The ability to explore an environment in a structured and interactive way was seen as a significant step forward for emergency response applications. Users concluded that while the clarity and simplicity of the current visualization should be maintained, incorporating additional object classes would improve SA without overwhelming the user with unnecessary complexity. This finding underscores the importance of balancing detailed information with intuitive usability, ensuring that emergency responders can quickly and effectively interpret critical elements within the environment.

Obstacle Differentiation

The obstacle differentiation visualization builds on the foundation of the basic visualization by introducing a key improvement: differentiation between electrical objects and other structural components. Compared to basic visualization, this additional level of distinction significantly improves the interpretability of the environment and improves SA for emergency responders. Across all user groups, this visualization was widely regarded as a step forward compared to the basic visualization, as it provided more actionable insights and made it easier to assess potential hazards in emergency scenarios.

Higher-level control room personnel noted that this visualization provided a clearer overview of the environment, particularly when distinguishing potential electrical hazards from general obstacles. The ability to identify these elements at a glance was seen as a crucial improvement for real-time decision-making, as it allowed higher-level control room personnel to relay more specific information to responders in the field. Compared to the basic visualization, which provided only a general overview, this visualization added depth to the understanding of space without overwhelming the user with excessive detail.

Local responders, including firefighters and police, particularly appreciated the added differentiation between electrical objects and other components. Electrical hazards pose significant risks in fire-related incidents, making their clear identification essential. Firefighters emphasized that a further step could be taken by distinguishing between fixed and movable objects, which would help them assess whether an obstacle could be cleared from an evacuation path or if it was a permanent structure that had to be navigated around. Furthermore, they noted that emergency equipment, such as fire extinguishers or defibrillators, should be highlighted, as these are critical in certain emergency response scenarios.

CoPI commanders also found this visualization to be a significant improvement over the basic version. The ability to quickly differentiate between structural obstacles and electrical equipment provided them with more precise SA when evaluating building layouts and potential risks. However, several respondents noted that further classification refinements could improve its usability even more. Although electrical devices were now clearly distinguishable, other crucial objects such as emergency exit signs, fire extinguishers, and movable versus fixed objects remained undifferentiated. Adding these elements to the visualization could further improve its practical value for tactical decision making.

Although obstacle differentiation visualization already represents a major improvement over basic visualization, these findings suggest that further differentiation could make it even more effective and increase SA even further. In particular, the distinction between fixed and movable objects, as well as the emphasis on emergency exits and fire safety equipment, was frequently mentioned as potential improvements. Despite this, the overall response to this visualization was highly positive, as it struck a balance between providing additional detail without introducing excessive complexity.

Compared to the basic visualization, the obstacle differentiation visualization was unanimously regarded as more effective in all user groups. By introducing hazard differentiation, it significantly improved SA and provided emergency responders with a more structured and actionable representation of the environment. Although additional refinements could further improve its utility, this visualization already demonstrated clear benefits in supporting emergency decision making.

Detailed Object Classification Visualization

The detailed object classification visualizations are based on the visualization of obstacle differentiation by introducing a more granular breakdown of objects within the environment. This approach distinguishes between a wide range of object categories, such as furniture, equipment, and structural elements, offering an extensive level of detail. Although this additional differentiation theoretically improves the completeness of the spatial representation, in practice, the results indicate that this level of detail introduces excessive complexity, making the visualization overwhelming for emergency response purposes.

Across all user groups, respondents reported that this visualization contained too much information to be effectively interpreted in high-pressure situations. The control room personnel found that while the classification provided a rich dataset, it also increased cognitive load, making it harder to extract key spatial insights at a glance. In contrast to the obstacle differentiation visualization, which provided a structured and actionable overview, this approach fragmented the visualization into too many elements, making it less practical for real-time coordination. Respondents noted that distinguishing between too many object types made it harder to focus on the most critical spatial relationships, which are essential for decision making in emergencies.

Local responders, including firefighters and paramedics, found this visualization particularly overwhelming. Firefighters said that the sheer amount of object differentiation made it harder to focus on the most relevant hazards. The excessive number of categories demanded too much cognitive effort, which is not

practical in high-stress, time-sensitive situations. Instead of providing immediate and actionable insights, the detailed classification led to information overload, forcing users to filter irrelevant details rather than focus on critical navigation and safety concerns. Furthermore, it was noted that the distinction between different types of furniture was of little relevance to their decision-making process, further strengthening the idea that this level of object classification does not align with the needs of emergency response personnel.
CoPI commanders expressed similar concerns, emphasizing that this level of granular classification slowed their ability to assess risks and plan tactical responses. Although the ability to differentiate between various types of furniture and structural elements might be beneficial for post-incident analysis, it was not useful for real-time emergency response. The additional cognitive effort required to interpret the cluttered visualization was considered a drawback, as it reduced the speed of decision making, which is crucial in fast-evolving emergency scenarios.

A key limitation of this visualization is its impact on cognitive load and SA. Although the previous visualizations allowed users to rapidly extract key information, the detailed object classification required additional time and effort to process, reducing its overall efficiency. The increased extraneous cognitive load, the effort required to process irrelevant or excessive information, was frequently cited as the reason why this visualization was less useful in real-time emergency situations.
Compared to obstacle differentiation visualization, detailed object classification visualization was widely considered impractical for emergency response applications. Although it provided a comprehensive breakdown of the environment, the additional level of detail negatively impacted usability, slowing down decision making, and increasing cognitive load. For emergency situations where speed, clarity, and efficiency are essential, this visualization was too complex to be used effectively in the field. The results suggest that, while some degree of differentiation is beneficial, excessive classification detracts from SA rather than improving it.

Effectiveness of Different Color Schemes

Functional Color Scheme
The functional color scheme was the most effective across all types of visualization, particularly when used in the visualization of obstacle differentiation. The clear separation between structural elements, obstacles, and hazards allowed users to extract meaningful insights quickly and efficiently. Control room personnel, CoPI commanders, and local responders found that this approach provided the best balance between clarity and usability without overwhelming cognitive load.

Contrast Based Color Scheme
The contrast-based color scheme was consistently perceived as overwhelming and impractical for real-time decision making. Although it ensured strong visual separation between objects, it lacked contextual meaning, making interpretation more difficult. The extreme contrasts disrupted spatial comprehension, increased cognitive load, and slowing decision making processes. Across all user groups, this scheme was regarded as the least effective approach.

Realistic Color Scheme
The realistic color scheme was useful for spatial orientation, but it did not provide sufficient emphasis on critical elements such as doors, stairs, and hazards. Control room personnel and CoPI commanders found that while it helped with general environmental recognition, it lacked the level of differentiation needed for an effective emergency response. Local responders noted that, while it provided an intuitive representation, it did not sufficiently highlight potential risks, limiting its overall effectiveness in high-pressure scenarios.

Most Effective Approach for Improved Situational Awareness

Although initial expectations based on sub-question 4 suggested that different user groups would require distinct visualization methods tailored to their specific roles, the evaluation results demonstrated a clear consensus among all participants. Regardless of their operational responsibilities, control room personnel, CoPI commanders, and local responders identified obstacle differentiation visualization using the functional color scheme as the most effective approach to improving SA and supporting real-time decision making. This visualization method provided a clear and structured overview of the environment, allowing users to quickly recognize key structural elements, obstacles, and hazards. The improved differentiation

between navigable areas and electrical objects significantly improved SA, while the balanced level of detail ensured that users could extract the most relevant information without unnecessary cognitive load.
The findings suggest that while different emergency response roles have distinct responsibilities, their fundamental visualization must align. The ability to quickly and intuitively interpret critical information is a shared requirement among all groups. Continuing, the development of role-based customization options and adaptive visualization layers could further optimize this approach, ensuring that emergency response personnel have access to customized visualization methods that meet their specific operational needs.

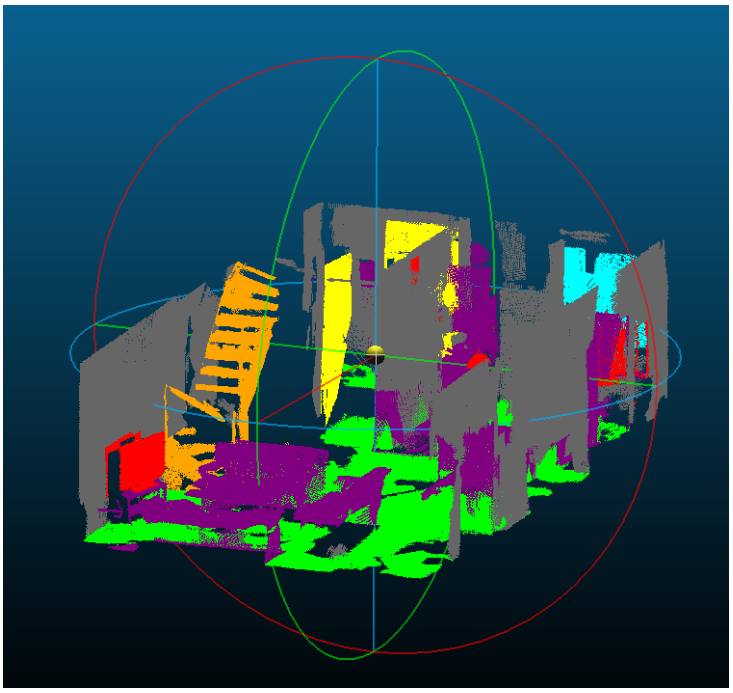


Figure 5.17: Visualization of Author's House Obstacle Differentiation in the Functional Color Scheme.

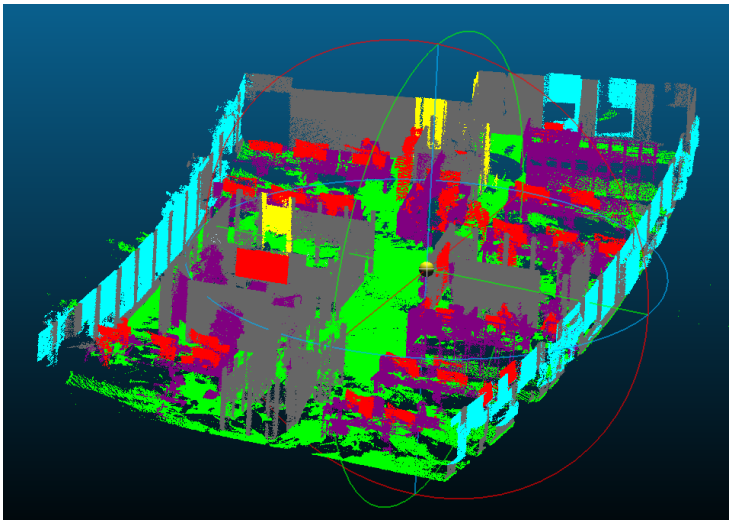


Figure 5.18: Visualization of the CGI Office Obstacle Differentiation in the Functional Color Scheme.

5.4 Challenges

The results of this research highlight several significant challenges and insights in achieving (near) real-time point cloud segmentation and effective visualization for first responders. This section examines the key issues that were encountered during this research.

5.4.1 Model Performance on Self-Acquired Data

One fundamental challenge was the failure of the segmentation model to perform on self-collected point cloud data, despite its strong results on benchmark datasets. The chosen model had demonstrated high accuracy on the S3DIS dataset, and it was possible to reproduce these results on the training data, according to the model's viability under controlled conditions. However, when the model was applied to a point cloud captured with an iPhone LiDAR in a real-world indoor environment, the output was unusable and all points were misclassified into a single generic class ("clutter"). This stark contrast to the rich segmentation scores available in the S3DIS dataset indicates a serious generalization problem. The key factors contributing to this failure likely include the discrepancies between the training data and the self-acquired data, such as higher noise levels, lower point density, and differences in spatial layout and object shapes in the scan. In essence, the model struggled to cope with the messiness and variability of real-world data outside its training distribution.
It is important to discern whether this failure was due to inherent limitations of the model or to issues in the implementation systems and process. On the one hand, the model architecture was not explicitly tailored to handle the specific characteristics of the sensor and environment used. The broad category of "clutter" in the model label set may have made it easy for the network to lump unknown or unclear objects into that class when faced with noisier input, rather than confidently identifying different classes. This points to limitations in the robustness of the model: it was trained on well-annotated static data sets and was not resistant to domain shift or sensor noise.
At the same time, the author acknowledges that implementation challenges and limited expertise also played a role, as unresolved errors in the data processing or inference pipeline could have contributed to poor results. Despite a careful pre-processing routine (including computing point normals and cropping to the area of interest) to fit the expected input format of the model, segmentation still failed, suggesting that something in the pipeline did not meet the model's requirement or the model required further tuning. The author's inability to pinpoint and fix the problem within the project time frame (due in part to their current programming skills) underscores that this was not a trivial integration.

5.4.2 Real-Time Processing Constraints

Achieving true real-time semantic segmentation proved to be beyond the reach of the current setup. Although the model should be able to perform in (near) real time, in practice the inference speed on the personal hardware did not meet the demands of live emergency response. High-performance architectures such as Point Transformer or Point-SAM are computationally intensive, making it challenging to run them on a laptop or mobile device. The pipeline from the point cloud capture to segmented output involves heavy data processing, and even minor inefficiencies introduce latency. In fact, fully seamless real-time operation was explicitly outside the scope of this research due to technical limitations and time constraints. This is a significant practical hurdle: In a fire or search-and-rescue scenario, responders cannot wait many seconds for a scene to update. The current performance gap suggests that either algorithm optimizations or specialized hardware would be required to truly achieve low-latency SA.

5.4.3 Visualization Usability and Cognitive Load

Developing the SA visualization revealed a tension between detail and usability. Rich, information-dense visualizations can overwhelm the user under stressful conditions. During expert evaluations, it became clear that more detail is not always better. In fact, the excessive complexity of the segmented view hindered effectiveness in high-pressure situations. Users faced with a multitude of colored classes and symbols took longer to interpret the scene, indicating an increased cognitive load. This challenge was to find an optimal level of abstraction: enough information to be useful but not so much that it overloads the responders attention. Ensuring that the visual output could be parsed at a glance (to support quick decision making) required simplifying the scene portrayal and carefully considering human cognitive limits. This limitation is aligned with cognitive load theory, which warns that too much information can reduce overall SA instead of improving it.

Chapter 6

Conclusions

This thesis aimed to further explore the possibilities of (near) real-time indoor segmentation and also to explore how to visualize these segmented point cloud data to increase situational awareness and decision making for first responders. The main research question was as follows:

To what extent is it possible to segment (near) real-time 3D point cloud data using an existing deep learning model, and how can this segmented point cloud be effectively visualized to increase situational awareness for first responders?

6.1 Sub-Questions

Before answering the main research question, all the sub-questions are answered.

6.1.1 Sub-Question 1: Model Selection

- **Which existing deep learning segmentation model is best suited for (near) real-time point cloud segmentation in indoor environments?**

A comprehensive review of nine state-of-the-art deep learning models for indoor semantic segmentation was conducted, focusing on segmentation accuracy and real-time applicability. At the time of writing, two models stood out: Point Transformer V3 + PPT, which achieved an S3DIS 6-fold mean IoU of 80.8, and Point-SAM, which reached a mIoU of 86.2 on the same benchmark.
Both models demonstrated strong segmentation performance in indoor environments, but their real-time suitability differs. Point Transformer V3 + PPT is better suited for (near) real-time applications because its serialized attention mechanism and efficient memory management allow for faster inference while maintaining high segmentation quality. Point-SAM, while achieving state-of-the-art accuracy, is computationally demanding, limiting its feasibility for real-time deployment, particularly on personal hardware.
The trade-off between segmentation accuracy and inference speed suggests that Point Transformer V3 + PPT is the best available model for (near) real-time indoor point cloud segmentation, although true real-time performance remains challenging on personal hardware. Achieving seamless real-time segmentation will require further advances in both model efficiency and hardware acceleration, as current deep learning models still require significant computational resources for real-world, dynamic operational environments.

6.1.2 Sub-Question 2: Reproducibility

- **To what extent is it possible to integrate the chosen segmentation model on a personal device and reproduce the accuracy results reported by the authors on the training data?**

The integration of the selected model onto a personal device was successful, with performance results closely matching those reported by the authors. The model maintained high accuracy on benchmark datasets, which confirmed the feasibility of deploying such models on personal hardware. However, minor discrepancies were observed due to differences in hardware configurations and pre-processing steps.

Although the model performed well under controlled conditions, achieving the exact reported accuracy required careful optimization of several parameters and computational resources. These findings indicate that reproducing state-of-the-art results is feasible, but hardware limitations and implementation details can introduce slight variations in performance.

6.1.3 Sub-Question 3: Segmenting Self-Acquired Data

- **How can the selected segmentation model be integrated and tested on self-acquired data?**

Integrating the selected segmentation model with self-acquired data proved to be a significant challenge, primarily due to difficulties in formatting the data correctly for inference and the model's limited ability to generalize beyond its training datasets. Despite extensive preprocessing efforts, including adjustments to data normalization and format conversion, the model failed to detect and segment the self-acquired data properly. Even with the assistance of CGI colleagues, meaningful segmentation was not achieved.
This issue could arise from errors in the preprocessing pipeline, a fundamental limitation in the model's ability to adapt to new, real-world datasets, or a combination of both. The findings suggest that Point Transformer V3 + PPT struggles with self-scanned LiDAR data, probably due to differences in point density, sensor noise, or scene variability not present in the benchmark datasets. However, the author's limited programming experience has also played a role in the challenges faced during data preparation and integration. Although an extensive effort was made to correctly format and preprocess the data, the complexity of deep learning-based point-cloud segmentation requires advanced debugging and optimization skills, which may have affected the final result.

6.1.4 Sub-Question 4: Defining User Needs

- **What are the key information needs during emergencies to support decision making and situational awareness for various stakeholders?**

An effective emergency response is based on clear and structured spatial information tailored to the needs of different user groups. This research identified three primary stakeholders: control room personnel, local responders, and CoPI commanders, each with distinct information requirements for visualized building data.
Lower-level control room personnel require rapidly interpretable overviews of structural layouts, hazard locations, and responder movements to facilitate efficient dispatch and coordination. Local responders operating in high-risk environments need simplified, actionable visualizations that emphasize building layouts, access points, and hazard zones for real-time decision making. CoPI commanders and higher-level control room personnel benefit from integrated multisource visualizations that offer a comprehensive operational overview, enabling effective resource management and interagency collaboration.

Although different stakeholders have different information needs, the findings indicate that a single structured visualization approach, combining key structural elements and hazard differentiation, was preferred in all user groups. Although an adaptive system with toggling capabilities was explored, respondents favored a streamlined and unified visualization that balanced clarity and critical information display to avoid cognitive overload. Furthermore, key structural elements, such as doors, stairs, elevators, windows, walls, and floors, were identified as crucial for navigation and hazard assessment. Their inclusion in visualizations significantly improves situational awareness, efficiency, and effectiveness of emergency response operations. The results further underscore the importance of maintaining a clear visual hierarchy, ensuring that critical spatial elements are prioritized while minimizing unnecessary complexity that could hinder rapid decision making. These findings highlight the need for user-centered, intuitive visualization strategies that align with the fast-paced, high-stakes nature of emergency response scenarios.

6.1.5 Sub-Question 5: Effective Visualization

- **How can scientific principles and cartographic methods be applied to effectively communicate segmented 3D point cloud data to diverse user groups?**

To ensure that segmented 3D point cloud data effectively supports emergency response operations, this research applied cartographic principles and visualization frameworks to improve clarity, usability, and situational awareness. The map use curve, cognitive load theory, and the new map communication model guided the development of a unified visualization system that improves situational awareness while addressing the specific needs of different user groups.
The map use cube was used to classify user groups based on their level of interaction and purpose. Lower-level control room personnel required moderate interaction, allowing them to navigate between rooms, structural elements, and responder locations. Local responders operating under extreme time constraints needed low-interaction visualizations emphasizing key navigational elements. CoPI commanders and higher-level control room personnel benefited from high-level overviews, integrating multiple data sources for a complete operational picture.
The cognitive load theory was applied to ensure that visual complexity remained manageable. The intrinsic load was controlled by prioritizing essential structural elements (walls, doors, and hazard zones), using progressive disclosure to reveal details as needed. The extraneous load was reduced by eliminating redundant elements and optimizing color schemes to display only relevant information. The germane load was improved by using contrast-based color schemes and interactive components that highlighted critical spatial features, such as evacuation routes and structural risks.
The new map communication model introduced an iterative design process, in which visualizations in Krita were tested with CGI colleagues and emergency response professionals. Their feedback helped refine usability, clarity, and overall effectiveness. Participants emphasized that a single structured visualization was more effective than role-specific adaptations, leading to refinements that ensured that the final design was clear, intuitive, and adaptable to different user needs rather than requiring multiple presets.
These findings led to the development of a unified visualization system that accommodates different levels of complexity while maintaining a consistent design across all user groups. Control room personnel can analyze detailed, multilayered data, local responders can quickly interpret simplified hazard indicators, and CoPI commanders can integrate live updates for strategic decision making. By applying cartographic and cognitive principles, the system ensures that segmented point cloud data remains actionable, improving both situational awareness and operational efficiency in emergency response scenarios.

6.1.6 Sub-Question 6: Proof of Concept

- **How can a proof-of-concept visualization be developed to demonstrate the effective communication of segmented point cloud data, tailored to different user needs?**

A proof-of-concept visualization was developed to assess the effective communication of segmented 3D point cloud data for emergency response. The visualization incorporated three levels of detail, progressively refining the representation of indoor environments. The first level provided a basic structural visualization, emphasizing key architectural elements such as walls, doors, and staircases. The second level introduced obstacle differentiation, distinguishing electrical objects from general obstacles to support hazard identification. The third level applied detailed object classification, offering fine-grained segmentation to facilitate more comprehensive situational awareness.
The development process began with conceptual sketches in Krita, where different segmentation strategies were explored to determine the most effective way to structure and visualize the data. These conceptual designs informed the manual segmentation and refinement process in CloudCompare, where point cloud data was classified into distinct structural and navigational components. To improve interpretability, three color schemes were implemented: a functional color scheme that assigned colors based on emergency relevance, a contrast-based scheme emphasizing maximum visual differentiation, and a realistic color scheme that replicated real-world colors for intuitive recognition.
This structured approach ensured that segmented point cloud data was effectively categorized and visually represented, providing a foundation for further evaluation and refinement in emergency response applications.

6.1.7 Sub-Question 7: Reflection and Evaluation

- **How effective are the proof-of-concept visualizations in increasing situational awareness through the communication of segmented point cloud data?**

The visualizations demonstrated significant potential for improving situational awareness by focusing on spatial recognition, navigation, and hazard detection, thus supporting a safer work environment. The visualization of obstacle differentiation using the functional color scheme emerged as the most effective, improving interpretability and situational awareness by maintaining a balance between clarity and essential detail. In contrast, the contrast-based scheme, while visually distinct, was overwhelming and hindered rapid comprehension, making it less suitable for high-pressure scenarios. The realistic color scheme, although familiar, failed to emphasize critical navigational elements, reducing its effectiveness in urgent decision-making contexts.
Although different user needs were identified, all emergency responders ultimately preferred the same structured visualization approach, which prioritized clarity, reduced cognitive load, and emphasized key structural elements and hazards. Selective hazard differentiation proved especially useful in highlighting risks without introducing unnecessary complexity. Furthermore, the results reinforce that excessive detail can impede rapid cognition, underscoring the need for a balanced approach that provides only the most relevant and actionable spatial information.
By maintaining a streamlined yet informative design, the proof-of-concept visualization allowed emergency responders to quickly extract relevant spatial information, assess risks, and make more informed decisions in high-stakes situations. Ultimately, increasing situational awareness and creating a safer workspace for first responders.

6.2 Main Research Question

- **To what extent is it possible to segment (near) real-time 3D point cloud data using an existing deep learning model, and how can this segmented point cloud be effectively visualized to increase situational awareness for first responders?**

Although (near) real-time segmentation of indoor 3D point clouds using existing deep learning models is theoretically feasible, practical deployment remains challenging due to difficulties in processing self-acquired data and computational constraints on personal hardware. Further improvements in preprocessing workflows, model adaptation for real-world data, and hardware acceleration are necessary to bridge this gap.
Effective visualization requires simplified designs that prioritize key structural elements while minimizing cognitive overload. Although different user needs were identified, all respondents ultimately preferred the same structured visualization approach, emphasizing the importance of clarity and intuitive design over role-specific adaptations. By presenting only the most relevant spatial information in an easily interpretable format, visualization improves situational awareness by allowing emergency responders to quickly assess environments, identify hazards, and make informed decisions under time pressure.
Furthermore, the integration of segmented and RGB data is crucial for improving interpretability, as segmentation alone provides geometric structure but lacks the visual context necessary for rapid and effective decision making in emergency scenarios. By combining both data types, the visualization ensures that first responders receive a comprehensive, real-time understanding of their surroundings, ultimately supporting faster response times and improved operational efficiency in high-stakes environments.

6.3 Future work

6.3.1 Improving Real-Time Segmentation Performance

Techniques to achieve faster inference on 3D data are essential. Future efforts could explore model compression, more efficient architectures, or streaming approaches (processing the point cloud incrementally) to reduce latency. Optimizing the code for parallel processing and leveraging GPU acceleration more effectively would help address the real-time challenge identified (where current models struggled to run at speed on personal devices).The goal would be to reach a point where segmentation can keep up with data acquisition, enabling a true SLAM+segmentation workflow.

6.3.2 Exploring Alternative Models or Hardware Optimizations

Given the failure of the model in noisy self-scans, trying different deep learning models could be fruitful, for example, models known for robustness to sparse data or ones trained in datasets closer to our use case. Additionally, hardware solutions should be considered: running the segmentation on powerful edge computing devices or cloud servers (with results streamed back to responders) could bypass the current device limitations. Specialized hardware like AI accelerators or lighter-weight frameworks (for example, using a smaller neural network or a knowledge-distilled version of the model) might achieve a better speed-accuracy trade-off. Research could compare the performance of the chosen model with emerging approaches to find a better fit for on-the-fly use.

6.3.3 Improving Visualization for Better Situational Awareness

Future research should expand the visualization component, making it more adaptive, intuitive, and tailored to different operational needs. One direction is to implement adaptive legends and emphasis, for example, dynamically highlighting the most relevant symbols or categories based on the current context. If the system detects numerous heat sources (possible fires) or certain hazards, the interface could automatically bring those to the forefront (larger icons, blinking, or sorted to top of legend) to draw user attention. Another improvement could be the integration of hybrid mapping capabilities in 2D and 3D. As prior work suggests, responders can benefit from both a top-down floor-plan view and a 3D perspective. A future system might allow seamless switching between a 3D point cloud view and a 2D map overlay (for example, a floor layout with icons for hazards and team members). This gives flexibility: A commander can quickly check the 2D overview for the overall status of the incident, then dive into the 3D view for details of a particular room. In addition, role-specific user interfaces need to be further developed. Although this project introduced the concept of role-based layers, it can be taken to the next level by designing distinct interface modes for each type of user. A firefighter interface might be simplified to the essentials, while an incident commander interface could include rich controls to query and filter information, and a control room operator interface might emphasize tracking of personnel and resources on the map. Importantly, the visualization should also adapt in real-time to changing data. As new information comes in from the field (for example, updated scans or sensor readings), the map should update or re-symbolize accordingly, ensuring the view is always current. In summary, making the visualization smarter, able to adjust what it shows and how it shows based on context, user role, and live data, will further improve SA and user experience.

6.3.4 Conducting Real-World Usability Testing

Once the technical and visualizations improvements are in place, it is crucial to validate the system under realistic conditions. Future research should involve field trials with first responders, such as training exercises in which teams use the 3D segmentation tool during a simulated incident. Such studies can measure tangible benefits (for example, reduced time to locate victims or hazards) and reveal any usability issues that only become apparent in context. It would also allow collecting direct feedback from the target end-users about what information they find most valuable or what features they wish to have. This real-world evaluation would complement the expert interviews by adding observational data on performance and user behavior. Insights from these tests could then loop back into further refining the tool and moving the system closer to an operational product.

Chapter 7

Appendix

Appendix 1: Test.py

Listing 7.1: Point Cloud Parsing Script

```
"""
Main Testing Script

Author: Xiaoyang Wu (xiaoyang.wu.cs@gmail.com)
Please cite our work if the code is helpful to you.
"""

from pointcept.engines.defaults import (
    default_argument_parser,
    default_config_parser,
    default_setup,
)
from pointcept.engines.test import TESTERS
from pointcept.engines.launch import launch


def main_worker(cfg):
    cfg = default_setup(cfg)
    tester = TESTERS.build(dict(type=cfg.test.type, cfg=cfg))
    tester.test()


def main():
    args = default_argument_parser().parse_args()
    cfg = default_config_parser(args.config_file, args.options)

    launch(
        main_worker,
        num_gpus_per_machine=args.num_gpus,
        num_machines=args.num_machines,
        machine_rank=args.machine_rank,
        dist_url=args.dist_url,
        cfg=(cfg,),
    )


if __name__ == "__main__":
    main()
```

Appendix 2: Slow pace iPhone scan of the authors home

Figure 7.1: Slow pace iPhone scan of the authors home Scan

Appendix 3: Fast pace iPhone scan of the authors home

Figure 7.2: Fast pace iPhone scan of the authors home Scan

Appendix 4: preprocessing_normals.py

Listing 7.2: Point Cloud Parsing Script

```
1
2  import os
3  import argparse
4  import numpy as np
5  import open3d as o3d
6
7  def parse_room(
8      ply_file,
9      save_path,
10     save_format,
11     batch_size=400_000,
12     max_ram_gb=8,
13     clean_mesh=True
14 ):
15     print(f"=== Parsing '{ply_file}'... ===")
16
17     # Load the point cloud
18     pcd = o3d.io.read_point_cloud(ply_file)
19
20     # Extract coordinates and colors
21     room_coords = np.asarray(pcd.points)
22     room_colors = np.asarray(pcd.colors)
23     if len(room_colors) == 0:
24         room_colors = np.zeros_like(room_coords)
25
26     # Calculate bounding box and crop (if needed)
27     x_min, y_min, z_min = np.min(room_coords, axis=0)
28     x_max, y_max, z_max = np.max(room_coords, axis=0)
29     padding = 0.1
30     max_bound = np.array([x_max, y_max, z_max]) + padding
31     min_bound = np.array([x_min, y_min, z_min]) - padding
32     bbox = o3d.geometry.AxisAlignedBoundingBox(min_bound=min_bound, max_bound=max_bound)
33     cropped_pcd = pcd.crop(bbox)
34
35     assert len(pcd.points) == len(cropped_pcd.points), "Cropping bbox messed things up"
36
37     # Compute normals directly on the cropped point cloud
38     if not cropped_pcd.has_normals():
39         print("Normals missing, computing directly on point cloud...")
40         cropped_pcd.estimate_normals(search_param=o3d.geometry.KDTreeSearchParamHybrid(
               radius=0.1, max_nn=30))
41
42     # Extract normals from the point cloud
43     room_normals = np.asarray(cropped_pcd.normals)
44
45     # Save in S3DIS format
46     if save_format == 's3dis':
47         # Create numpy arrays for S3DIS format
48         zeros = np.repeat(0, room_coords.shape[0]).reshape([-1, 1])  # Create empty
               labels
49
50         # Save in the S3DIS format (usually .npy files)
51         np.save(os.path.join(save_path, 'coord.npy'), room_coords)  # 3D coordinates
52         np.save(os.path.join(save_path, 'color.npy'), room_colors)  # RGB colors
53         np.save(os.path.join(save_path, 'normal.npy'), room_normals)  # Normals
54         np.save(os.path.join(save_path, 'semantic.npy'), zeros)  # Semantic ground truth
               (empty)
55         np.save(os.path.join(save_path, 'instance.npy'), zeros)  # Instance ground truth
               (empty)
56
57         print(f"=== Completed parsing {ply_file} in S3DIS format ===\n")
58         return
59
60     print("Unsupported save format. Only 's3dis' is supported in this script.")
61     return
62
63 if __name__ == "__main__":
64     parser = argparse.ArgumentParser()
65     parser.add_argument(
```

```
        "--raw_ply_path", required=True,
        help="Path to the raw PLY file"
    )
    parser.add_argument(
        "--save_path", required=True,
        help="Directory to save output files"
    )
    parser.add_argument(
        "--save_format", required=True,
        choices=['s3dis'],
        help="Output format (currently only supports 's3dis')"
    )
    parser.add_argument(
        "--batch_size", type=float, default=400_000,
        help="Batch size for processing (default: 400,000)"
    )
    parser.add_argument(
        "--max_ram_gb", type=float, default=8.0,
        help="Maximum RAM usage in GB (default: 8 GB)"
    )
    args = parser.parse_args()

    print(f"Raw PLY Path: {args.raw_ply_path}")
    print(f"Output Path: {args.save_path}")
    print(f"Maximum RAM Usage: {args.max_ram_gb} GB\n")

    os.makedirs(args.save_path, exist_ok=True)
    parse_room(
        args.raw_ply_path,
        args.save_path,
        args.save_format,
        args.batch_size,
        args.max_ram_gb
    )
```

Appendix 5: Inference Script

Listing 7.3: Point Cloud Parsing Script

```

import os
import numpy as np
import torch
import torch.nn.functional as F
from collections import OrderedDict
from pointcept.datasets import build_dataset
from pointcept.models import build_model
from pointcept.datasets.transform import Compose
from pointcept.utils.config import Config
from pointcept.utils.visualization import save_point_cloud
from pointcept.utils.comm import get_world_size
from pointcept.datasets.utils import point_collate_fn

from tensorboardX import SummaryWriter

# Define the color map for each class
CLASS_COLOR_MAP = {
    "ceiling": (255, 0, 0),     # Red
    "floor": (0, 255, 0),       # Green
    "wall": (0, 0, 255),        # Blue
    "beam": (255, 255, 0),      # Yellow
    "column": (255, 0, 255),    # Magenta
    "window": (0, 255, 255),    # Cyan
    "door": (128, 0, 0),        # Dark Red
    "table": (128, 128, 0),     # Olive
    "chair": (0, 128, 128),     # Teal
    "sofa": (128, 0, 128),      # Purple
    "bookcase": (0, 128, 0),    # Dark Green
    "board": (128, 128, 128),   # Gray
    "clutter": (64, 64, 64)     # Dark Gray
}

CLASS_NAMES = list(CLASS_COLOR_MAP.keys())
CLASS_COLOR_MAP_INDEXED = {i: CLASS_COLOR_MAP[name] for i, name in enumerate(CLASS_NAMES)
    }

data_config = dict(
    type="S3DISDataset",
    split="Area_6",
    data_root="/XXXX/XXXX/pointcept/data/dokstraat_slecht",
    transform=[
        dict(type="CenterShift", apply_z=True),
        dict(
            type="GridSample",
            grid_size=0.03,
            hash_type="fnv",
            mode="train",
            return_grid_coord=True,
        ),
        dict(type="CenterShift", apply_z=False),
        dict(type="NormalizeColor"),
        dict(type="ToTensor"),
        dict(
            type="Collect",
            keys=("coord", "grid_coord", "segment", "name"),
            feat_keys=("color", "normal"),
        ),
    ],
    test_mode=False,
    test_cfg=dict(  # Added test_cfg
        voxelize=dict(
            type="GridSample",
            grid_size=0.02,  # Adjust grid size if necessary
            mode="test",
        )
    ),
)

```

```python
model_config = dict(
    type="PT-v3m1",
    in_channels=6,
    order=("z", "z-trans", "hilbert", "hilbert-trans"),
    stride=(2, 2, 2, 2),
    enc_depths=(2, 2, 2, 6, 2),
    enc_channels=(32, 64, 128, 256, 384),
    enc_num_head=(2, 4, 8, 16, 32),
    enc_patch_size=(1024, 1024, 1024, 1024, 1024),
    dec_depths=(2, 2, 2, 2),
    dec_channels=(64, 64, 128, 256),
    dec_num_head=(4, 4, 8, 16),
    dec_patch_size=(1024, 1024, 1024, 1024),
    mlp_ratio=4,
    qkv_bias=True,
    qk_scale=None,
    attn_drop=0.0,
    proj_drop=0.0,
    drop_path=0.3,
    shuffle_orders=True,
    pre_norm=True,
    enable_rpe=False,
    enable_flash=True,
    upcast_attention=False,
    upcast_softmax=False,
    cls_mode=False,
    pdnorm_bn=False,
    pdnorm_ln=False,
    pdnorm_decouple=True,
    pdnorm_adaptive=False,
    pdnorm_affine=True,
    pdnorm_conditions=("ScanNet", "S3DIS", "Structured3D"),
)

checkpoint = "/XXXX/XXXX/pointcept/exp/s3dis/s3dis-semseg-pt-v3m1-1-ppt-extreme/model/
    model_best.pth"
keywords = "backbone."
replacement = ""
idx = 300

if __name__ == "__main__":
    model = build_model(model_config).cuda()
    checkpoint = torch.load(
        checkpoint, map_location=lambda storage, loc: storage.cuda()
    )
    weight = OrderedDict()
    for key, value in checkpoint["state_dict"].items():
        if not key.startswith("module."):
            key = "module." + key  # xxx.xxx -> module.xxx.xxx
        if keywords in key:
            key = key.replace(keywords, replacement)
        if get_world_size() == 1:
            key = key[7:]  # module.xxx.xxx -> xxx.xxx
        weight[key] = value
    load_state_info = model.load_state_dict(weight, strict=False)
    print(load_state_info)
    dataset = build_dataset(data_config)
    data = dataset[idx]
    for key in data.keys():
        if isinstance(data[key], torch.Tensor):
            data[key] = data[key].cuda(non_blocking=True)
    output = model(data)

    # Determine which classes are present in the point cloud
    segment_labels = data['segment'].cpu().numpy()  # Get segment labels
    unique_classes = np.unique(segment_labels)
    present_classes = [CLASS_NAMES[i] for i in unique_classes if i < len(CLASS_NAMES)]
    print("Classes present in the point cloud:", present_classes)

    # Map segment labels to colors directly using indexed color map
    color_map = np.array([CLASS_COLOR_MAP_INDEXED.get(label, (255, 255, 255)) for label
        in segment_labels]) / 255.0

```

```
    # Save the point cloud with class-specific colors
    os.makedirs("/XXXX/XXXX/pointcept/output/vis_class_color", exist_ok=True)
    point_cloud_path = f"/XXXX/XXXX/pointcept/output/vis_class_color/segmented_color{idx}.ply"
    save_point_cloud(
        coord=output.coord,
        color=color_map,
        file_path=point_cloud_path,
    )
```

Appendix 6: RGB point cloud of the CGI office

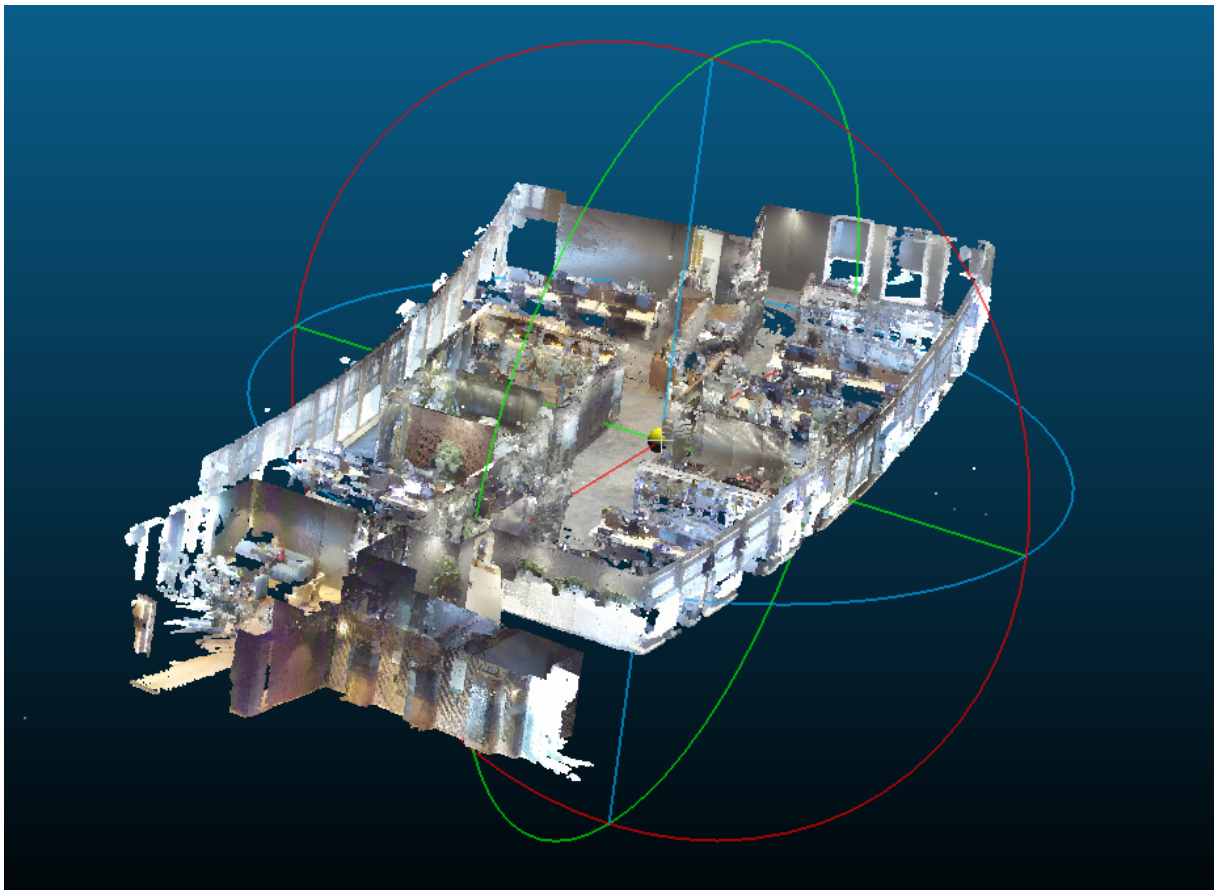


Figure 7.3: RGB point cloud of the CGI office

List of Figures

List of Tables

Bibliography

Armeni, I., Sener, O., Zamir, A. R., Jiang, H., Brilakis, I., Fischer, M., & Savarese, S. (2016). 3d semantic parsing of large-scale indoor spaces. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1534–1543. https://openaccess.thecvf.com/content_cvpr_2016/papers/Armeni_3D_Semantic_Parsing_CVPR_2016_paper.pdf

Bassier, M., Van Genechten, B., & Vergauwen, M. (2019). Classification of sensor independent point cloud data of building objects using random forests. *Journal of Building Engineering*, *21*, 468–477. https://doi.org/10.1016/j.jobe.2018.04.027

Bavle, H., Sanchez-Lopez, J. L., Cimarelli, C., Tourani, A., & Voos, H. (2023). From slam to situational awareness: Challenges and survey. *Sensors*, *23*(10), 4849. https://doi.org/10.3390/s23104849

Bertin, J. (1983). *Semiology of graphics: Diagrams, networks, maps.* Madison: University of Wisconsin Press.

Bhargava, A., & Bansal, A. (2021). Fruits and vegetables quality evaluation using computer vision: A review. *Journal of King Saud University - Computer and Information Sciences*, *33*(3), 243–257. https://doi.org/10.1016/j.jksuci.2018.06.002

Boykov, Y., & Funka-Lea, G. (2006). Graph cuts and efficient nd image segmentation. *International Journal of Computer Vision*, *70*(2), 109–131. https://doi.org/10.1007/s11263-006-7934-5

Bunch, R. L., & Lloyd, R. E. (2006). The cognitive load of geographic information. *The Professional Geographer*, *58*(2), 209–220. https://doi.org/10.1111/j.1467-9272.2006.00527.x

Camilus, K. S., & Govindan, V. K. (2012). A review on graph based segmentation. *International Journal of Image, Graphics and Signal Processing*, *4*(5), 1. https://doi.org/10.5815/ijigsp.2012.05.01

Cheng, B., Misra, I., Schwing, A. G., Kirillov, A., & Girdhar, R. (2022). Masked-attention mask transformer for universal image segmentation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1290–1299. https://openaccess.thecvf.com/content/CVPR2022/papers/Cheng_Masked-Attention_Mask_Transformer_for_Universal_Image_Segmentation_CVPR_2022_paper.pdf

Dai, A., Chang, A. X., Savva, M., Halber, M., Funkhouser, T., & Nießner, M. (2017). Scannet: Richly-annotated 3d reconstructions of indoor scenes. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 5828–5839. https://openaccess.thecvf.com/content_cvpr_2017/papers/Dai_ScanNet_Richly-Annotated_3D_CVPR_2017_paper.pdf

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*. https://arxiv.org/pdf/2010.11929

Endsley, M. R. (1995). Measurement of situation awareness in dynamic systems. *Human Factors*, *37*(1), 65–84. https://doi.org/10.1518/001872095779049499

Endsley, M. R. (2015). Situation awareness misconceptions and misunderstandings. *Journal of Cognitive Engineering and Decision Making*, *9*(1), 4–32. https://doi.org/10.1177/1555343415572631

Endsley, M. R. (2021). A systematic review and meta-analysis of direct objective measures of situation awareness: A comparison of sagat and spam. *Human Factors*, *63*(1), 124–150. https://doi.org/10.1177/0018720819875376

Endsley, M. R., Bolté, B., & Jones, D. G. (2003). *Designing for situation awareness: An approach to user-centered design.* CRC press.

Fang, H., Xin, S., Zhang, Y., Wang, Z., & Zhu, J. (2020). Assessing the influence of landmarks and paths on the navigational efficiency and the cognitive load of indoor maps. *ISPRS International Journal of Geo-Information*, *9*(2), 82. https://doi.org/10.3390/ijgi9020082

Forsyth, D. A., & Ponce, J. (2002). *Computer vision: A modern approach.* Prentice Hall.

Furtado, P. G. F., Hirashima, T., & Hayashi, Y. (2018). Reducing cognitive load during closed concept map construction and consequences on reading comprehension and retention. *IEEE Transactions on Learning Technologies*, *12*(3), 402–412. https://doi.org/10.1109/TLT.2018.2861744

Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., & Garcia-Rodriguez, J. (2017). A review on deep learning techniques applied to semantic segmentation. *arXiv preprint arXiv:1704.06857*. https://doi.org/10.48550/arXiv.1704.06857

Grilli, E., Menna, F., & Remondino, F. (2017). A review of point clouds segmentation and classification algorithms. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLII-2/W3*, 339–344. https://doi.org/10.5194/isprs-archives-XLII-2-W3-339-2017

Gupta, S., Davidson, J., Levine, S., Sukthankar, R., & Malik, J. (2017). Cognitive mapping and planning for visual navigation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2616–2625. https://openaccess.thecvf.com/content_cvpr_2017/papers/Gupta_Cognitive_Mapping_and_CVPR_2017_paper.pdf

Hamilton, K., Mancuso, V., Mohammed, S., Tesler, R., & McNeese, M. (2017). Skilled and unaware: The interactive effects of team cognition, team metacognition, and task confidence on team performance. *Journal of Cognitive Engineering and Decision Making*, *11*(4), 382–395. https://doi.org/10.1177/1555343417731429

Helmy, B. E.-D. (2019, June). Image processing: Graph-based segmentation [Accessed: 2025-03-24].

Janai, J., Güney, F., Behl, A., Geiger, A., et al. (2020). Computer vision for autonomous vehicles: Problems, datasets and state of the art. *Foundations and Trends® in Computer Graphics and Vision*, *12*(1–3), 1–308. https://doi.org/10.1561/0600000079

Kapucu, N., & Garayev, V. (2011). Collaborative decision-making in emergency and disaster management. *International Journal of Public Administration*, *34*(6), 366–375. https://doi.org/10.1080/01900692.2011.561477

Karthick, S., Sathiyasekar, K., & Puraneeswari, A. (2014). A survey based on region based segmentation. *International Journal of Engineering Trends and Technology*, *7*(3), 143–147. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=753ae5aab292ada416fe361bf7387aea3df71066

Kent, A. (2018). Form follows feedback: Rethinking cartographic communication. *Westminster Papers in Communication and Culture*, *13*(2), 96–112. https://doi.org/10.16997/wpcc.296

Khan, A. I., & Al-Habsi, S. (2020). Machine learning in computer vision. *Procedia Computer Science*, *167*, 1444–1451. https://doi.org/10.4108/eai.21-4-2021.169418

Kirillov, A., He, K., Girshick, R., Rother, C., & Dollár, P. (2019). Panoptic segmentation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 9404–9413. https://openaccess.thecvf.com/content_CVPR_2019/papers/Kirillov_Panoptic_Segmentation_CVPR_2019_paper.pdf

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., et al. (2023). Segment anything. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 4015–4026. https://openaccess.thecvf.com/content/ICCV2023/papers/Kirillov_Segment_Anything_ICCV_2023_paper.pdf

Kolodiazhnyi, M., Vorontsova, A., Konushin, A., & Rukhovich, M. (2024). Oneformer3d: One transformer for unified point cloud segmentation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 20943–20953. https://openaccess.thecvf.com/content/CVPR2024/papers/Kolodiazhnyi_OneFormer3D_One_Transformer_for_Unified_Point_Cloud_Segmentation_CVPR_2024_paper.pdf

Kraak, M.-J., & Ormeling, F. (2020). *Cartography: Visualization of geospatial data*. CRC Press.

Krygier, J., & Wood, D. (2024). *Making maps*. Guilford Publications.

Le, T. V., Kulikowski, C. A., & Muchnik, I. B. (2008). A graph-based approach for image segmentation. *International Symposium on Visual Computing*. https://api.semanticscholar.org/CorpusID:8732846

Li, N., Becerik-Gerber, B., Krishnamachari, B., & Soibelman, L. (2014). A bim centered indoor localization algorithm to support building fire emergency response operations. *Automation in Construction*, *42*, 78–89. https://doi.org/10.1016/j.autcon.2014.02.019

Li, X., Ding, H., Yuan, H., Zhang, W., Pang, J., Cheng, G., Chen, K., Liu, Z., & Loy, C. C. (2024). Transformer-based visual segmentation: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. https://doi.org/10.1109/TPAMI.2024.3434373

Liu, Z., Tang, H., Lin, Y., & Han, S. (2019). Point-voxel cnn for efficient 3d deep learning. *Advances in Neural Information Processing Systems (NeurIPS)*, *32*. https://proceedings.neurips.cc/paper/2019/file/5737034557ef5b8c02c0e46513b98f90-Paper.pdf

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 3431–3440. https://doi.org/10.1109/CVPR.2015.7298965

Lorenz, A., Thierbach, C., Baur, N., & Kolbe, T. H. (2013). Map design aspects, route complexity, or social background? factors influencing user satisfaction with indoor navigation maps. *Cartography and Geographic Information Science*, *40*(3), 201–209. https://doi.org/10.1080/15230406.2013.807029

MacEachren, A. M. (1995). *How maps work: Representation, visualization, and design*. Guilford Press.

Maturana, D., & Scherer, S. (2015). Voxnet: A 3d convolutional neural network for real-time object recognition. *Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 922–928. https://doi.org/10.1109/IROS.2015.7353481

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, *5*, 115–133. https://doi.org/10.1007/BF02478259

Minaee, S., Abdolrashidi, A., Su, H., Bennamoun, M., & Zhang, D. (2023). Biometrics recognition using deep learning: A survey. *Artificial Intelligence Review*, *56*(8), 8647–8695. https://doi.org/10.1007/s10462-022-10237-x

Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., & Terzopoulos, D. (2021). Image segmentation using deep learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(7), 3523–3542. https://doi.org/10.1109/TPAMI.2021.3059968

Nikoohemat, S., Diakité, A. A., Zlatanova, S., & Vosselman, G. (2020). Indoor 3d reconstruction from point clouds for optimal routing in complex buildings to support disaster management. *Automation in Construction*, *113*, 103109. https://doi.org/10.1016/j.autcon.2020.103109

NIPV. (2009). Kwalificatiedossier calamiteitencoördinator meldkamer [Aangeboden aan het ministerie van Binnenlandse Zaken en Koninkrijksrelaties in december 2009]. https://archief.nipv.nl/wp-content/uploads/sites/2/2022/03/calamiteitencoordinator-meldkamer-kwalificatiedossier-poc-2009.pdf

NIPV. (2020). Kwalificatiedossier centralist meldkamer [Aangeboden aan het ministerie van Binnenlandse Zaken en Koninkrijksrelaties]. https://nipv.nl/wp-content/uploads/2022/11/20200320-Kwalificatiedossier-Centralist-meldkamer.pdf

Nöllenburg, M. (2007). Geographic visualization. *Human-Centered Visualization Environments: GI-Dagstuhl Research Seminar, Dagstuhl Castle, Germany, March 5–8, 2006, Revised Lectures*, 257–294. https://doi.org/10.1007/978-3-540-71949-6_6

Nossum, A. S. (2011). Indoortubes: A novel design for indoor maps. *Cartography and Geographic Information Science*, *38*(2), 192–200. https://doi.org/10.1559/15230406382192

Nossum, A. S. (2013). Developing a framework for describing and comparing indoor maps. *The Cartographic Journal*, *50*(3), 218–224. https://doi.org/10.1179/1743277413Y.0000000055

Nurunnabi, A., Belton, D., & West, G. (2016). Robust segmentation for large volumes of laser scanning three-dimensional point cloud data. *IEEE Transactions on Geoscience and Remote Sensing*, *54*(8), 4790–4805. https://doi.org/10.1109/TGRS.2016.2551546

Oh, J., Hebert, M., Jeon, H.-G., Perez, X., Dai, C., & Song, Y. (2019, October). Explainable semantic mapping for first responders [arXiv preprint]. https://doi.org/10.48550/arXiv.1910.07093

Ohta, Y. I., Kanade, T., & Sakai, T. (1978). An analysis system for scenes containing objects with substructures. *Proceedings of the Fourth International Joint Conference on Pattern Recognition (IJCPR)*, 752–754. https://www.cs.cmu.edu/~efros/courses/LBMV09/newpapers/OhtaKanadeSakai1978.pdf

Osman, T., Psyche, S. S., Ferdous, J. M. S., & Zaman, H. U. (2017). *Intelligent traffic management system for cross section of roads using computer vision*. IEEE. https://doi.org/10.1109/CCWC.2017.7868350

Poux, F. (2019). *The smart point cloud: Structuring 3d intelligent point data* [Doctoral dissertation, Universite de Liege (Belgium)]. https://www.proquest.com/openview/68fd56e6a4febf7971a9313e44e1ba77/1.pdf?cbl=2026366&diss=y&pq-origsite=gscholar

Poux, F., Neuville, R., Hallot, P., & Billen, R. (2016). Smart point cloud: Definition and remaining challenges. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *IV-2/W1*(W1). https://doi.org/10.5194/isprs-annalen-IV-2-W1-1-2016

Qi, C. R., Su, H., Mo, K., & Guibas, L. J. (2017a). Pointnet: Deep learning on point sets for 3d classification and segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pat-*

tern Recognition (CVPR), 652–660. https://openaccess.thecvf.com/content_cvpr_2017/papers/Qi_PointNet_Deep_Learning_CVPR_2017_paper.pdf

Qi, C. R., Yi, L., Su, H., & Guibas, L. J. (2017b). Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*, *30*. https://doi.org/10.1109/ICCV.2017.16

Qian, G., Li, Y., Peng, H., Mai, J., Hammoud, H., Elhoseiny, M., & Ghanem, B. (2022). Pointnext: Revisiting pointnet++ with improved training and scaling strategies. *Advances in Neural Information Processing Systems (NeurIPS)*, *35*, 23192–23204. https://proceedings.neurips.cc/paper_files/paper/2022/file/9318763d049edf9a1f2779b2a59911d3-Paper-Conference.pdf

Rantakokko, J., Rydell, J., Strömbäck, P., Händel, P., Callmer, J., Törnqvist, D., Gustafsson, F., Jobs, M., & Grudén, M. (2011). Accurate and reliable soldier and first responder indoor positioning: Multisensor systems and cooperative localization. *IEEE Wireless Communications*, *18*(2), 10–18. https://doi.org/10.1109/MWC.2011.5751291

Robert, D., Raguet, H., & Landrieu, L. (2023). Efficient 3d semantic segmentation with superpoint transformer. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 17195–17204. https://openaccess.thecvf.com/content/ICCV2023/papers/Robert_Efficient_3D_Semantic_Segmentation_with_Superpoint_Transformer_ICCV_2023_paper.pdf

Salmon, P. M., Stanton, N. A., Walker, G. H., Jenkins, D., Ladva, D., Rafferty, L., & Young, M. (2009). Measuring situation awareness in complex systems: Comparison of measures study. *International Journal of Industrial Ergonomics*, *39*(3), 490–500. https://www.sciencedirect.com/science/article/abs/pii/S0169814108001625

Schmidt, S., & Götze, H.-J. (1998). Interactive visualization and modification of 3d-models using gis-functions. *Physics and Chemistry of the Earth*, *23*(3), 289–295. https://www.gravity.uni-kiel.de/Curso-Caracas/UebungModellierung/Egs97.htm

Smit, B.-P. (2020). *Creating remote situation awareness of indoor first responder operations using slam* [Master's thesis, Utrecht University] [Master's thesis]. https://www.gdmc.nl/publications/2020/MScThesisBart-PeterSmit.pdf

Stanton, N. A., Chambers, P. R. G., & Piggott, J. (2001). Situational awareness and safety. *Safety Science*, *39*(3), 189–204. https://bura.brunel.ac.uk/bitstream/2438/1804/1/Situation_awareness_and_safety_Stanton_et_al.pdf

Strudel, R., Garcia, R., Laptev, I., & Schmid, C. (2021). Segmenter: Transformer for semantic segmentation. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 7262–7272. https://openaccess.thecvf.com/content/ICCV2021/papers/Strudel_Segmenter_Transformer_for_Semantic_Segmentation_ICCV_2021_paper.pdf

Sultana, F., Sufian, A., & Dutta, P. (2020). Evolution of image segmentation using deep convolutional neural network: A survey. *Knowledge-Based Systems*, *201-202*, 106062. https://doi.org/10.1016/j.knosys.2020.106062

Sweller, J. (2011). Cognitive load theory. In B. H. Ross (Ed.), *Psychology of learning and motivation* (pp. 37–76, Vol. 55). Academic Press. https://doi.org/10.1016/B978-0-12-387691-1.00002-8

Tashakkori, H., Rajabifard, A., & Kalantari, M. (2015). A new 3d indoor/outdoor spatial model for indoor emergency response facilitation. *Building and Environment*, *89*, 170–182. https://doi.org/10.1016/j.buildenv.2015.02.036

Teixeira, H., Magalhães, A., Ramalho, A., Pina, M. d. F., & Gonçalves, H. (2021). Indoor environments and geographical information systems: A systematic literature review. *SAGE Open*, *11*(4), 21582440211050379. https://doi.org/10.1177/21582440211050379

Tyner, J. A. (2014). *Principles of map design*. Guilford Publications.

van der Meer, T. (2018). *Geovisualization for the dutch fire brigade* [Master's thesis, Utrecht University] [Master's thesis]. https://www.gdmc.nl/publications/2018/MSc_thesis_Tom_van_der_Meer.pdf

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *30*. https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep learning for computer vision: A brief review. *Computational Intelligence and Neuroscience*, *2018*, 1–13. https://doi.org/10.1155/2018/7068349

Wang, J., Ma, Y., Zhang, L., Gao, R. X., & Wu, D. (2018). Deep learning for smart manufacturing: Methods and applications. *Journal of Manufacturing Systems*, *48*, 144–156. https://doi.org/10.1016/j.jmsy.2018.01.003

Wang, Y., Sun, Y., Liu, Z., Sarma, S. E., Bronstein, M. M., & Solomon, J. M. (2019). Dynamic graph cnn for learning on point clouds. *ACM Transactions on Graphics*, *38*, 1–12. https://doi.org/10.1145/3326362

Wolf, D., Prankl, J., & Vincze, M. (2015). Fast semantic segmentation of 3d point clouds using a dense crf with learned parameters. *Proceedings of the 2015 IEEE International Conference on Robotics and Automation (ICRA)*, 4867–4873. https://doi.org/10.1109/ICRA.2015.7139875

Wu, X., Jiang, L., Wang, P.-S., Liu, Z., Liu, X., Qiao, Y., Ouyang, W., He, T., & Zhao, H. (2024). Point transformer v3: Simpler, faster, stronger. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 4840–4851. https://openaccess.thecvf.com/content/CVPR2024/papers/Wu_Point_Transformer_V3_Simpler_Faster_Stronger_CVPR_2024_paper.pdf

Yu, D., Wang, H., Chen, P., & Wei, Z. (2014). Mixed pooling for convolutional neural networks. *Proceedings of the 9th International Conference on Rough Sets and Knowledge Technology (RSKT 2014)*, 364–375. https://doi.org/10.1007/978-3-319-11740-9_34

Yu, H., Yang, Z., Tan, L., Wang, Y., Sun, W., Sun, M., & Tang, Y. (2018). Methods and datasets on semantic segmentation: A review. *Neurocomputing*, *304*, 82–103. https://doi.org/10.1016/j.neucom.2018.03.037

Zhang, B., Tian, Z., Tang, Q., Chu, X., Wei, X., Shen, C., et al. (2022). Segvit: Semantic segmentation with plain vision transformers. *Advances in Neural Information Processing Systems*, *35*, 4971–4982. https://proceedings.neurips.cc/paper_files/paper/2022/file/20189b1aaa8edbb6d8bd6c1067ab5f3f-Paper-Conference.pdf

Zhou, Y., Gu, J., Chiang, T. Y., Xiang, F., & Su, H. (2024). Point-sam: Promptable 3d segmentation model for point clouds. *arXiv preprint arXiv:2406.17741*. https://doi.org/10.48550/arXiv.2406.17741