applied sciences

MDPI

Review

Methods and Applications of Space Understanding in Indoor Environment—A Decade Survey

Sebastian Pokuciński and Dariusz Mrozek *

Department of Applied Informatics, Silesian University of Technology, 44-100 Gliwice, Poland; sebastian.pokucinski@polsl.pl
* Correspondence: dariusz.mrozek@polsl.pl; Tel.: +48-32-237-13-39

Abstract: The demand for digitizing manufacturing and controlling processes has been steadily increasing in recent years. Digitization relies on different techniques and equipment, which produces various data types and further influences the process of space understanding and area recognition. This paper provides an updated view of these data structures and high-level categories of techniques and methods leading to indoor environment segmentation and the discovery of its semantic meaning. To achieve this, we followed the Systematic Literature Review (SLR) methodology and covered a wide range of solutions, from floor plan understanding through 3D model reconstruction and scene recognition to indoor navigation. Based on the obtained SLR results, we identified three different taxonomies (the taxonomy of underlying data type, of performed analysis process, and of accomplished task), which constitute different perspectives we can adopt to study the existing works in the field of space understanding. Our investigations clearly show that the progress of works in this field is accelerating, leading to more sophisticated techniques that rely on multidimensional structures and complex representations, while the processing itself has become focused on artificial intelligence-based methods.

Keywords: decade survey; indoor environment analysis; BIM; machine learning; room segmentation; room classification

check for updates

Citation: Pokuciński, S.; Mrozek, D. Methods and Applications of Space Understanding in Indoor Environment—A Decade Survey. *Appl. Sci.* **2024**, *14*, 3974. https://doi.org/10.3390/app14103974

Academic Editors: Ryan Gibson and Hadi Larijani

Received: 5 April 2024
Revised: 25 April 2024
Accepted: 29 April 2024
Published: 7 May 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Nowadays, the digitalization of building structures is a common practice. In many countries, newly started public projects are required to define a Building Information Modeling (BIM) workflow. This requirement results in electronic documentation describing even the most detailed aspects of the facility. It allows for project coordination from the construction stage to the operation and maintenance of already functioning property. For companies operating during the Fourth Industrial Revolution (Industry 4.0), it is clear that such models are crucial for optimizing manufacturing processes, automating space management, tracking the most valuable resources, and improving the safety of employees. The demand for digital descriptions and electronic representations of buildings has led to the extensive development of diverse software systems. These systems help manage all the information and make it easily accessible—not only during property construction but at any time when needed. In general, BIMs may comprise various data: 3D models, asset descriptions, parameters of used materials, available equipment, etc. However, a straightforward question is yet to be answered: *What about already existing properties?* Private facilities built before the digital era have only physical building plans, usually stored in paper form. Proper documentation is more likely to be found for public properties and factories, but there may be no reliable documentation other than the building itself. As BIM plays the role of an information repository, there is always a possibility to conduct an inspection and gather all required data, even manually. An existing facility can be upgraded with the Internet of Things (IoT) sensors, smart assistants, and intelligent robots.

We can construct a digital twin of the building and create a 3D model of it. The floor plan can be reconstructed using drones flying through the factory. This is all possible but requires excessive time and (most likely human) resources. The problem becomes urgent when we consider the number of measurements and other data demanded by modern machine-learning-based solutions to work efficiently. To achieve the best performance possible, all the mentioned smart systems require a complete semantic understanding of their application field and its high-quality models. For example, a wearable tracking bracelet needs to know what room the calculated employee's coordinates represent. This creates a niche for systems capable of moving building data from any available source (like paper, pictures, or the existing property itself) to e-resources in an automated way. The best possible scenario would be to find a system that creates the documentation without user interaction other than data input. It should recreate not only the model but also understand its components and retrieve their meaning, most likely with extensive use of artificial intelligence and machine-learning-based decision models. One of the aspects of data processing in such systems is how to accomplish the operation of space segmentation (e.g., into rooms) and its categorization (e.g., into types of rooms). Although multiple potential solutions are available, they all noticeably differ in terms of used data structures and methodologies. In the literature, even a fast search for solutions based on a specific input data type (like a floor plan or point cloud) returns many possibilities and different approaches. They stretch from typical ones, such as the spatial data of 3D point clouds or colorful 2D images, to extraordinary ones, like sound chirps or energy measurements.

The **primary motivation** for our work is this unprecedented diversity of possibilities. The industry needs to choose proper input data and select it based on the expected implementation reality, such as the limited computing power of sensors or the short battery life of robots. However, there is a visible lack of proper systematic reviews and surveys that would organize already proposed, implemented, and tested solutions with a particular focus on the processed data, utilized methods, and their actual applications. To the best of our knowledge, in the last decade, there has been only one survey discussing the topic of room segmentation on sample input data type of floor plan images—the one conducted by Bormann et al. [1]. Similarly, during the research, we found only two other surveys focused on spatial data and discussing 3D building modeling [2,3]. Searching for a survey focused on the diversity of available data structures yielded no results. There is expected to be a significant improvement in digitalization technology and both the quantity and quality of available data sources each year, not to mention each decade. We can clearly observe the need for a catalog of the newest possibilities. This would have to be massive and comprehensive so that the companies (at all stages of their digital transformation) can use it as a guide through the extensiveness of data they can analyze and applications they can benefit from. It should present what types of data are worth gathering and how they can be used. It is also worth discussing the commonly found challenges and mistakes so that others do not repeat them.

To answer this need, in this paper we present a set of taxonomies dedicated to space understanding in the indoor environment. These taxonomies organize the current state of knowledge on three different levels: processed input data types, high-level category of the processing method, and actual low-level application. We summarize each of the taxonomies with extensive discussion and conclude the observations in a section dedicated to visible research challenges. We propose a custom review protocol and present literature research on space segmentation and classification systems across diverse input data structures, with a primary focus on the processing of 2D images. In this analysis, we followed the instructions of a Systematic Literature Review (SLR) [4–6]. A guide to this methodology was presented by Xiao and Watson [7]. Their process is grounded in eight consecutive steps executed one after another. It allows for minimizing the researcher bias that can occur in the case of an entirely subjective choice of papers. Based on their proposals, we adapted the idea, created a set of twelve execution steps, and used them analogically—to prepare a new research protocol, adjusted and dedicated to our needs. Previous works

were overly task-focused and mainly discussed only the newly presented algorithms and under specific conditions. To widen the research base, we wanted to avoid focusing on any particular solution and expected the whole research to be data-focused instead. We expected the analyzed systems to fulfill just one essential requirement: solutions needed to work correctly in the indoor environment.

In summary, in this paper we extend the current state of literature by:

- a customized SLR on the topic of room segmentation and classification,
- three interdisciplinary taxonomies for the research done so far and summarizing research findings,
- extended discussion of the application scenarios of found solutions,
- description of observed challenges and research directions that need further analysis.

The remainder of this paper is structured as follows: Section 2 describes the methodology we used, the proposed survey execution protocol, and its conduction results. Sections 3–5 present the taxonomies of input data type, high-level abstraction category of the performed process, and its low-level application (taxonomy of the accomplished task), respectively. Section 6 provides a bibliometric analysis of the articles found. Section 7 focuses on the encountered challenges the researchers are trying to overcome, Section 8 discusses the newest trends in the area, and Section 9 summarizes this document with a short review of the findings and a general survey conclusion.

2. Review Methodology and Conduction

To make the selection of the literature as objective as possible, the proposed review methodology followed the SLR steps for the planning and organization of the research. Initial assumptions and requirements were defined and described before the actual study was conducted. We asked three research questions that formulated the discussed problem in an organized manner (Section 2.1). Presented sets of filtering criteria narrowed the search results to only scientifically important papers (Section 2.2). We used multiple publication libraries and searched each one with the same, previously prepared query (Section 2.3). Finally, we combined all the elements to develop a step-by-step review protocol, which can be easily followed and reviewed at any time (Section 2.4).

2.1. Research Questions (RQ)

The main goal of this study was to create an overview of the room segmentation and classification methods that could be easily visualized. Questions we raised in this section were formulated to organize the results into a set of taxonomy diagrams. Each diagram represents knowledge gathered from answering one of the questions asked. To make the taxonomies cover a possibly wide range of solutions, they had to be generic enough to cover many possible relevant answers. At the same time, the questions had to be precisely formulated so as not to include too many papers unrelated to the topic.

RQ1: What input data structures are in use? The approach chosen for the entire survey was to be primarily data-driven. It created the need for a list of input data structures, which are typically available and used in room segmentation and classification tasks. We expected to find data structures of multiple dimensions, such as 2D-floor plan images or 3D point cloud models. Additional subtype divisions were specified as needed.

RQ2: What is the high-level abstraction category of the performed process? After identifying the input data structure, we separated different classes of solutions. Found methodologies proved to be highly diverse and incomparable between one another without additional grouping. We included instance segmentation or multi-class classification to the expected categories.

RQ3: What is the accomplished task? The third taxonomy was dedicated to the actual practical application of the analyzed methodology. We expected the same category of performed process to be used in multiple, often very different types of tasks, like 3D model reconstruction, floor plan prediction, or robot localization.

We answered all three of the stated research questions for each of the papers found during the conducted survey. This way, the preparation of the taxonomies and proper paper classification was faster and more likely to be truly objective. Gathered research papers were automatically organized by the answers given, going top to bottom by (1) input data type, (2) solution category, and (3) implemented task.

2.2. Filtering Criteria

To specify the rules of paper acceptance, we formulated filtering criteria. We split them into two categories. *Exclusion Criteria* (EC) described attributes that a publication **could not have**. In contrast, the *Inclusion Criteria* (IC) specified the requirements that **had to be fulfilled**. The order of the filtering directly influenced the amount of work required for the subsequent analysis steps. At the very beginning of the review protocol, we applied filters to the static attributes of the papers. As *'static'*, we understood attributes that were not a subject for discussion and could be easily found, such as the publication date or its language. The more time-consuming the evaluation of the exclusion criteria was, the later it was executed since its analysis potentially would not have been required any longer if the easier evaluated criteria had filtered the result beforehand. All the specified filtering criteria, including their numbers, titles, and descriptions, are presented in Table 1.

Table 1. Filtering criteria used during the research, split into Exclusion Criteria (EC) and Inclusion Criteria (IC).

No.	Title	Description
EC1	Published before 2012	To keep the research up to date, the survey conducted was focused on the newest methodologies—from the last decade only.
EC2	Duplicated article	As we searched multiple publication databases, the same article could be found in many different sources but was supposed to be analyzed only once.
EC3	Not written in English	English was chosen as the only accepted language. It was important to check the whole paper, as it happened to find results with English titles and abstracts but foreign language content.
EC4	Not concerning a topic, at least potentially related to the room segmentation or classification	Although we used a precise search query, the found papers' relevance was not guaranteed. We checked them manually and verified if the general topic of the article discussed floor plan analysis, spatial data processing methods, or at least an issue that could lead to room segmentation in any different type of data.
EC5	Full text not found	Reliable paper analysis requires the publications to be read and understood. Titles or abstracts alone were not enough.
EC6	Does not describe the process in detail	Papers without a precise description of the methodology used were rejected. The presentation of only the research results was not enough to fully answer the research questions.
EC7	Describes only ideas, discussions, or interviews	The objective of this study was to include publications of substantial value and precise descriptions of the papers. They were required to be implemented reliably, tested, and their results had to be available.
IC1	The topic must indicate the, at least potential, use in the indoor environment	This paper focuses on closed spaces, which can segment rooms inside of a building, not areas outside of it. This criterion filtered out solutions dedicated to large-scale outdoor applications, like the analysis of aerial photos.
IC2	Method must include some form of automated processing	The idea is to compare systems of somehow unsupervised data processing. Descriptions of fully manual processes, design guidelines, or manually carried out reports were omitted.
IC3	Article must reference at least 10 other papers	As the survey should be based only on reliable and scientifically important articles, analyzed papers were expected to be based on at least ten reviewed references.
IC4	Solution must process room—or higher structure—level data	We want to filter out solutions focused on internal single-room analysis. An example of such a scenario was the furniture segmentation task or wall décor recognition. To fulfill this criterion, the algorithm had to be able to segment at least one instance of a room or one class for the whole room needed to be recognized.
IC5	Article must describe the achieved performance and datasets used	Only papers with reliable results presentations were accepted. To fulfill this criterion, a description of the performance evaluation method had to be presented. The public availability of the datasets was not required, but their description was.

2.3. Search Query and Data Sources

The construction of a proper search query, used during the survey, was one of its most essential tasks. Overly vague phrases would overflow us with irrelevant publications, while too precise ones would result in an excessive restriction of the search results. For this survey, we proposed to use two groups of examples to build the search query: '*room [activity]*' and '*floor plan [activity]*'. For the room group, there was a set of seven activities chosen: '*segmentation*', '*classification*', '*clustering*', '*structure*', '*detection*', '*recognition*' and '*labeling*'. For the floor plan group, the set contained only three activities: '*segmentation*', '*clustering*', and '*analysis*'. This way, we expected the returned results to provide a wide range of possible differentiated input data types (because of the room examples group) but with an additional focus on processing 2D images (because of the floor plan examples group). Adjusted search phrases were used within the advanced search engines of the selected databases as a single merged query. We searched six different databases of publications. They all collect numerous publications, are fully digital, and are available via web browsers. In many cases, they are also accessible to academic researchers free of charge via their institutions.

The selected databases were as follows:

- Association for Computing Machinery Digital Library (ACM DL),
- Institute of Electrical and Electronics Engineers Xplore (IEEE Explore),
- Digital Bibliography & Library Project (dblp),
- Scopus,
- Elsevier Science Direct (SD),
- Springer Link (SL).

2.4. Review Protocol

All requirements and data sources were combined into a unified review protocol. We specified two flows of processing: main flow and survey extension. The first one determined and processed articles returned from the initial search. The second one focused on 2D image processing and extended the survey with potentially relevant articles found as referenced in the initial set and described as highly important. The entire procedure we proposed consisted of twelve steps: ten of the main flow and two of the survey extension. Each protocol step was done sequentially and extended with notes about its execution. A summary of the articles gathering is presented in the diagram in Figure 1. We represented every single step of the protocol with one dedicated block of the diagram. Each block comprises the step number, its name, and the filtering criteria used (if any were applied). Red numbers are dedicated to the main survey flow. Blue numbers indicate the survey extension process and the number of additionally reviewed papers. Details of each step's execution are described in the following part of this section.

STEP 1: Initial search. The combined search phrase was used to query all six proposed databases. The number of returned results was noted separately for each database. Without a filter for publication year, we found over three thousand records. Most were found in the Springer Link database, and the least in IEEE Explore.

STEP 2: Attributes filtering. Depending on the actual features of the database search engine, the article's publication date (EC1) was manually checked or introduced additional parameters to the search query. Applying the first exclusion criteria reduced the number of overall records returned by over 35%. The reduction ranged from 22% for IEEE Explore to 45% for Springer Link.

STEP 3: Results concatenation. All papers found were aggregated to form one larger initial set of articles for further filtering. The entire set was checked, and duplicated findings were removed (EC2). The concatenated results constructed a collection of over two thousand records. Duplicate removal excluded 236 papers from further analysis, leaving 1854 records in the set.

STEP 4: Titles screening. The first level of actual paper reading was to examine its title. Initially, it was checked for the language used (EC3). Later, the title's topic was assessed for conformity to the room segmentation and classification problem (EC4) and

indoor environment usage (IC1). Over two-thirds of the papers were removed, leaving 634 records in the set.

STEP 5: Abstracts screening. Abstracts were assessed and checked for language (EC3) and topic (EC4, IC1). Additionally, we confirmed that the proposed methodology worked without extended user involvement (IC2). Abstract screening reduced the set of papers by 36%, leaving 402 records in the collection.

STEP 6: Full-text obtainment. If all previous criteria were satisfied, the publication was checked for access availability (EC5). The whole paper had to be downloaded as a file or easily readable in an online reader.

STEP 7: Full-text screening. We checked the publication the same way as titles and abstracts (EC3, EC4, IC1, IC2), but this time with the use of the whole article's text. The number of filtered records was noted for steps 6 and 7 combined, reducing the total set by an additional 57%, leaving 172 papers in the collection.

STEP 8: Bucketing by input data type. In this step, we answered the first research question (RQ1). For each article, we specified the processed input data type and used it to categorize the papers. The analysis led to the creation of five buckets of publications. Selected for further analysis, the bucket of 2D images turned out to be the largest one. Including the survey extension, we assigned 87 articles to it. We found a similar number of papers discussing 3D spatial data. The rest of the discovered data types were significantly less numerous. Only two additional review papers were found on room segmentation and classification. A deeper analysis of the taxonomy created based on the input data type is presented in Section 3.

STEP 9: Bucketing by solution category. In this step, we answered the second research question (RQ2). We split the initial buckets of articles into subgroups based on the presented high-level abstraction category (performed process) of the analyzed solutions. We were able to specify four such categories and describe them in Section 4.

STEP 10: Bucketing by accomplished task. The last step of the main protocol flow was focused on the low-level task of the analyzed methodology and its actual application in real life. It answered the third research question (RQ3). We found several different applications and organized them into the third proposed taxonomy. It is described in detail in Section 5.

STEP 11: Full text analysis. The first step of the survey extension was to read the entire article and validate that its full content presented a detailed solution (EC6) and was not a discussion, idea description, or interview (EC7). It had to process entire rooms (IC4) and describe the achieved performance (IC5). This step was based on a deep, precise understanding of the article and methodology analysis. From the initially selected 72 articles processing 2D images, 51 papers were determined as passing all stated filtering criteria and were to be checked with further referenced articles analysis.

STEP 12: Survey extension. The goal of the last step of the review protocol was to extend the set of publications with articles that were important and related to the analyzed topic. They could be the basis for found solutions or be referenced in them and described as highly important. This step was executed only once—not to include too many papers recursively. The extension suggested the following (in decreasing order): 36 new articles for the 2D Images bucket, 9 for the Feature Set, 3 for 3D Spatial Data, and two for Graph Structure and Review Papers.

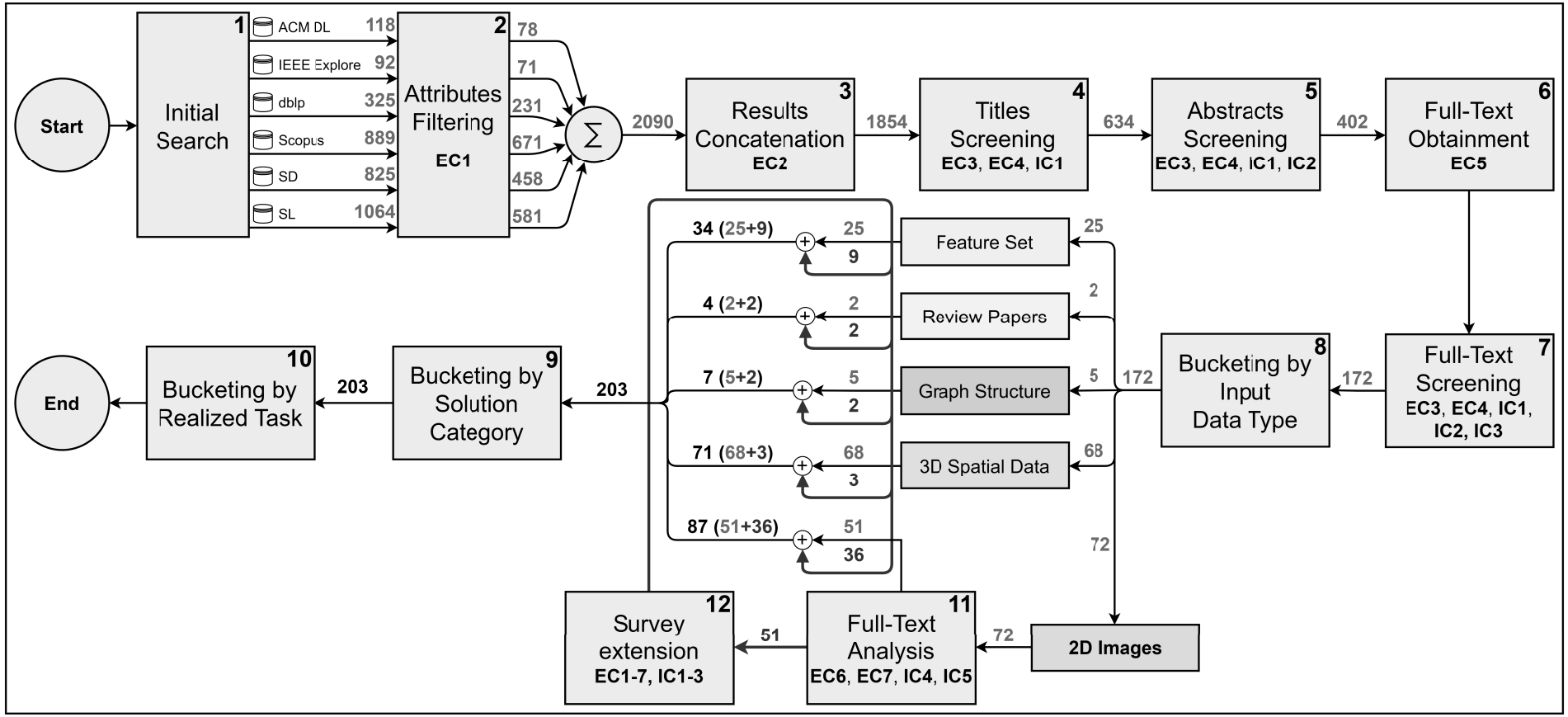

Figure 1. Review protocol diagram with details representing the number of papers after each step's execution. Red numbers provide information about the main execution flow and blue ones about the survey extension. Introduced color-coding for input data types as follows: 2D Images in green, 3D Spatial Data in blue, Graph Structures in red, Review Papers in gray, and a Feature Set in yellow.

3. Taxonomy of Input Data Types

The first taxonomy organizes the analyzed works by the input data type, which is related to the equipment used to obtain the data and further influences the processing algorithms used. It allows readers to familiarize themselves with the most common representations of data structures and the wide range of their kinds.

3.1. Taxonomy Presentation

The found input data types, as from the first bucketing step of the survey execution protocol, can be seen in Figure 2. Review papers were omitted from the analysis because they do not represent any actual type of input data themselves.

The four main types of structures we specified are:

- 2D Images,
- 3D Spatial Data,
- Graph Structures,
- Mixed Feature Sets.

2D Images were further divided into three and the feature sets into six sub-types of input data structures. In total, twelve types of data were distinguished. In Table 2, we present referenced publications organized by the processed input data type. As 2D images were the most commonly used and are generally the easiest to obtain, we decided to study them more closely. They serve as the primary input data type for most object detectors, are relatively easy to analyze, and are utilized in various scenarios, both academic and non-academic alike. Due to their popularity, flat images are handled in many different ways and by algorithms of various computing complexity. We extend our survey in the direction of 2D images to be able to fully cover their applications and present their popularity in found solutions.

Table 2. Referenced articles organized by the input data type. Numbers in brackets indicate the number of articles found in the main survey's flow (first number) and its extension (second number).

Input Data Type	Subtype	Found Papers
3D Spatial Data (68 + 3)	-	[8–78]
Graph Structure (5 + 2)	-	[79–85]
2D Images (51 + 36)	Floor Plan / Sketch (26 + 15)	[86–126]
	Occupancy Map (17 + 6)	[1,127–148]
	Environment Picture (8 + 15)	[149–171]
Feature set (25 + 9)	CAD-Like Data (1 + 0)	[172]
	Energy Consumption (2 + 0)	[173,174]
	Laser Range Measurements (5 + 0)	[175–179]
	Mixed (10 + 9)	[180–198]
	Radio Signal Fingerprint (2 + 0)	[199,200]
	Sound echo, chirp, RF (5 + 0)	[201–205]

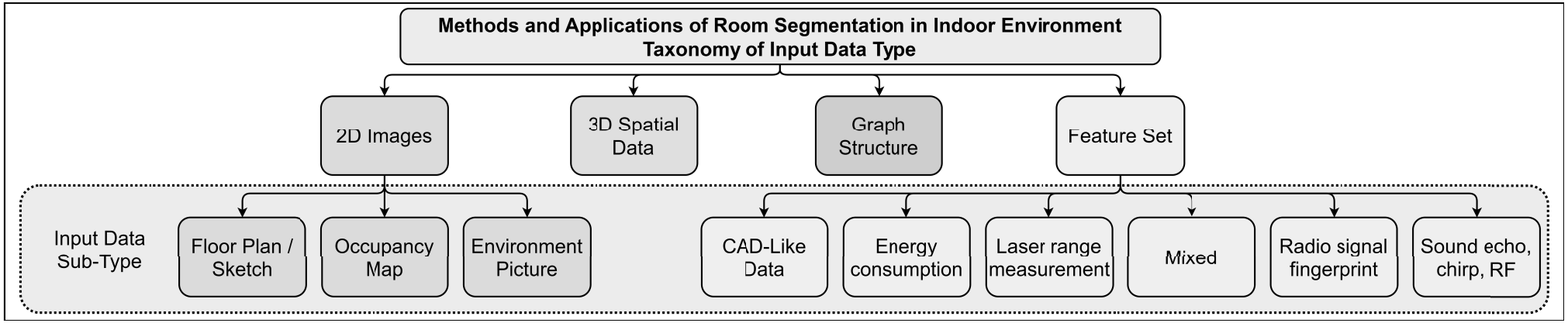

Figure 2. Diagram presenting first constructed taxonomy—based in article grouping by input data structures.

The **2D Images** group represents all the data that could be delivered as images, pictures, document scans, or sketches. It covers the rather raw environment representation, such as its planar depictions, as well as clearly more sophisticated inputs, like real pictures taken by an agent. They visibly differed in the applied processing methodology—with a vision-based approach dedicated to real photos and algorithmic processing for the more symbolic representations. Their essential aspect was that each 2D image could be described as a matrix consisting of columns and rows of numerical data and processed in a pixel-based manner. This way, the group covered a wide range of possible image representations regardless of the number of information channels they provided. It could be a typical black-and-white picture (one channel, 0–255), an occupancy grid map (one channel, empty/occupied/unknown), a colorful image (three channels, RGB, 0–255), or even a depth-aware image (four channels, RGBD, 0–255). As long as the processing remained at the pixel level (was aware of the pixel's x and y coordinates), the input data were classified as a 2D image. To somehow organize the group, we specified three subtypes: floor plan/sketch, occupancy map, and environment picture. The first one covered pictures and scans of the floor plans. They could be images of professional architectural plans from dedicated software or even pictures of freehand sketches. The second subtype grouped occupancy maps generated, e.g., from the laser sensors of home cleaning robots. The third one combined all the pictures of the real indoor environment, photographs taken with a mobile phone, professional indoor panoramas, or frames of videos from surveillance cameras.

The **3D Spatial Data** group gathers input data types that somehow describe the spatial positioning of the measurements. This group is dominated by point-cloud-based solutions, so we specified no additional subtypes. Other data structures we found were significantly less numerous, with a 3D mesh model of a building as an example. Not to overcomplicate the taxonomy, we decided to classify the three-dimensional trajectory as spatial data as well. This way, we extended this group with solutions processing readings from tracked Inertial Measurement Units (IMUs), mobile devices, or mobile scanners. A specific case that we assigned to this group was a way of depth-aware RGBD image processing. In the solution found, images were not analyzed as matrices of data (pixel-examination was abandoned) but were used to construct a point cloud instead.

The **Graph Structures** group combines methodologies processing input data represented as graphs of dependencies. In this case, the most popular solution was to use the spatial instances of building structures (rooms, corridors, areas) as graph nodes and their adjacency or connections as graph edges. The aspect determining if a found solution should be a member of this group was the usage of graph theory. If the data provided were formatted in such a way that graph theory could be used, the method was classified as processing graph structures.

The **Feature Set** group is the most diverse. Because of this diversity, we separated it into six different subtypes: CAD-like data, energy consumption measurements, laser range measurements, mixed input data, radio signal fingerprints, and the combined group of sound, echo, chirping, and radio frequencies-based input. Subtypes differ significantly from one another. The Computer Aided Design (CAD) data were much like the models typically used by architects, e.g., exported from an already existing BIM project. Energy consumption measurements were based on home energy meters and the individual energy profile of the device. Laser range measurements represented all the input data that were not an occupancy grid and came from laser object-to-obstacle distance measurements. Mixed input data combined situations where multiple, not directly related parameters, such as a picture of the building façade and expected room sizes, were provided. Radio signal fingerprints were used as distance determinants between a signal source and reference stations. The last subtype combined relatively uncommon data—measurements obtained via sound analysis. An example was a smartphone's microphone readings after a chirp generation from the smartphone's loudspeaker.

3.2. Taxonomy Discussion

Summarizing, the created taxonomy showed that there is virtually no limitation or minimal requirement for the input data complexity to perform room segmentation and recognition successfully. The range of inputs stretched from trivial series of numbers (one dimension) and matrixes of values (two dimensions), through the spatial location of measurements (three dimensions), to complicated mixes of different information with an unlimited number of features. This variety allowed researchers to choose the data structure that best suited the hardware capabilities of their equipment. The least powerful devices generated simplified measurements, such as power usage readings or radio signal strength. We can observe that the more computational power the sensor has, the more complex data it returns. Worth mentioning is a visible change in the approach, from offline data gathering and later processing (if the device was not powerful enough) to edge-computing and transfer to the cloud of only already deducted conclusions. Mobile platforms with better processors were capable of online analysis of even complicated visual input or continuous readings, such as video streams from cameras or LiDAR-based distance calculations. Nevertheless, we can summarize that there is not yet enough computational power on mobile platforms to process complex spatial data, such as point clouds, in a fully online manner. The best industrial laser scanners generate so much data, and their processing is so complicated, that neither data analysis on the device nor instantaneous transfer to the cloud is possible. In such a scenario, the data were only collected during the acquisition phase and processed later, offline and using dedicated machines.

4. Taxonomy of High-Level Abstraction Category of Performed Process

The second taxonomy arranges the existing works according to high-level categories of performed processes. Such an arrangement allows readers to become acquainted with the general fields a particular research work was focused on.

4.1. Taxonomy Presentation

The second taxonomy groups input data types from Section 3 into buckets of the high-level abstraction category of performed processes. As articles using the same input data type completed different categories of tasks, each data structure could be assigned to more than one bucket simultaneously. Visualization of this taxonomy is presented in Figure 3. We distinguish four buckets of categories here: segmentation (without class recognition), segmentation with a simplified classification (room/corridor/unknown), segmentation with precise classification (kitchen/bathroom/etc.), and precise classification (without space segmentation). There was no additional splitting into smaller buckets as this would over-complicate the taxonomy. The separation criterion was the achieved level of semantic meaning assigned to the analyzed data.

The **Segmentation** category groups solutions that detect building structures (rooms) in the source data but do not provide any deeper semantic meaning to it. These algorithms mainly separate areas and rooms from a more extensive context but without any details of what kind of room it is. All the segmented objects represented the same class, just a "*room*". As a result, there were, e.g., *room1*, *room2*, and *room3* found. In multiple cases, the segmentation process took the form of data clustering.

The **Segmentation + Simplified Classification** category expands on plain segmentation with no room-type analysis. It represents a bucket with solutions that recognize more than one class of object from the whole set of input data but without a complete semantic understanding of them. These solutions can determine simplified class separation, e.g., into two or three classes of rooms. As a result, there were, e.g., '*room1*', '*corridor1*', '*room2*', and '*outside*' found.

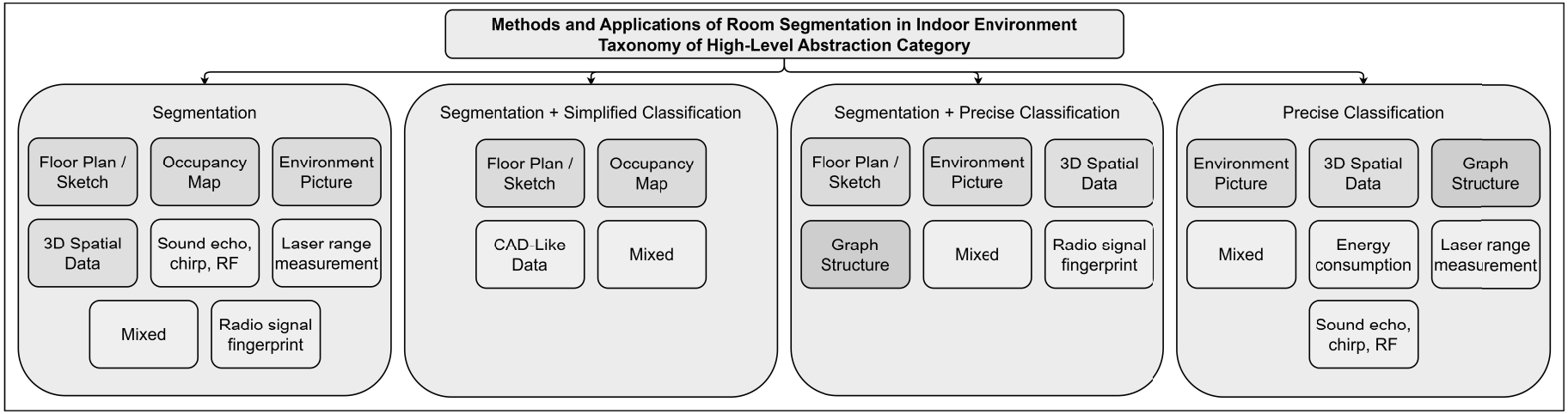

Figure 3. Diagram presenting the second constructed taxonomy—based on the high-level abstraction category of found solutions.

The **Segmentation + Precise Classification** category groups solutions that further expand the functionalities of methods from previous buckets. This time the methods can not only segment rooms but also recognize their precise function. They add semantic meaning to each segmented object from a broad set of class labels. These solutions can specify that the found room was a kitchen, living room, bedroom, bathroom, etc. The number of known classes varied between solutions.

The **Precise Classification** bucket groups methods that implement multi-class classification with a wide range of available class labels but without their segmentation from the whole set of input data. Previously described segmentation methods could detect multiple instances of room structures in the input data and assign meaning to them. Methods from this bucket are entirely focused only on the second part, i.e., just on the class assignment. An example could be the room-type classification as presented in the input picture taken with a smartphone. The class was assigned to the entire input and not to separate fragments of it. A sample result was a classification: "this is a kitchen". Table 3 presents a matrix of referenced articles, but this time organized by the processed input data type and the high-level solution category.

Table 3. Referenced papers grouped by the input data type and high-level solution category.

	High-Level Solution Category			
Data Type	**Segmentation**	**Segmentation + Simplified Classification**	**Segmentation + Precise Classification**	**Precise Classification**
Floor Plan/Sketch	[86–88,90–97,103,108,113,116,118,120–122,124,125]	[102,106,110,126]	[89,98–101,104,105,107,109,111,112,114,115,117,119,123]	-
Occupancy Map	[1,129–132,134,135,137–139,142,145,146,148]	[127,128,133,136,140,141,143,144,147]	-	-
Environment Picture	[151,153,154,170]	-	[156]	[149,150,152,155,157–169,171]
3D Spatial	[8–14,16–21,23,25–32,34–49,51–55,57–68,70,71,74–77]	-	[22,24,33,50,56,69,78]	[15,72,73]
Laser Range Measurement	[176,177]	-	-	[175,178,179]
Mixed	[181–185]	[189]	[180,187,190–195,197,198]	[186,188,196]
Radio Signal Fingerprint	[200]	-	[199]	-
Sound Echo, Chirp, RF	[204]	-	-	[201–203,205]
CAD-Like Data	-	[172]	-	-
Graph	-	-	[79,80,82,83]	[81,84,85]
Energy	-	-	-	[173,174]

4.2. Taxonomy Discussion

While preparing the second taxonomy, we focused more on the processing methods. We can draw a series of conclusions and observations thanks to the analyzed algorithms and the variety of results they generate.

Main methodology. There is a tendency to change the most popular and effective algorithm according to the processed category. Generally, analyzed solutions focused on two approaches—algorithmic processing or machine-learning methods. Pure segmentation, in which no weight was given to room-type recognition and without deeper semantic recognition of spaces, was performed mainly with repeatable algorithms. They process the readings just like any other multidimensional data, e.g., using clustering methods, variations of watershed algorithms, Generalized Voronoi Graphs, and many others. Repetitive processing of relatively simple input data yielded satisfactory results. On the other hand, solutions focused solely on classification are most effectively implemented by using complex input structures, large neural network models, and different machine-learning

approaches. Intermediate solutions combine both features and perform individual subtasks algorithmically or based on artificial intelligence and then combine the results. Other approaches, although present, were noticeably less popular or effective.

Semantic meaning. The more semantically rich the result needed to be, the more difficult it was to achieve it. Simple segmentation algorithms are successfully used in solutions where the collected information presents some abstract form of the world. An example can be the occupancy map that stores only three types of value in its pixels—empty, occupied, or unknown. Algorithmic methods perform well in such data, but it is a very limited representation of the real world, with no deeper semantic meaning that can be retrieved. When the actual space is represented with 'real input', that is, photos taken by agents, video streams, or point clouds, and the task was to "understand", the simplest algorithms cease to work. The input of even the same type is too diverse, and there is no easy way to specify definite criteria for the classification. Even humans cannot always say what determines a space membership to a specific category. We subconsciously 'know' it, but it is not enough to prepare an algorithm. In such a situation, machine-learning solutions begin to dominate. Due to their nature and simultaneous processing of millions of parameters, they can detect rules and dependencies that are invisible even to humans. In general, the more semantic meaning we want to retrieve, the more real-life input needs to be processed. For such input, the machine-learning-based solutions are expected to perform much better than other methods. Solutions that recognize objects from their point cloud representations and infer the type of the room based on its equipment are also very promising. However, such processing is an example of a very complex data structure that requires advanced and resource-intensive processing.

Machine-learning popularity. We can observe that the newer the solution, the more likely it is to use some form of artificial intelligence and machine-learning processing. This applies to both one-dimensional and multi-dimensional data inputs. Algorithmic solutions are much more vulnerable to even the smallest changes in the structure of processed data. Machine-learning-based solutions are better adjusted to nuances in the input and less likely to return completely false results. Their performance would worsen in a relatively progressive way and not with a single visible error threshold.

Hardware capabilities. We can observe an adjustment of the chosen processing method to the available hardware. This final conclusion is a continuation of the observation about the increasing complexity of input data with the upgrade of used hardware. We observe the same with the processing methods used in each distinguished category. For example, we can point to the precise classification bucket of solutions. Algorithmic processing requires fewer resources and is successful even in online solutions but requires more straightforward input data (e.g., graphic structures or radio readings). A powerful hardware platform was needed to carry out the same task category but for more complex structures (e.g., point clouds of whole buildings). However, in general, the more advanced the approach and the larger the amount of data to be processed, the more promising the results will be.

5. Taxonomy of Accomplished Tasks

The third taxonomy arranges the existing works according to the specific task for which the particular data types and the performed high-level analysis processes are used. This taxonomy shows how broad the application areas are and allows the reader to become familiar with the solutions developed for various problems of indoor environment segmentation and the discovery of its semantic meaning.

5.1. Taxonomy Presentation

The third taxonomy is focused on specifying the accomplished task of the found solutions. It is presented in Figure 4. For this analysis, we used the extended set of papers (including the survey extension). Eleven main tasks were found, with three of them requiring further separation into subtasks. We analyzed papers that led to 17 different

end results of data processing. Fourteen of them were successfully applied to 2D image processing and are marked on the diagram with a green background of the box. We describe all the papers analyzed in the rest of this section, presenting them grouped by the completed task.

To facilitate the understanding of the content, we decided to develop a single, universal example that the reader may follow and that we could reuse in each of the subsections. It was meant to serve two purposes: graphical and practical. We wanted something visually simple, schematically showing the idea, yet actually used in real applications. We decided to use a sample floor plan representing an apartment consisting of a living room, two bedrooms, a toilet, a kitchen, and a corridor. Such an example was selected for three primary reasons: (1) the leading input data type of this study is a 2D image, (2) it is a data structure that could be applied to all the main described tasks, and (3) it is subjectively easy in both preparation and later reading.

To further simplify the navigation across this document, in Table 4 we present a matrix of referenced articles, this time organized by the low-level actual application and the processed input data type. It allows for direct navigation into the subsection describing the task of interest or applications available to held data structures.

Table 4. Referenced papers grouped by their application and processed data type.

		Processed Data Type			
Task	**Section**	**3D Spacial Data**	**2D Images**	**Graph Structure**	**Feature Set**
3D Model Reconstruction	5.1.1	[9–12,17,18,20,21,23,26–30,35,36,42,45–49,51,52,54,55,57,60,61,63,64,74–76]	[94,110,111,116,154]	-	[195]
CBIR	5.1.2	-	[97,98,114]	-	-
Environment Desc. Creation	5.1.3	-	[90,95,96,99,103,107,115,119]	-	-
Floor Plan Vectorization	5.1.4	-	[104,117,120,122]	-	-
Floor Plan Predict./Gen.	5.1.5	[8,25,31,33,37,44,53,58,62,68,71,78]	[89,151,153,170]	[79,83]	[77,172,176,177,183–185,189,204]
Graph Generation	5.1.6	[13,14,19,67]	[138,139]	[80,82]	[180,190,191,194]
Room Classification	5.1.7	[15,73]	[143,149,150,152,155,157–169]	[81,84,85]	[173–175,178,179,186,188,201–203,205]
Change Detection	5.1.8	[43]	-	-	-
Map Segmentation	5.1.9	[50]	[1,82,127,128,135,136,140–142,144–147]	-	[192,193,197,198,206]
Plan Segmentation		-	[86–88,91,100–102,105,106,108,109,112,113,118,121,123–126]	-	-
Point Cloud Segmentation		[16,32,34,38,39,41,59,65,66,69,207,208]	-	-	-
VR/AR	5.1.10	[40]	-	-	-
Robot Expl./Localization		-	[129,131,132,134,156,171]	-	[182]
Path Planning		[70]	[133]	-	-
Localization		[22,56]	[93]	-	[24,181,196,199,200,209]
Map Alignment/-Matching	5.1.11	-	[187]	-	[130,148]
Plan Alignment/-Matching		-	[92]	-	-

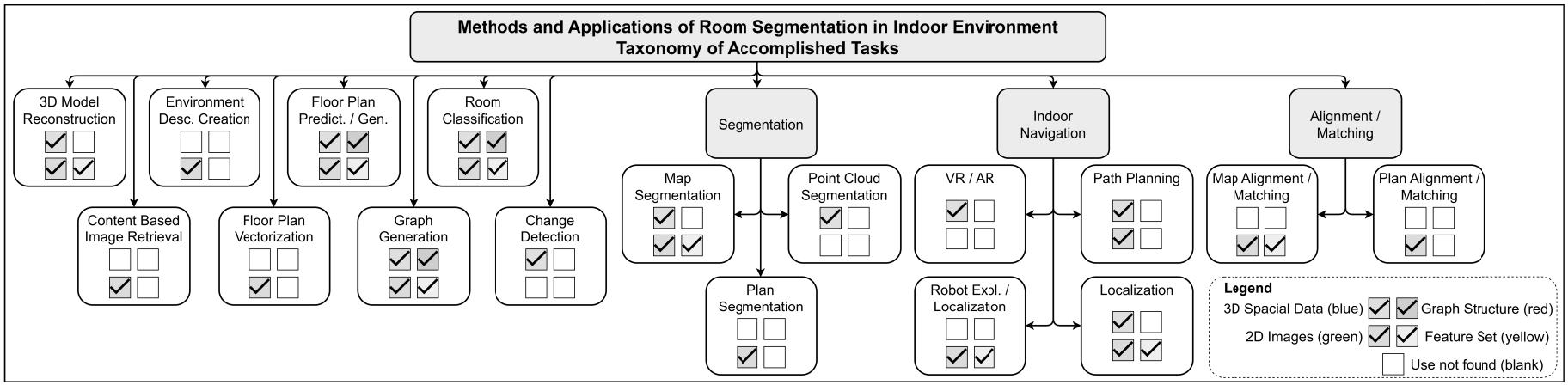

Figure 4. Diagram presenting third constructed taxonomy: accomplished task.

5.1.1. 3D Model Reconstruction

This group of solutions is focused on generating three-dimensional spatial representations of properties. Depending on the type of processed data, they could take the form of a labeled point cloud, mesh model, or be an upgraded visualization like a floor plan taken from a 2D image into a 3D model by adding the walls with an assumed height as the third dimension (case presented in Figure 5). It is one of the largest groups of articles.

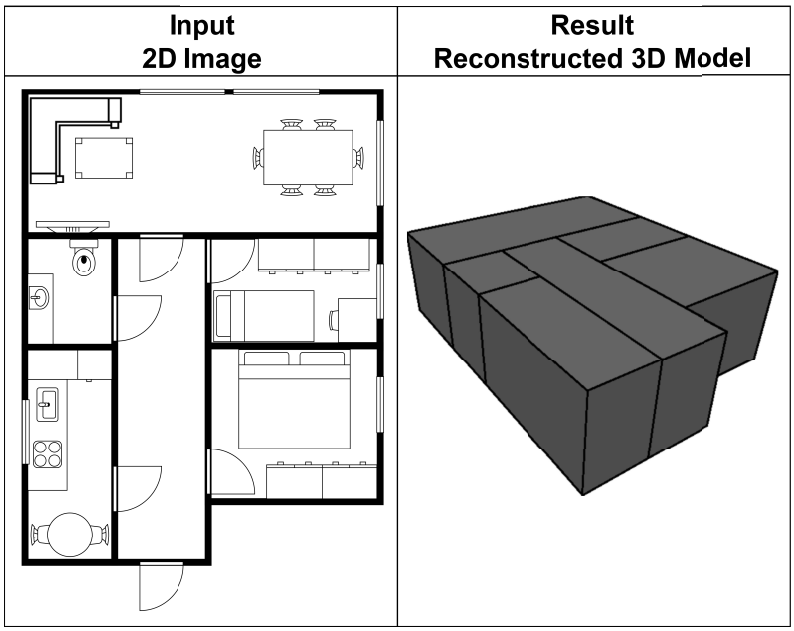

Figure 5. Visualization of the 3D Model Reconstruction task. From the input 2D image (**left**), there is a simplified 3D model generated (**right**).

Three-Dimensional Spatial Data was the most represented input data type. Thirty-four different articles implemented model reconstruction based on it. They all seem very similar, but the difference is the main focus of the processing and implementation details. The largest group of papers represented the task of generating models without additional emphasis on the specification of its concrete type. Such articles produced the least complicated representations, similar to those from CAD-like software wireframe models. Nevertheless, we can further differentiate papers by the exact type of resulting model or the way of acquiring input data. Model generation from point cloud was presented by Ochmann et al. [23], Shi et al. [48], and He et al. [63]. All the solutions generate 3D models aware of rooms and their connections described as doors or openings. Yang et al. [51] paid special attention to the structural constraints of the generated model. They used three levels of constraints, semantic, geometric, and topological, which were introduced to the process of room segmentation for their easier recognition from the point cloud. Similarly, the topological consistency of the generated model was the topic presented by Ai et al. [60]. The first specific model subtype, the mesh model, was developed in the works of Turner et al. [17] and Turner and Zakhor [18]. In this representation, a 3D model consists of vertices, edges, and faces that build a mesh of polygons to reconstruct the scanned object. Another subtype of the model was a watertight model. A sample solution generating such was presented by Cai and Fan [61]. A model is classified as watertight when each triangle building the mesh of the model has exactly two neighbors and leads to a structure with no holes. The same resulting model was found in the solutions of Wang et al. [29] and Nikoohemat et al. [45], with the distinctive aspect being the type of laser scanner—in both articles, a Mobile Laser Scanner (MLS) was used to capture the initial point cloud in a real-time manner. On the market, there are multiple types of scanners available. They all generate the same output, a point cloud, but with different operating concepts. The Standard Static Laser Scanner (SLS) is a ground-based device,

typically mounted on a tripod placed on a flat surface and in a reference point. It scans the environment with high precision, generating a point cloud including ceilings and floors. It can be moved, but only discretely, from place to place. The MLS changes the approach—it is mounted on a moving vehicle instead. It can be a cleaning robot, a trolley, or a flying drone. This way, the scanning is a relatively continuous operation. SLSs were also used in the papers of Nikoohemat et al. [28] and Xie and Wang [30]. They utilized a specific feature of the scanning process, the trajectory of the scanner. In this way, the point cloud was supported by information regarding consecutive scanner positions in the test environment. A specific subtype of the MLS is a Backpack Laser Scanner (BLS), a version in which a human operator carries the device as a backpack. Such a solution was found in the publication of Ryu et al. [57]. They successfully used indoor point cloud data gathered by a BLS to generate a geometric representation of found rooms, utilizing a two-step procedure based on the data refinement and later processing. Murali et al. [27] and Jung et al. [35] targeted the creation of 'as-built' BIM from 3D point clouds. This means that they processed data for buildings that did not have a proper 'as-designed' BIM created during the construction. A similar application to the generated model was described by Ochmann et al. [46], but as just one of many possible applications of volumetric models. The BIM generation was also further investigated by Otero et al. [55]. Their solution focused on the generated output files. They processed LiDAR data, prepared a BIM model, and saved it to a gbXML file representing an energy-efficiency-focused structure, which could be used in further energy studies. Mura et al. [9,12] and Tang et al. [49] discussed model generation in highly cluttered input data scenarios with severe occlusion. They presented scenarios of, e.g., wall parts placed behind obstacles, large-scale scan artifacts, or outlier points in the acquired point cloud. Many publications implicitly or explicitly work with the Manhattan-world assumption [210] (highly regular environment). It can be used in building structure simplification. One example is the article of Previtali et al. [47]. They discussed indoor layout regularization and proposed a method for room hierarchy estimation, assuming their high similarity and regular shapes. On the other hand, the work of Yang et al. [52] explicitly focuses on processing multi-room building structures with curved walls, which is not a typical layout. They proposed a solution using horizontal slicing of the acquired point cloud, which allowed them to easily ignore the furniture present in the building and simplified the detection of straight and curved walls. Object detection and classification can also be separate topics. Armeni et al. [20] proposed segmenting the point cloud into "disjoint spaces", representing objects found in the building, such as chairs, sofas, or tables. They detected gaps in the density of recorded points and deduced the potential edges of separated objects. Maurović et al. [11] and Manfredi et al. [21] presented solutions to a highly different task. Both proposed methods for autonomous robot exploration of an unknown environment. The main problem was to somehow plan an efficient path and set of scanning points so that a robot could acquire the most precise model of the building. Similarly, a mobile robot was used by Bormann et al. [10] to create a thermal model of the explored building. The achieved solution augmented the captured point cloud with colors that indicated the temperature of each of the points. By grouping them, it was possible to reveal potential heat sources. In their implementation, they used the idea of space voxelization. A voxel can be described as being like a pixel of an image but in a three-dimensional world. It is the smallest part of a regular, grid-adjusted volume. The topic of voxelization was also found in the work of Hübner et al. [64]. For an indoor environment, their suggested solution for model reconstruction is voxelization with a 5 cm resolution. This way, the generated grid was dense enough to successfully reconstruct even complex room layouts, like those with curved walls, complicated openings, or a height of more than one floor. The generation of multistory building models was found as the topic of papers by Macher et al. [26], Li et al. [36], and Cui et al. [42]. They all process massive amounts of data, as each consecutive level of scanned buildings multiplied the number of acquired points. Implemented approaches are quite similar. A typical process segments the whole building into separated floors first and into distinctive rooms second. In all found implementations,

the floor clustering was based on the density of points across the height axis of the models. It was effective but introduced a specific limitation—buildings with more complex structures of ceilings were expected to be processed incorrectly, e.g., lofts or rooms with entresol. This problem was partially solved in the work of Nikoohemat et al. [54], in which 3D models were used for disaster management and emergency exit path planning. Trajectory, captured by an MLS during scanning, was used in the later processing. The proposed method segmented the trajectory into separated levels and combined the result with scanner position timestamps, allowing the associated parts of the point cloud to be separated as well. All data gathered from mobile devices were processed in three further articles. Franz et al. [76] implemented a solution for collaborative scene modeling using multiple mobile devices compatible with Google Tango technology. One of the proposed applications was to quickly and easily scan a crime scene for further offline investigation of possible clues. A smartphone was also used in the article by Liu et al. [74]. The recorded magnetometer and gyroscope measurements provided values for the azimuth and pitch angles of the device, which allowed researchers to reconstruct the environment with the use of geometry algebra. Weinmann et al. [75] used an even more sophisticated device. They conducted a survey focused on the possible applications of the Microsoft HoloLens. The device itself is a head-worn mixed reality device consisting of glasses with an integrated IMU, multiple color cameras, depth cameras, and many other sensors. The survey discusses diverse applications of the generated models and acquired measurements: indoor localization, spatial mapping, scene model reconstruction, and semantic segmentation.

Two-Dimensional Images and solutions generating 3D models from them were the second most numerous. We found five such articles. Four of them involved generating the model from scans of floor plans. As expected, they struggled with specifying the third dimension, namely the height of the rooms. Gimenez et al. [116] let the user specify this dimension and assumed all components to be equal in height (walls and openings). The work of Dodge et al. [94] presented a neural network approach to floor plan parsing, but with the necessary model height provided by the authors themselves. The same was found in the work of Park and Kim [110], where the height was set to 240 cm. The first paper that actually calculated the height dimension was the one by Lv et al. [111]. They implemented a complex scale calculation procedure based on character recognition in the scanned images. The proposed method first searched for available scale information and (if found) calculated the rest of the model parameters. If the scale was not found, they assumed the median door size to be equal to 90 cm and used this value to calculate the global scale of the floor plan. Basing the model sizing on the proper scale calculations was also found in the work of Pintore et al. [154]. They proposed a method to process panorama images of the indoor environment. The lack of scaling information was solved by calculating the scanner distance from the floor and the ceiling. This way, a global scale was found and used for the proper 3D model scaling.

The **Feature Set** was the least represented. Based on a feature set of visual cues about the environment, the work of Kostavelis and Gasteratos [195] proposed a complex spatial model, extended with a scanner trajectory and room-type probabilities presentation in a place-by-place manner, that resulted in a 3D metric map of the explored space.

5.1.2. Content-Based Image Retrieval (CBIR)

Generally speaking, CBIR is a search method in which the query is formulated using a sample image instead of keywords. For this group, implemented solutions were based only on 2D image processing. A sample picture was provided as input data. The solution searched for other floor plans, classified as similar, and returned a set of possibly closely matched results. An example that returns the TOP3 most similar results is presented in Figure 6. There were two types of input images found. In the works of Sharma and Chattopadhyay [97] and Yamasaki et al. [98], the query was formulated with the use of a sample floor plan. Ahmed et al. [114] proposed a solution that works with a user-provided

freehand sketch. All these systems analyzed the input image, processed it, and utilized a generated graph representation for similarity search across the database of other floor plans.

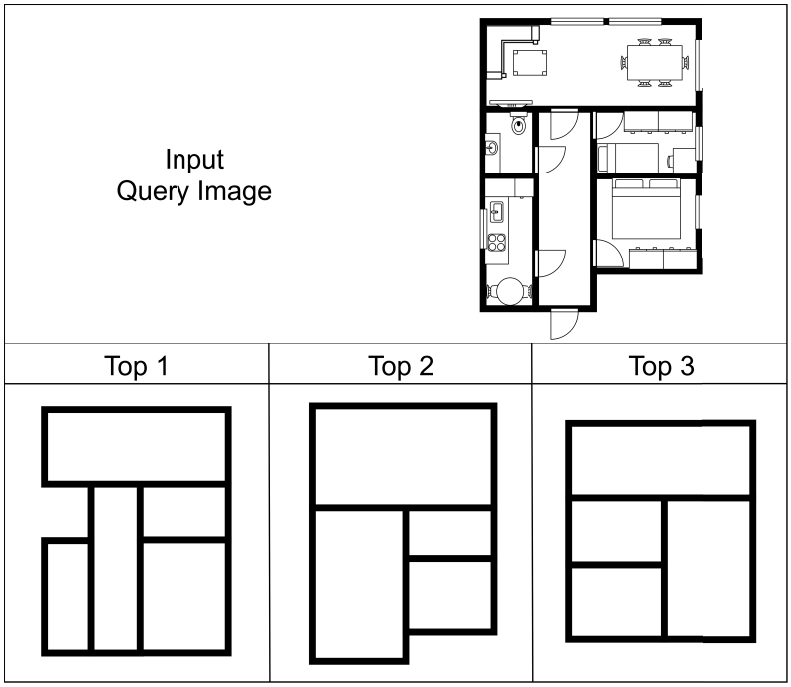

Figure 6. Visualization of the Content-Based Image Retrieval (CBIR) task. A query picture (**top**) is used to search a database for TOP3 most similar floor plans (**bottom**).

5.1.3. Environment Description Creation

This task grouped solutions that resulted in easily readable extended text descriptions of the analyzed buildings. Their main targeted usage was dedicated to supporting visually impaired people, in independent environment exploration and path planning or building structure understanding before an actual visit. The sample result is presented in Figure 7. Goncu et al. [90] created a system that converted the given floor plan into an interactive graphic displayed on the touch screen of a tablet device. A visually impaired person could tap the screen to hear a textual description or other sound indicators of touched building components. The achieved approach was further extended by Madugalla et al. [95,103]. The authors upgraded the proposed solution to a complex system providing the user with text descriptions of whole floors, separated rooms, and many other features, like an editing tool for a sighted person to help manually correct possible recognition errors. Paladugu et al. [115] presented a web browser-based system that generates extended textual descriptions of floor plans found as a result of the search phrase provided by the user. Similar whole-floor descriptions were the results of consecutive solutions developed by Goyal et al. [96,99,107,119]. The first one [96] proposed a framework called "Plan2Text" that allowed the processing of the input floor plan and generated a textual description from a first-person perspective. "SUGAMAN", presented later in [99], expanded the research and added, e.g., navigation features that described to the user how to plan a walk from one room to another. The third paper [119] introduced and tested a building plan repository called "BRIDGE", which contained over 13,000 images of floor plans and the descriptions generated for them. The last article [107] extended the plan repository project even further and presented machine-learning models for simultaneously extracting visual and textual features.

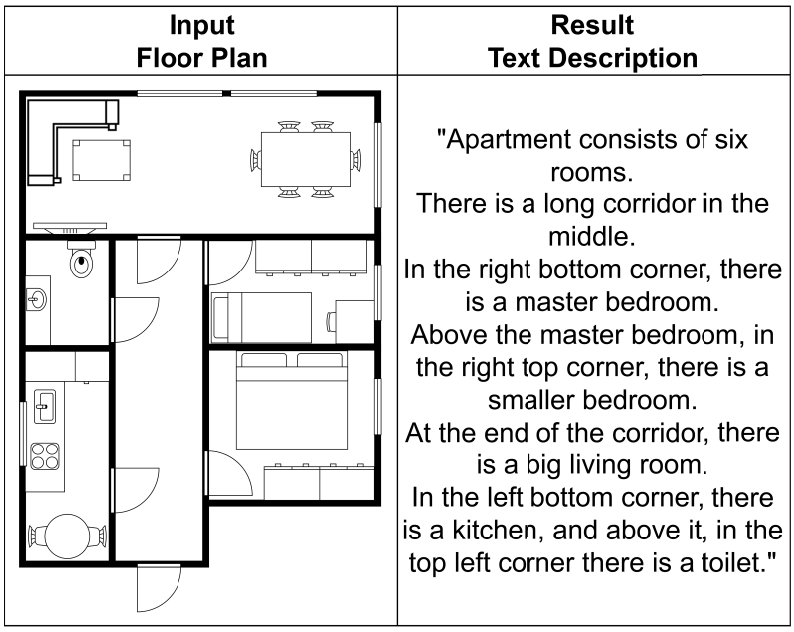

Figure 7. Visualization of the Description Generation task. An input floor plan (**left**) is analyzed and converted to a set of sentences describing its components and their location (**right**).

5.1.4. Floor Plan Vectorization

Vectorization was a task performed only on 2D images of architectural floor plans. As the input data in such scenarios consists of a rasterized image, they are hardly scalable, with processing possibilities limited to computer vision techniques. Vectorization of such a plan changes its representation from a matrix of pixels into a structured file of vector descriptions. The processed image should preserve quality (even if scaled up) and simplify further data processing. A sample result is presented in Figure 8. Chronologically, the first article found was that of Liu et al. [117]. The "Raster-to-Vector" approach they proposed grounded the processing in a convolutional neural network and integer programming techniques. The later articles were similar and differed mainly in the proposed algorithm and the learning method. Jang et al. [120] used a combination of many different neural networks, such as Global Convolutional Network (GCN), Pyramid Scene Parsing Network (PSPNet), DarkNet53, and Deeplab v3. Surikov et al. [122] researched a processing method integrating UNet, Faster-RCNN, statistical component analysis, and the Ramer–Douglas–Peucker algorithm. The last article found, published by Dong et al. [104], relied on a newly designed version of a Generative Adversarial Network (GAN) called EdgeGAN, which was found to process data significantly faster than other compared solutions.

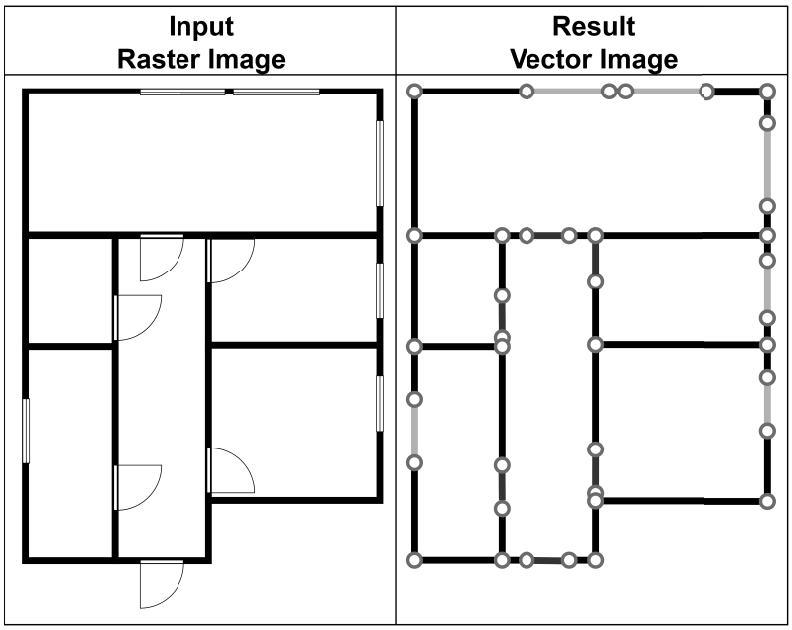

Figure 8. Visualization of the Floor Plan Vectorization task. An input raster image (**left**) is converted to a vector representation (**right**). Converted walls are marked with black lines, doors with blue lines, windows with green lines and joints with red circles.

5.1.5. Floor Plan Prediction/Generation

The task of floor plan deduction from gathered input data was found in 27 articles spread over all the analyzed buckets of data types. As a visualization, in Figure 9 we present the generation of floor plans based on the processing of the recorded trajectory of a user. The first articles found were those by Liu et al. [79], with their solution for the generation of interior layouts for precast concrete-based buildings, and Santos et al. [8], who contributed by proposing an approach for translating data from a grid map into a topological one. In 2014, Gao et al. published research on mobile device-based floor plan generation [77]. They proposed to use a crowd-sensed combination of vision and inertial data from mobile phones. Turner and Zakhor [176] published the first version of floor plan generation from laser range data and continued their work a year later [177]. The presented solution was based on "wall-samples" processing and worked with even multistory buildings. Camozzato et al. [89] switched to working with freehand sketches. They procedurally generated a floor plan from a user-provided paper drawing. An article from 2016 by Pintore et al. [170] discussed indoor map creation from omnidirectional images. They captured one image per room and processed them to generate a map of the whole floor. Their later work [153] presented an extension that even allowed them to create a complete 3D floor plan. A much different approach was introduced by Loch-Dehbi et al. and Dehbi et al. [183–185]. They proposed basing the floor plan processing on a set of sparse observations, assumed constraints, estimated parameters, and stochastic reasoning. Fleer [151] constructed a system for a mobile floor-cleaning robot, which generated a segmented floor plan based on laser readings and a hemispherical camera directed at the ceiling. Jung et al. [25] made a step towards classical techniques, but with modern data—they applied morphological processing to the captured point cloud of the environment. Ambruş et al. [31] introduced a method for room segmentation from a raw, unstructured 3D point cloud, later used in 2018 by Brucker et al. [33]. They applied it to a semantic room labeling task based on RBGD images. Liu et al. used their previous work on floor plan vectorization and proposed a framework for plan reconstruction from 3D scans [78]. Chen et al. [71] evolved the solution to an automated approach for CAD-like plan generation. A similar evolution can be seen in the project of Magri and Fusiello [37] that proposed reconstructing the walls of analyzed buildings from their scanned point clouds, and a later variation of the method presented by Maset et al. [44] focused on better results with lower user dependence. He et al. [53]

proposed a solution for furniture-free room mapping using a mobile robot equipped with multiple 3D LiDAR. They combined the source data acquired from vertical and horizontal laser scans. Zhou et al. [189] presented a method of creating floor plans using smartphone sensors and recorded user traces. In a 2020 article, Nauata et al. [83] proposed a solution for house layout generation according to an input bubble diagram (graph) of expected rooms and connections between them. Users could specify types of rooms as well as connections between them and obtain a set of proposed house plans. Phalak et al. [58] presented a novel technique for floor plan generation from point clouds using deep neural networks trained on purely synthetic data. The publication of Resuli et al. [204] introduced a solution relying on input data unseen before—radio frequency readings from sensors. They created a system of transmitters and radars that allowed room structure reconstruction thanks to the signal reflections and their processing. Floor plan prediction task analysis was concluded with articles by Fang et al. [62], Simonsen et al. [172], and Cai et al. [68]. The first one [62] presented a space partitioning approach to floor plan generation from the point cloud, similar to and compared with the already mentioned work of Chen et al. [71]. The second article [172] focused on processing CAD floor plans exported from architectural software as DXF files. It was desired to skip the rasterization part of floor plan generation and work with a graph structures recognized from the CAD primitives. The last article [68] proposed to process an input point cloud and generate an accurate floor plan using geometric priors—instead of relying solely on the density of points to recognize the building structure, they proposed combining them with indoor area recognition, normal information, and other geometric analysis.

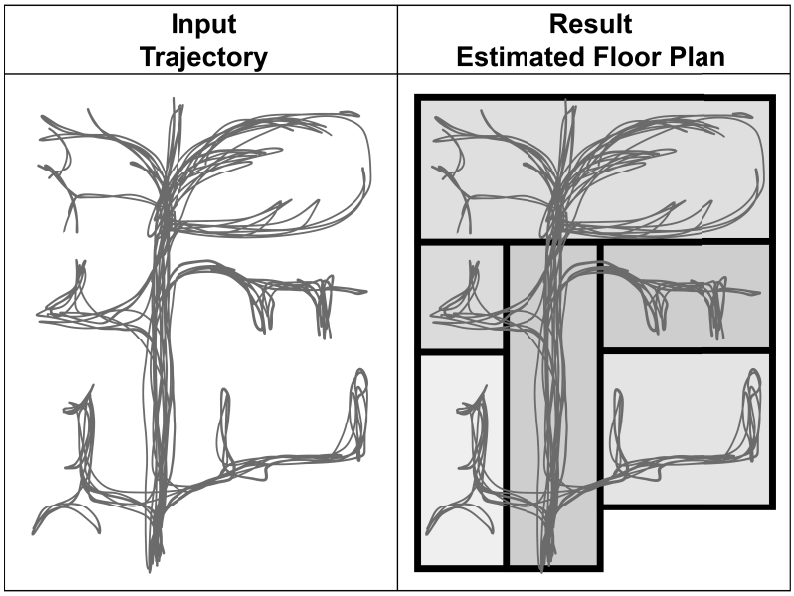

Figure 9. Visualization of the Floor Plan Generation task. Input data (recorded user trajectory, **left**) is used to estimate a possible floor plan—relying on the paths' density and their clustering (**right**).

5.1.6. Graph Generation

These methods took advantage of the graph theory applied to input data structures. They modeled the environment into graph structures and were able to predict the room segmentation or classification with graph processing techniques, like node clustering or adjacency statistics. An example is presented in Figure 10. Graph generation was found in 12 articles. Pronobis and Jensfelt [190] and Kostavelis et al. [191,194] used them to process the trajectory of a robot in the environment to add semantic meaning to visited places. Similarly, Ikehata et al. [19] used a motorized tripod for RGBD panoramic picture acquisition. They processed the data to generate a graph of the environment structure for the room segmentation process. Wu et al. [180] equipped their robot with vision sensors prepared for artificial label recognition and used it to create a multilevel graph,

representing the global topology of the building, and to detect connections between small items in the same local area. Ochmann et al. in [13,14] focused on graph generation for topological building representation and its structure decomposition from a pure 3D point cloud. Luperto et al. in [80] and later Luperto and Amigoni in [82] took the input structure of the undirected graph representing the indoor environment and discussed how to analyze the graph to segment rooms and predict the global structure of a building. Hou et al. [139] used a Voronoi diagram method for automatic graph creation from a 2D grid map. The proposed "Area Graph" was used as a segmentation technique. Schwertfeger and Yu [138] presented a similar solution, but added Alpha Shape-based processing to the initial diagram and used it for topology graph improvement. Another complex solution was presented by Yang et al. [67]. They combined a neural network with a distance transform algorithm to reconstruct the topological meaning of indoor space partitions from input RGBD pictures and their colorful point cloud representations.

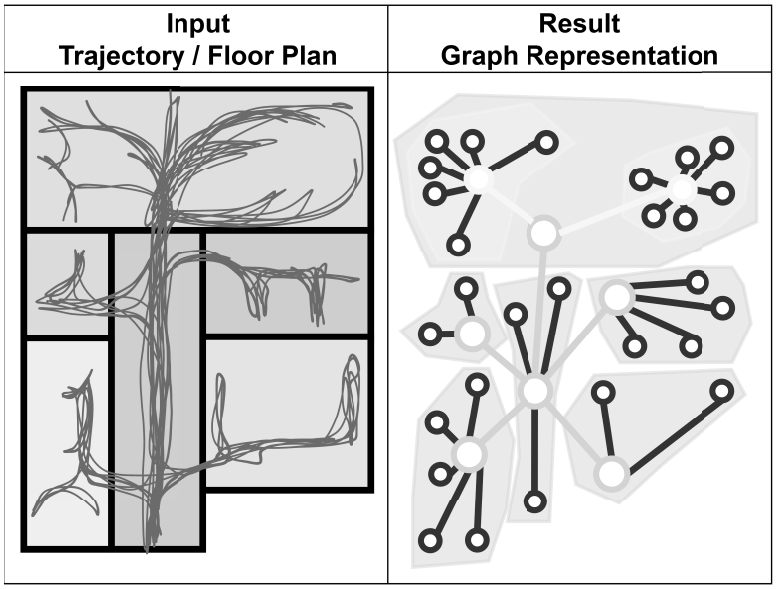

Figure 10. Visualization of the Graph Generation task. Input data (recorded user trajectory with some understanding of the building, (**left**)) is used to construct a graph structure (**right**) representing rooms and areas as graph nodes, and their adjacency as graph edges.

5.1.7. Room Classification

In the room classification bucket, the analyzed solutions predicted the kind of room the given input represented. Again, this task covered all distinguished input data types and delivered numerous articles. Recognition of the room type can be achieved using an image of the indoor environment, calculated energy measurements, or even the recorded sound echo. The presented example (Figure 11) analyzed a raster image, searched for characteristic elements, and deducted the placement and types of rooms.

Indoor pictures, as a sub-bucket of data types, was the largest one and was dominated by machine-learning approaches. Young et al. [152] presented a CNN-based solution for a situated robot's recognition of the environment around it. They combined picture understanding with a web mining approach. Room type was deducted from a set of object pictures and their co-occurrence in the web resources. Similarly, Ursic et al. [150] extracted parts of the pictures and analyzed them as regions of interest. Instead of a holistic categorization, they classified each part and deduced the end result, which appeared to be robust to image distortions. Othman and Rad [155] compared multiple CNN architectures in the room classification task for a humanoid robot exploring the environment. In their work, a custom multi-binary classifier was proposed and combined with a CNN for better performance. The article of Rubio et al. [149] was the only one avoiding a network of convolutions—they proposed and compared systems of Support Vector Machines (SVMs)

and Bayesian Network Classifiers (BNC). The reported results indicated that with the proper processing, BNCs outperformed SVM classifiers. The survey extension returned even more results than the main flow; we found 13 further articles. Ranganathan [159], in one of the oldest articles analyzed, processed images coming from a stream of measurements, like a moving robot capturing a video. Changepoint detection was performed sequentially in consecutive frames and used to classify entire video segments. One article analyzing the robot's camera orientation was the one by Erkent and Bozma [158]. In their article, they proposed a "Bubble Space" that represented the nodes of topological maps of an environment as abstract representations of the robot base and its viewing directions. Mozos et al. [162] used a readily available Kinect camera and presented its use on a mobile service robot. Combining the depth and grayscale images made it possible to gather data in which patterns could be detected. Histograms of patterns were then simplified into a feature vector used in a supervised indoor space classification method. A visible direction in the room classification techniques was to analyze not the whole image at once but part by part. Parizi et al. [160] presented a solution based on picture regions. They assumed that a picture of a specific class of objects has a set of regions expected to contain a particular data type. An example of this might be the blue sky in the top part of the beach image. With the use of a reconfigurable version of Bag of Words (BoW), their model learns interesting regions of images and uses them to classify the observed scene. Similarly, Sadeghi and Tappen [157] presented a solution based on the idea of spatial pyramid processing. Proposed Latent Pyramidal Regions (LPRs) were used to describe the characteristics of parts of the processed scene image. The later searching through of all possible subwindows of the input image allowed using learned representations for scene classification. A year later, Juneja et al. [163] proposed a solution combining the BoW technique with a novel approach called Bag of Parts (BoP). In their work, attention is placed on the part-learning process. Detection of important parts of the picture is achieved automatically and is used without extensive user involvement. Hierarchical dependencies between parts of the pictures were also analyzed by Sadovnik and Chen [161]. They assumed each scene was built using some substructures, separated objects, and connections between them. In their work, a Minimum Description Length (MDL) principle was used to deduce the scene type from a set of its building blocks. In this way, a scene could be classified as a kitchen if its recognized components were chairs next to tables and cabinets under the stove. Margolin et al. [164] proposed a novel local descriptor for low-level image representation, significantly different from the solutions mentioned above, i.e., Oriented Texture Curves (OTC). In OTC processing, each patch of the image was analyzed for its texture and color variation along different orientations. Constructed curves were processed into a single descriptor that was used for scene classification. A lower level of processing was proposed by Zuo et al. [165]. Instead of processing subparts, they prepared a pixel-level-based solution named Discriminative and Shareable Feature Learning (DSFL). Their main idea was to generate a bank of filters that represent common features of images and encode information from all raw pixel values. In a typical analyzed picture, only a specific subset of all the features was present, which allowed for scene classification. The opposite approach, the holistic analysis of pictures, was presented by Dixit et al. [169]. They presented a solution modeling the scene as a bag of object semantics achieved by a CNN trained on the ImageNET dataset. A constructed Fisher Vector (FV) summarizing object descriptors was reported to present better performance than an FV of low-level features. Similarly, the use of CNNs was further researched by Jie and Yan [166]. Aware of CNN's sensitivity to image distortions, they focused on proposing a multilevel processing pipeline that fine-tuned the networks in a cascade approach, relying on differently scaled input images. Cross-level processing has been reported to improve the robustness to scale transformation of pictures. The lack of adequate datasets dedicated to scene recognition was addressed by Zhou et al. [168]. They presented the Places dataset consisting of over 7 million labeled pictures of scenes. While presenting the dataset, they also prepared a solution called Places-CNN and reported competitive performance. On the other hand, Mesnil et al. [167] presented a solution focused

on unsupervised learning. Based on the image detectors trained on popular image datasets, they studied the possibility of higher-level feature processing on top of the returned object representations.

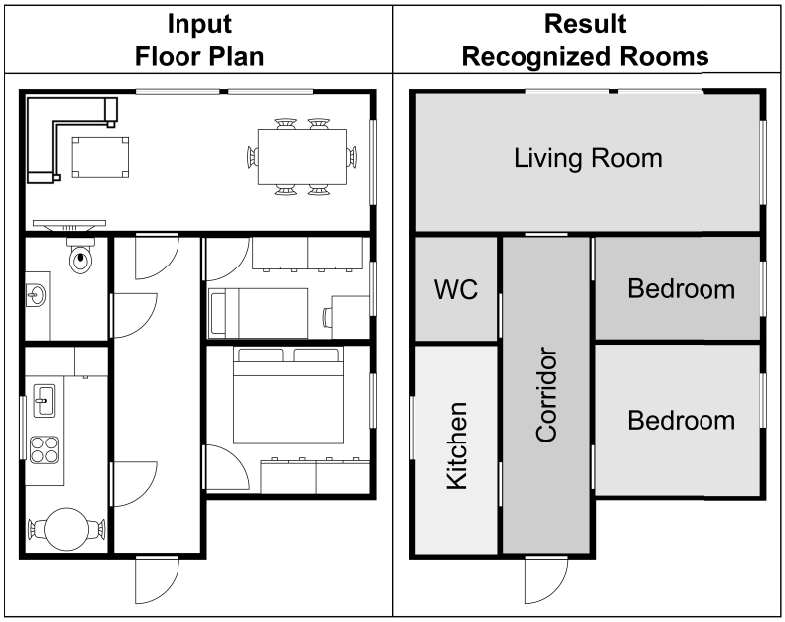

Figure 11. Visualization of the Room Classification task. A sample input floor plan (**left**) is analyzed and searched for room areas. Each area is classified and marked with a room-type label (**right**).

Graph data were processed by three of the found papers. The oldest one, by Luperto et al. [81], was from 2017. They analyzed the graph structure of a building using Statistical Relational Learning (SRL). The KLog [211] algorithm was reported to be successfully applied to room and building type classification. Paudel et al. [84] presented graph processing techniques with an extended comparison of various Graph Neural Networks (GNNs) analyzing floor plan graphs. GraphSAGE [212] and Topology Adaptive Graph Convolutional Network (TAGCN) [213] were reported to outperform solutions using the Graph Convolutional Network (GCN) [214], Graph Attention Network (GAT) [215], and Multi-Layer Perceptron (MLP). The latest research by Wang et al. [85] continued the idea and proposed an improved version of the GraphSAGE algorithm called SAGE-E. It used graph node features and edge features, allowing it to outperform other solutions.

Three-dimensional spatial data were the least represented. Swadzba and Wachsmuth [15] investigated the possibility of point cloud processing using a novel spatial feature vector describing the properties of the environment. In their scenario, a room type was recognized from a limited number of inputs—a view recorded only by a depth-aware camera situated at the door frame while entering the room. Real-time scanning was the topic analyzed by Matez-Bandera et al. [73]. They focused on a moving robot equipped with adjustable cameras. The main objective was to optimize the sensor line of sight and maximize the amount of collected data. They combined the processing of gathered knowledge, like robot paths, detected objects, and taken pictures, for quick and accurate place categorization.

Sound echoes, chirping, and radio frequencies were the input data types for which we found four articles. The oldest one, by Peters et al. [201] from 2012, presented a system that identified the room type in an audio sample or video recording. Each sample included acoustic features and came from a musical signal or a speech recording. The proposed solution was based on a Gaussian Mixture Model (GMM). In 2018, Song et al. [202] published an article exploring the usage of a smartphone's loudspeaker in indoor localization tasks. The presented solution classified the room by analyzing the echoes recorded in response to a generated inaudible chirp. The best accuracy was reported while using a two-layer CNN processing spectrogram of recorded echoes. A similar solution was

presented by Au-Yeung et al. [203] in 2020. They proposed using a smartphone with a dedicated application to capture room acoustic profiles. The app generated and transmitted acoustic signals that reflected from objects in a room, creating a unique profile of a room. Learned features were later used by a classifier for room recognition. The newest work by Dziwis et al. [205] was focused on Augmented Reality (AR) applications. In their research, the authors analyzed the task of virtual sound generation. They proposed combining a machine-learning system based on a CNN with Binaural Room Impulse Responses (BRIRs) to provide possibly realistic room reverberations in an augmented environment. Laser range readings represented a specific type of feature set. Known for environment picture processing, the hierarchical approach was presented in the research of Uršič et al. [175,178]. In their first publication, an algorithm called Spatial Hierarchy of Parts (sHoP) was used to construct a representation of the environment from laser scans. Initial measurements were transformed into images, which were hard to understand, even for a human. For room classification, the authors used a Histogram of Compositions (HoC), and the whole solution consisted of three processing layers. In the later article, the idea was further developed, and on top of the hierarchical processing was the Multi-Category-Affinity-Based-Exemplars algorithm added as a fourth layer, reporting better performance. In 2018, He et al. [179] presented a machine-learning approach. In their work, a laser range scan of a mobile robot was formatted into polar images, which were further processed by a technique named Local Receptive Field-Based Extreme Learning Machine (ELM-LRF). It effectively merged the ideas of splitting image processing into its sub-parts (receptive fields) and using an extreme learning machine (a neural-network machine with just one hidden layer). The Mixed input was represented by two articles. In 2018, Dehbi et al. [186] proposed using sparse observations of the environment and openly accessible data in combination with a Bayesian Classifier. They could infer floor room shapes and their functions through the room area and its direction (orientation). Hu et al. [188] used a similar yet extended information set. They used training data consisting of area, length, width, and room type with a created set of grammar rules and a geometric map of the floor to estimate the semantic meaning of the rooms. Their solution grounded the processing in the Bayesian inference. It was utilized in the calculations of room-type probabilities as well as in the creation of a parse forest structure. Energy consumption models were used in two articles. In 2015, Wei et al. [173] published a paper researching the use of a neural network in room classification for office buildings. Their solution analyzed electricity consumption measurements from three main categories of devices: sockets, lights, and air-conditioners. The presented Echo State Network (ESN) was a recurrent neural network reported to achieve good effects in practical applications. The work was later continued and improved by Shi et al. [174]. The extended version combined two layers of ESNs—a set of three networks for each of the modeled consumption categories and a fourth one, on top of the others, combining previous results into a single classification. Grid map was found to be processed by only one additional article. The solution of Shi et al. [143] grounded the map analysis in a semi-supervised learning algorithm. A combination of Support Vector Machine (SVM), Conditional Random Fields (CRFs), and a Generalized Voronoi Graph (GVG) resulted in three-class office room-type recognition.

5.1.8. Change Detection

This was a very specific task achieved for the environment monitoring in the case of building structure updates. Its main added value was an increasing analysis of how a building has evolved over time. This task was found in the work of Koeva et al. [43]. A sample visualization is presented in Figure 12.

Figure 12. Visualization of the Change Detection task. Initial and updated images (**top**) are analyzed and searched for differences between them. Detected mismatched elements are marked in red (**bottom**).

5.1.9. Segmentation

Segmentation represented the first group, which needed further separation. We specified three subgroups depending on the processed data type. They were all focused on plain room segmentation but within floor plans, occupancy grids/maps, and point clouds, respectively. For floor plans and occupancy maps, as they are two-dimensional inputs, the results were in the form of room instances segmented on the pixel level of the picture or polygons stretched on sets of points of the analyzed map. This sample scenario is visualized in Figure 13. The point cloud subgroup, as a three-dimensional representation of the whole environment, represented the spatial data. Although more complicated, the segmentation process was quite similar, with the main difference being the dimensionality of the segmented spaces—room instances were described at the point level as clustered sets of points in the point cloud or closed volumetric cuboids.

Floor plan analysis was the most represented task. In the found solutions, multiple articles represented the same team of researchers improving presented works, comparing the results, or trying to solve the problem in a different way. The first group of publications were Ahmed et al. [86,112] and de las Heras et al. [87,91]. Their work initially segmented rooms in a plan with the use of a text segmentation method, followed by wall and object detection. The results were also supported by an Optical Character Recognition (OCR) framework for room label recognition. The later solution introduced patch-based processing, which categorized pixels into one of three classes and constructed a graph of found structures. The analysis of the graph allowed for room segmentation via path planning. The idea was further developed from another perspective in the last article, presenting an Attributed Graph Grammar (AGG) analyzed with a greedy search algorithm for the best possible room segmentation representation. The second solution that improved over the years was the one by Liu et al. [113] and Liu and von Wichert [88,124]. The initial work presented a probabilistic solution to extracting semantic models from a grid map generated by a Simultaneous Localization And Mapping (SLAM) process. The proposed algorithm, based on the Markov Chain Monte Carlo (MCMC) sampling procedure, searched all possible world models for the best fit (in the biggest likelihood sense). On such a basis, the authors proposed an extended solution that combined the already tested MCMC procedure

with the idea of a Markov Logic Network (MLN). These were used to prepare a set of rules describing the task-specific context and were reported to improve the performance achieved. Two different approaches to floor plan analysis and environment model generation were presented by Kim et al. [121,125]. The first article focused on the problem of non-existing standardization in the floor plan format. They stated that research in this field is closely bounded to the processed dataset, and the achieved results are hardly transferable. To solve the issue, they proposed neural network architecture, implementing a style transfer task on the input images based on the idea of Generative Adversarial Networks (GANs). The vectorization process performed on unified floor plan formats was reported to work correctly, even with new, previously unseen forms. The second article dealt with large-scale complex buildings. A patch-based approach was introduced using a CNN architecture to segment the plan and generate a vectorized image that was later used to create a 3D model. The article of Mewada et al. [102] used a different approach. It was the first in this task to use the Alfa Shape algorithm to extract rooms from a floor plan. In the detected rooms, their actual areas were calculated and used in a linear regression model that deducts the type of room. All other published works found used some sort of neural network. The architecture of Convolutional Neural Networks (CNNs) was widely found. In 2019, Kalervo et al. [100] focused on the problem of the low availability of annotated floor plans that could be used by researchers. The presented dataset of images, called CubiCasa5k, made 5000 large-scale floor plans available for further analysis. While introducing the dataset, they also showed a CNN for a baseline results comparison. In the same year, Sandelin and Sjöberg [118] presented research on the use of Mask R-CNN in the room feature segmentation. Although they were aware of CubiCasa5k's publication, it was too late to use it. The research was concluded based on 700 other annotated images, documenting the applicability of Mask R-CNN to the problem. Both articles were referenced by Murugan et al. in [109]. Their publication presented an improved approach based on a combination of Cascade R-CNN and Keypoint R-CNN. The final floor plan parser was reported to achieve better performance on the CubiCasa5k dataset than the initial author's baseline results. A specific variation of implemented tasks was presented by Gan et al. [106]. In their solution, only bedrooms were recognized and counted in the analyzed floor plans. The described algorithm had extensive initial image processing (with Otsu thresholding, morphological operations, Hough transformation, and the application of the FAST algorithm), with the final bedroom detection achieved with a CNN. Five architectures were tested, and GoogLeNet was reported as the best-performing one. Zeng et al. [101] used the spatial context of elements found in the floor plan. The prepared multitask neural network used convolutional layers to carry out subtasks of room-type and room-boundary predictions. The later combination of results with spatial context awareness of structures hierarchy improved the achieved performance. Zhang et al. [123] tried to detect not only rectangular but also circular rooms. The presented Adversarial Network was constructed using direction-aware, additive convolutional kernels, which reported improving performance in irregular-shaped room detection. Foroughi et al. [105] presented an architecture called MapSegNet—a variation of the typical encoder–decoder solution. Using the technique of skip-connections between different layers of the network, they reported comparable results with higher accuracy or lower computation costs compared to other architectures. Lu et al. [108] focused their research on rural areas and presented a network processing the floor plans of 800 residences from the China region. As object recognition and text detection are typically two different tasks, their architecture primarily tried to reduce the time needed to train two different networks and combine them into one bigger system. Song and Yu [126] made use of a Graph Neural Network. In their algorithm, an input image was first pre-processed (text removal and binarization), then vectorized, and finally represented as a Region Adjacency Graph (RAG). In such a graph, a GNN was used to find the semantic meaning of floor plan elements.

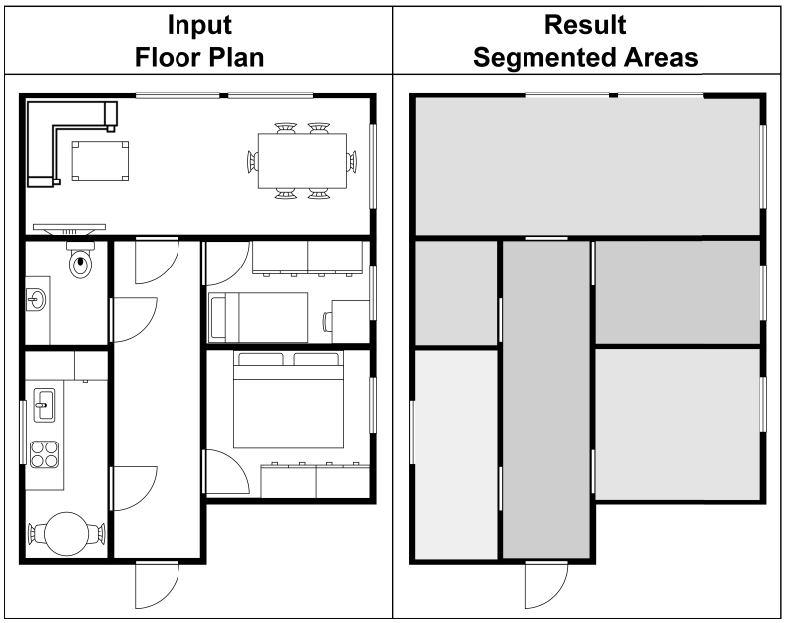

Figure 13. Visualization of the Segmentation task. An input image (**left**) is searched for separated areas representing individual rooms. Each room is marked with a polygon in a different color (**right**).

Grid maps, as the second subtype of analyzed data types, were also the second-most numerous and similar to floor plans in the data processing aspect. We found an example of an evolving solution in the works of Luperto et al. First found in [128], from 2014, the work presented how to build a semantic map of the environment based on the concept of building typology. In the article, the authors assumed that each property type has a different typical indoor structure of rooms. Using such an assumption, they proposed a classifier of rooms specific to a known building type. In the later article [82] from 2019, they tried to predict the layout of rooms only from a partially observed grid map. The main segmentation logic was based on the clustering process of the grid map's subparts ("faces"). Segments created by a set of processing operations (Hough transformation, contouring, wall identification, and representative line creation) were later joined together to build entire rooms. The final layout was the result of the detected room features' analysis. The newest publication [142], from 2022, indicates how much the approach evolved over a decade of research. The presented system, called Robust Structure identification and Room Segmentation (ROSE2), was reported to be capable of segmenting and identifying rooms even from a partial, heavily cluttered indoor environment's occupancy grid map. In 2012, Sjoo [144] published a semantic map segmentation prototype based on the idea of room function. Instead of focusing on the shape of the room, the main focus was on what the area is used for. The presented algorithm implemented the task using an energy maximization approach, where the energy function defined how well a tuple of room, label, and relational index described the space analyzed. In one of the first realizations, Hellbach et al. [127] presented research on semantic labeling using two elements: Non-negative Matrix Factorization (NMF) and Generalized Learning Vector Quantization (GLVQ). In their solution, a set of basic primitives and activities was achieved with NMF and used to create histograms of environment characteristics. The distance transformation applied to the combined results encoded the representation into a vector space. Finally, the GLVQ was used to predict the class labels for each primitive analyzed. Hemachandra et al. studied the text descriptions used in semantic labeling. An article [192] from 2014 presented a continuation of their previous work [206]. In the original version, a robot learned the environment representation in a guided tour from grid maps and textual descriptions. The discovered environment was decomposed into uniformly sized regions and represented using a semantic graph. In the later work, the inter-region connections were additionally analyzed and the scene classification factors were extended. The introduced use of the robot's camera and laser

rangefinders allowed for area-type reasoning, even without textual descriptions. In 2015, Capobianco et al. [145] presented a solution nowadays described as relatively simple. They proposed detecting walls in a grid map and using them as lines to create a matrix-like grid, where each adjacent cell was later merged with the use of expert-provided knowledge or the watershed algorithm. Also in 2015, Liu et al. [146] published an article documenting their research of the Generalized Voronoi Graph (GVG) application to an incrementally generated topological segmentation of an environment. In their solution, an environment-scanning robot explored an unknown area and constructed a GVG. The post-processing step clustered the initially generated, fine-grained set of areas into larger room-like structures. In the same year, Sünderhauf et al. [193] used a set of cameras and sensors placed on a robot platform exploring the environment to deduce the category of place and its meaning in a four-step procedure. The algorithm first used a CNN to classify taken photos, then classifiers to detect new scene classes, and finally a Bayesian filter and a mapping subsystem to create the resulting place labels. We also considered the very initial research article by Bormann et al. [1] for the segmentation subtask. They provided not only a review of used techniques but also their basic implementations. They compared four different approaches based on morphological operations, distance transformation, Voronoi Graph, and features processing. Their work was commonly referenced in later publications, and the analyzed techniques were widely used. It could be summarized that the algorithms described in the work are now the "classic" ones and are often used in the preprocessing phases of machine-learning-based approaches. Goeddel and Olson [147], in 2016, presented one of the first CNN-focused pieces of research on grid map segmentation. Although their network was trained to recognize only three distinctive classes (rooms, corridors, doorways), the presented architecture combining just two sets of convolutional and pooling layers with three fully connected layers was reported to implement the task successfully. Luo and Chiou [197], in 2018, presented an article focused on the construction of intelligent service robots. Their hybrid approach to semantic mapping combined two sources of data. The first one was an occupancy grid map from the robot's laser rangefinders, spatially segmented with a distance transformation. The second one was a topological map, created from pictures taken by the robot and searched for objects with the use of a CNN. Te combined results created a set of meaningful, spatially segmented topological nodes, classified by a probabilistic Bayesian Classifier, resulting in a generated semantic map. In 2018, Mielle et al. [135] presented a novel map segmentation method based on distance transformation. Their solution was reported to achieve better results than the referenced methodology of Bormann et al. [1]. In the proposed process, an algorithmic convolution of a circular kernel on a distance map was used to generate an extended set of region proposals. Such an over-segmented map was later merged into larger, straightened areas, representing rooms. The same test dataset was used by Hiller et al. [136] in 2019. They moved from a pure algorithmic to a learning-based approach. The proposed approach utilized a combination of a CNN and a segmentation network to generate the door-region hypotheses, later checked with computer vision algorithms. Based on the knowledge of where the doors were located, a segment classification into rooms and corridors was performed. A year later, in 2019, Wang et al. [50] presented a complex semantic mapping framework. In their scenario, an RGBD image was processed in three different ways and then connected to create a single result. Simultaneously, RGB images of the environment were searched for objects using a CNN, and depth data were combined with laser radar measurements to create a 2D semantic map. The acquired 3D space map was cast to a 2D navigation map and merged with the semantic map to create scene labels. In 2020, Tien et al. [140] presented a team of robots performing semantic mapping in a supervised learning approach. Their main focus was introducing P2P communication between robots and omitting the resource limitation problem. After that, they investigated the map segmentation task. Their high-level algorithm relied on initial preprocessing of the grid map to reduce noise and simplify a Voronoi Graph Map generation. Specific features of such a map were extracted and integrated into a classification algorithm. In the article, custom neural network- and

SVM-based approaches were tested, with the SVM solution reported to return slightly better results. One of the newest articles, by Jin et al. [198], presented a fused approach that combined scene recognition with object detection. Similarly to other multi-network architectures, they implemented subtasks separately and later fused the results. In this case, the EfficientNet-B2 [216] network was used for the initial classification of the scene, and the YOLOv5 model for the detection of objects. The proposed network added two fully connected layers processing the merged output of subnetworks. Additional room segmentation with MAORIS [135] and a combination of results led to obtaining a semantic metric map. In 2021, Zheng et al. [141] presented a system usable on low computing power devices like the Raspberry Pi. They proposed using a novel convolutional network architecture, LCNet, on indoor maps preprocessed with a watershed-based algorithm. Although deployable to a very specific device and under strict hardware limitations, the reported results were better than the Voronoi- or morphology-based solutions.

Point cloud and its segmentation were found in articles published at least a couple of years later than the lower border of the accepted time range. The first one, from 2015, by Macher et al. [16], reported research on creating as-built BIM models. In the article, the authors proposed a method for three-step, three-level segmentation. The first level was the whole floor—the distribution of points along the Z axis distinguished ceilings and floors. The second level segmented the rooms. First, it took part of the points building the ceiling layer and parsed it into 2D images of separated regions, and second, it applied the room information to all source points. The last processing step was to use the RANSAC algorithm to generally segment planes in the model, which exposed the internal structure of rooms. The reported results were promising, although requiring further research for noise reduction. In 2017, Bobkov et al. [32] presented a segmentation technique based on the anisotropic potential field. Their solution was reported to be applicable to 3D data representations, like point clouds generated by laser scanners, depth-aware sensor measurements, or even CAD models. The algorithm started with free space detection inside the 3D model. Computed for free voxels, values of anisotropic potential fields were stacked vertically, and maximum values were represented as a 2D map. Clustering of the map performed the actual room segmentation and was finally moved back to the 3D representation by labeling each point with a determined room assignment. Chang et al. [69], in 2017, published the largest found dataset of fully annotated RGBD panoramic images used for scene understanding tasks. The set, namely Matterport3D, containing over ten thousand panoramic views and built from almost 200,000 RGBD images, represents 90 building-scale scenes, their panoramic skyboxes, textured meshes, scanner locations, and other features, including point cloud representations. During the dataset presentation, not only data but also its application was presented—solutions to sample tasks of keypoint matching, view overlap prediction, estimation of the surface normal, region-type classification, and semantic voxel labeling. Regarding room segmentation, it was a CNN-based approach for scene classification. Sample baseline results were achieved with the use of ResNet-50 [207]. In 2018, Elseicy et al. [34] proposed combining (in the process of space subdivision) data from the point cloud and the scanner's trajectory. They used an Indoor Mobile Laser Scanner (IMLS) to capture the initial point cloud and record the scanning trajectory. First, the trajectory was used to detect staircases and separate stories—in the algorithm, timestamp-ordered trajectory points are assumed to represent the same segment based on the detected height. A later combination of trajectory and points analysis allowed for doorway detection, which finally separated the rooms. In the article by Nikoohemat et al. [38], the methodology was also tested with additional attention placed on reflective surfaces causing noise in the data and glass walls influencing the measurements. Also, in 2018, the work of Zheng et al. [39] researched a similar setup in the space subdivision task based on the scanline analysis process. The paper presented a novel method for opening extraction that searched for geometric regularities in scanlines and returned opening candidates. Based on the opening detections, the trajectory became segmented, and results were populated to the whole point cloud. In 2019, Cui et al. [41] presented a fully automatic

solution, performing unstructured point cloud segmentation through graph cuts based on semantic constraints. Although their scenario was to create a 3D model for 5G signal simulations in indoor environments, the segmentation method successfully generated meaningful floor plans. In a short summary, the point cloud was searched for openings and combined with the scanner trajectory for visible point cloud simulation and initial space subdivision. Additional processing, using an energy minimization function, clustered and extracted similar point clouds. Their representation as lines built a floor plan. Further processing clustered similar lines, grouped them, applied conditions, and finally generated a segmented model. In 2020, Frías et al. [59] presented a segmentation method for moving well-known and tested 2D image processing techniques into three-dimensional space. Their approach processed a 3D point cloud using 3D morphological operations. Initially, the point cloud was voxelized, and its contour was extracted using a concave hull. It left for further processing only the empty voxels that were initially filling the scanned structure. Applied morphological erosion broke connections between consecutive rooms, and a 3D-connected component algorithm clustered the remaining voxels, removing noise. Finally, a morphological dilation with the same structuring element was applied, which formed a whole opening operation and restored the segmented model. A similar approach was presented in the newest article by Yang et al. [66] from 2021. Aware of potential problems with memory usage, in the first step they stored the voxelized model in a structure called VDB (a variant of the B+ tree with dynamic topology). Next, Euclidean Distance Transformation (EDT) was applied, and based on its results the inner spheres were used to pack the space, resulting in the model being like a bubble-based one. Finally, the initial segmentation was achieved by analyzing the topological graph of created spheres, and the final result was obtained using the wavefront growth algorithm. Wang et al. [65], in 2021, presented a novel strategy for dense 3D model reconstruction with a high control of memory consumption. Their method, which is an extension of previous research [208], minimized the used resources by constructing the model on the fly, including the spatial understanding of space segmentation, by merging sub-maps of the entire scanned environment instead of storing all captured data. As the initial processing was already prepared, the paper describes only the newer part related to the environment sub-map management. The first novelty was the trigger for a new sub-map generation. The second was the method of sub-map overlap estimation, and finally, a new confirmation method for rejecting uncertain merges of sub-maps. The reported results indicated a reduction in memory usage by 50% compared to their baseline results.

5.1.10. Indoor Navigation

During the analysis, we split the indoor navigation task into four subcategories. Typically, the indoor navigation task achieved some variation of user guidance in environment discovery. Depending on the exact task, it could take the form of path planning from point to point (scenario visualized in Figure 14), user localization, robot navigation in an unknown environment, or extension of building visualization with Augmented/Virtual Reality (AR/VR).

Localization was the first subtask analyzed. We found four main groups of data types to be in use. Multiple solutions based their processing on more than one input type and combined the results to improve performance. Received Signal Strength (RSS) in a Wi-Fi network was one of the data types from the Feature Set group. In 2014, Schäfer [199] discussed the problem of concerns while using machine-learning algorithms in W-LAN fingerprinting. In the test scenario, three different smartphones were used to collect test samples, and multiple machine-learning approaches (KNN, SVM, Naïve Bayesian) were tested in the room labeling task. The reported results indicated no significant influence of the data preprocessing on the end results. Similar research was carried out by Zhang et al. [93]. In the article found, data types from the 2D Images bucket were used—floor plan analysis was presented to support the Adaptive Indoor Wi-Fi Positioning System known from [217]. The radio-based solution collected RSS measurements and extracted mobility patterns

and used them to create graphical and radio maps of the environment, later used to locate new samples. The proposed floor plan analysis extended the work with semantic meaning (room numbers) and simplified indoor navigation. A combination of floor plan analysis and radio fingerprinting was also found in the article of Laska et al. [200] from 2020. They focused on the crowd-sourced processing of location data accumulated over time. In their method, a floor plan was dynamically segmented based on the detected radio fingerprints and generated radio map. Detected locations were clustered together, creating a prediction of rooms, updated as the amount of data increased and adjusted with a parameter describing the sensitivity to individual areas. The point cloud was the subtype representing the 3D Spatial Data bucket. Although the data type suggested fully spatial input, we discovered that, in all three articles, the actual processing was based on RGBD images converted to point clouds. In the article by Liu et al. [196] from 2016, a wearable mobile device was used by a person with visual impairment to orientate in the environment. The semantic localization was deduced from a 3D indoor environment map created with the use of a Kinect sensor. Room classification performed by CNN gave the initial idea of the location and was corrected by the detection of representative objects—characteristic of specific rooms. The same year, an article by Martínez-Gómez et al. [22] presented a more universal work, a point cloud library-based framework for semantic localization. The implemented method relied on a Bag Of Words (BoW) technique. Input RGBD images were searched for valuable features, which created a dictionary of 3D words, later used to generate descriptors. Many configurations were tested, and an SVM with a combination of Harri3D (as keypoint detector) and PFHRGB (as feature extractor) was reported to score the overall highest location accuracy. The newest article found by Rusli et al. [56] proposed a full Simultaneous Localization and Mapping (SLAM) method. Their implementation processed two separated yet synchronized data samples for each analyzed timestamp—one from the RGBD sensor and one from the robot's odometry (position and orientation). An RGB image was searched for objects with the use of the YOLOv3 detector. The depth image was converted to a point cloud and used for wall detection and placing found objects in the space. During navigation, the proposed algorithm assumed that each room was built with walls and that if no walls were detected some virtual ones were accepted. If the robot crossed a wall, a new room was spawned, eventually generating an entire floor map with the robot's localization awareness. Mixed sensor readings were found as a second subtype from the Feature Set. In 2015, Hardegger et al. [181] presented a system based on two devices working together—a smartphone in the person's pocket and a foot-mounted IMU. Smartphone sensors provided data for the recognition of specific user actions, such as sitting, standing still, or walking up stairs. IMU readings were used for trajectory calculation and walked-path recognition. The actual room segmentation was carried out as in the floor plan's processing—using a location heatmap, which segmented parts of the user trajectory. The proposed solution achieved the subroom accuracy of localization in a multi-story building. Another system was created by Liu et al. [24] in 2016. They proposed to use multiple smartphone sensors to acquire data—camera, accelerometer, gyrometer, compass, magnetometer, and even Wi-Fi. The entire methodology consisted of two parts: offline training and online localization. First, a CNN model was trained for indoor scene classification, and a database of trajectories was constructed. Later, pictures from the camera were classified for the initial localization proposal. Both data sources were fused by a filtering algorithm and predicted the final location. In the article by Carrera V et al. [209] from 2018, a similar set of inputs was extended with a floor map. Instead of a CNN, they based the solution on a Monte Carlo Localization (MCL) with Bayesian filtering. Aware of reading instability, the localization problem was described as system state estimation from a sequence of noisy measurements. Newly introduced floor map data were combined with a filtering process to reduce the influence of noise on results. The whole system was equipped with an additional method for recovering from localization failure.

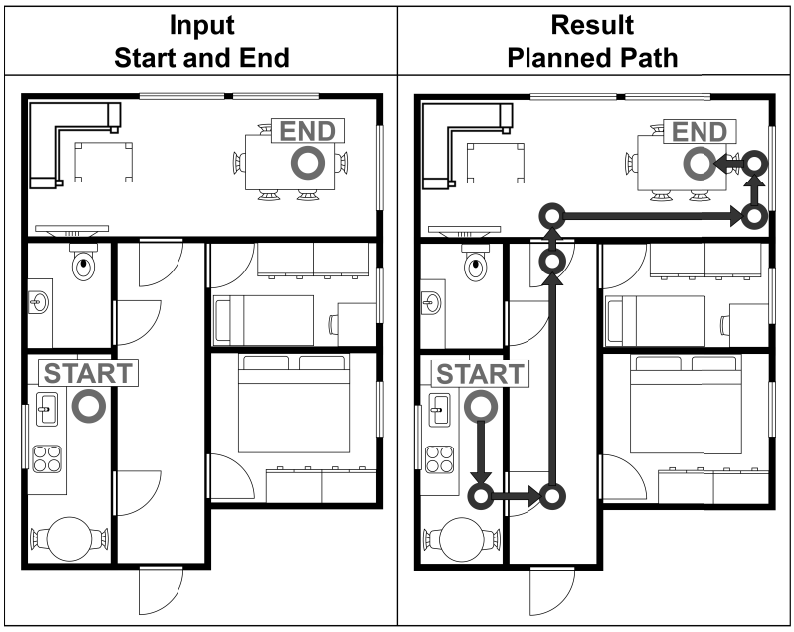

Figure 14. Visualization of the Indoor Navigation task. Input consists of the selected points of start and end (**left**). As a result of processing, a proposed path is generated and presented as a set of instructions (**right**).

The **Robot Exploration** task was mentioned in seven found articles and based on two types of input data. The first paper, from 2015, by Rojas Castro et al. [182], presented a solution for the real-time navigation of a humanoid robot. The method followed a human-like behavior: the robot first looked at a floor map of the environment, processed it, and then used it to navigate during the exploration. Map processing consisted of information segmentation, structural analysis, and semantic analysis, and the combined results were used as input to a neural system responsible for online navigation. Similarly, the processing of 2D images was found in two consecutive articles by Fermin-Leon et al. [129,132]. The first one presented a Contour-Based Topological Segmentation of a grid map. In the approach, a Dual Space Decomposition (DuDe) algorithm was used for segmentation and later improved with its custom incremental version. The idea was that new scans were aggregated to the already processed map—only the contours coming from new data were modified, leaving the rest unchanged. This allowed achieving similar quality results faster than before. The later article proposed to use it as a part of an online algorithm called TIGRE. It combined Graph-SLAM features and described contour-based segmentation, graph processing, and real-time decision-making to implement an autonomous robotic exploration. The implemented solution formulated the problem as a traversal of all graph edges generated online during environment exploration. The achieved results have shown similar error estimations as the off-line algorithms requiring prior full-graph awareness. Kleiner et al. [131], in their work from 2017, focused on a room-by-room coverage task. Their scenario was to improve how an autonomous cleaning robot explores the environment and its time optimization. A median filter, morphological closing operation, and a watershed algorithm applied to a distance-transformed map were used for region segmentation and feature extraction. The extended clutter removal process, region merging, and the environment presentation as a region graph were utilized for later planning of the cleaning path. The reported results indicated a reduction in mission execution time in both types of movement: cleaning and path following. A more general approach to the topic was presented by Cruz et al. [171] in 2018. In their article, the main focus was not directly on navigation but on the robot's learning process. As stated, the environment in which the robot functions is constantly changing, and a full retrain of a neural network is a highly time-consuming task. The presented work described several approaches to including new data in the trained model of a deep CNN. Two architectures of the system were

proposed and tested in a set of twelve different experiments. The reported results indicated that although a pure CNN achieved the best accuracy, its training time disqualified this approach, and a combination of CNN and a naïve classifier performed better. Another approach to segmentation based on graph analysis was presented by Liu et al. [134] in 2018. The proposed method focused on solving the "kidnapped robot" problem, i.e., restoring the robot's true global localization after an incorrect one. The method first cleaned the map with the use of a median filter and clustered free space using the Iterative Self-Organizing Data Analysis Technique Algorithm (ISODATA) [218] and ray casting [219]. Each cluster's center was used as a graph node representation and merged accordingly to their calculated connections. Searching for complete graphs led to the final segmentation of the region. The newest article by Balaska et al. [156] from 2020 approached the topic from the unsupervised learning point of view. In their proposal, a robot exploring the environment was processing consecutive frames of the video and combining them into a graph structure; they were then clustered into communities according to the Louvain Community Detection Algorithm [220]. Communities re-clustered with the robot's odometry readings provided the final segmentation of frames and led to topological map creation, which could be used in semantic localization.

The **Path Planning** subtask was found in two articles. Methods presented in both are similar in the processing procedure to some of the already described solutions. The article of Blöchliger et al. [70] was focused on processing point cloud data in a way similar to the plain room segmentation task: a topological map of the environment was created by segmenting the free space into a set of convex regions. This was achieved by point cloud voxelization, Truncated Signed Distance Fields calculation, and convex clustering, with the final clusters being merged into larger, enclosed areas (e.g., rooms). The Path Planning was accomplished with an A* searching performed on a topological graph, where the clusters were nodes, and their adjacency was represented as edges. Hang et al. [133], in 2018, proposed basing a multi-strategy path planner on the idea of space accessibility. An occupancy grid map was processed to identify rooms and hallways (with a GVG and Conditional Random Field model), later represented as a region topological map. The proposed method combined the grid and topological maps analysis and tried to create a planned path so that the hallways were preferred to be used. This was achieved using Dijkstra's algorithm, the mentioned GVG, and a custom implementation of the A* algorithm. The results were reported to be closely similar to the path selection of a human being.

The **AR/VR Application** was found in one article from 2018 only—by Sharma et al. [40]. Their system consisted of two main elements: a smartphone used inside the explored building and a server conducting calculations. Based on a 3D model of the building (generated from 2D architectural blueprints), they showed how to process it into a topology model and combined it with a Wi-Fi indoor positioning system. This method first created a 3D geometry replica of the building. Then, it processed it into two representations: a voxelized space for initial room segmentation and a skeleton graph for 3D navigation. The prepared data were stored in a database and, during the navigation, combined with readings from the smartphone's Wi-Fi, resulting in real-time localization information shown on the screen.

5.1.11. Alignment/Matching

In the alignment task, we divided the solutions into two subtasks based on the processed data type. Their common goal was to align a new input image to a reference one with the highest achievable precision. An example of the matching process, performed on a set of misaligned input floor plans, is presented in Figure 15.

Grid Map Alignment was a subtask represented by three articles found. One of the discovered use cases for such functionality was a service robot trying to match the current sensor readings and a newly created apartment grid map to a known map of the environment. Kakuma et al. [130], in 2017, presented such a solution based on the graph-matching procedure. The whole system first segmented regions in the grid map (using morphological operations), extracted a graph from them, and represented it as a

tree structure. The estimated transformation matrix was later checked for similarity, and a final alignment occurred. Similarly, in 2019, Hou et al. [148] presented an area graph-based solution. In the method, they first segmented the grid map using the Area Graph. Then, for each area a set of features like area size or convex hull were calculated, relying on a novel feature called the "passage distance". Later processing calculated a weighted sum of the extracted features, found the best matches between areas, and used the overlap of well-matched ones to find the final transformation needed to align the maps. The topic of aligning two maps was also discussed in 2018 in the article by Shahbandi et al. [187]. They presented a solution for the layout map (blueprint, floor plan) alignment with the sensor map (grid map) and focused on the grid map deformation corrections. The proposed method started with decomposition-based segmentation and presented results as a Doubly-Connected Edge List (DCEL).

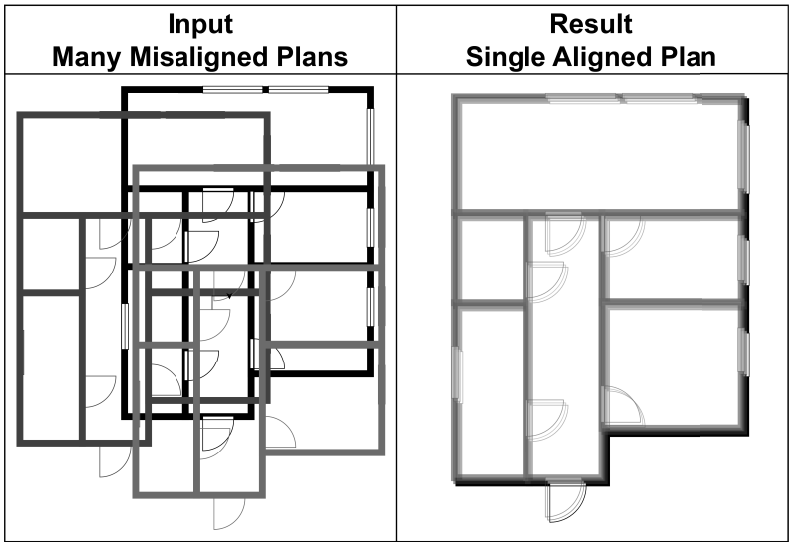

Figure 15. Visualization of the Map Alignment task. A set of misaligned input plans (**left**) is combined into a single aligned result (**right**).

The **Plan Alignment** task was found in the publication from 2016 by Sharma et al. [92]. Their main objective was to provide a solution for matching and retrieving similar floor plans, with a sample image being the search query. Like previous methods, this one also started with room segmentation (by boundary detection and morphological operations), adjacency recognition, and presentation in a topology graph form. A novel graph spectral embedding feature was proposed and used for floor plan representation, reducing the graph-matching computation time. The proposed method included information about the room décor in the retrieval process.

5.2. Taxonomy Summary and Discussion

In this section, we present a brief discussion of the discovered dependencies and observed tendencies. The number of implementations found is too large to summarize them all at once, so we do it task-by-task. A general discussion is held at the very end.

3D Model Generation. We can distinguish highly professional solutions (dedicated to commercial use) and solutions based on simplifications (often intentional, aimed at reaching non-professional users). The creation of a three-dimensional model was best carried out using three-dimensional data. In our opinion, the best results were achieved through point cloud processing. Such input is the best reflection of the actual world (the most precise), and its influence on the quality of the results was visible. Working with two-dimensional data lacks information about the third dimension, and its recovery is always fraught with errors. In one-dimensional data, it was impossible to recreate the model at all. Subjectively, the best solution currently achievable is a combination of algorithmic point

cloud preprocessing and machine-learning-based postprocessing. The first approach would achieve some kind of data preparation and cleanup, e.g., clustering and outlier removal. The latter would retrieve the actual environment structure.

Content-Based Image Retrieval (CBIR) was a relatively rare task, even though searching by image seems to be gaining popularity in the context of type recognition and floor plan analysis. All the solutions we found were processing the data in the same way. First, they analyzed the input image, then created a graph representation of it, and finally compared the structure for similarities. The obtained results were satisfactory, but we see a research niche here for some more sophisticated processing, especially for further development in the direction of recommendation systems.

Environment Description Creation is still not popular yet is a very important topic, especially for visually impaired people. The solutions we found in this area were implemented in disability support systems and were based solely on processing room plans, mainly using neural networks. There are unexpectedly few works in this research field and with a narrow scope of processed inputs. We can see great potential for developing systems that generate descriptions based on three-dimensional data. They could significantly improve the precision of generated explanations and reduce data loss, e.g., by providing details about the height of tables, countertops, chairs, or the steps of stairs.

Floor Plan Vectorization and its emergence was purely caused by the imperfections of the input data. By definition, raster images of floor plans do not scale well and lose quality while enlarging. The easiest solution would be to change the acquisition method. However, this is not always possible because vectorization is commonly used in digitizing existing documents. The works found here were trying to retrieve the lost information and switch the image format to a scalable one. This task was dominated by machine learning, mainly covering neural networks. Subjectively, the best results are currently achieved by Generative Adversarial Networks (GANs).

Floor Plan Prediction/Generation. We observed an immense expansiveness of the quality of the results. This was caused by the processing of data with virtually any degree of complexity. The most straightforward and least accurate predictions were achievable even on one-dimensional data, although the most accurate mapping was performed using three-dimensional input. With the use of spatial data, it was also the easiest one to implement. Technically, obtaining a 2D plan from a point cloud or 3D model is a kind of simplification. In such a case, we suggest moving from a more information-rich representation to a simpler one.

Graph Generation was the task implemented to utilize the advantages of graph structures, i.e., well-established analysis of relationships between the nodes. Here, we have a similar observation to the one from floor plan prediction—the creation of the graph takes place mainly based on input data with a higher degree of complication. The created graph is easier to process and better highlights simple dependencies, hard to notice in the data noise of more complex inputs. This task is not yet dominated by machine-learning solutions but with the increasing importance of the fairly novel idea of Graph Neural Networks and their variations.

Room Classification is a complex task carried out for every one of the found data types. Subjectively, the best results are achieved while processing authentic images of the environment, taken manually by a human or from a mobile platform/droid. The deep neural networks and their convolutional subtype remain the most popular and unbeatable in the achieved quality of results. The most promising solutions are those analyzing three-dimensional point clouds. However, their degree of maturity is too low to consider them as the default choice. They have great potential but are subject to an extensive calculation overhead, which is not yet justified by the quality of the results.

Change Detection was found to be a highly underestimated application that could noticeably improve the process of building lifecycle monitoring—only a single work focused on indoor environments. Most of the articles on this topic were filtered out of this research

because the analysis focused on processing aerial photos. The use of building models to detect, analyze, and document changes remains a field for further investigation.

The **Segmentation** task best shows the evolution of the input types and processing methods over time. Historically, the first was the segmentation of building plans. Their acquisition was the simplest and most widespread one. With time and the development of the available equipment (e.g., automatic cleaning robots), the segmentation issue expanded to occupancy maps. Finally, the most modern structure is the point cloud. Segmentation based on two-dimensional data is noticeably more explored than spatial data processing. Equipment capable of acquiring 3D representations of space is still a novelty that makes spatial data processing a leading topic of ongoing research in this field. It is worth noting that selected techniques that work well in two-dimensional data can be successfully adapted and applied to three-dimensional structures. This applies to transforming symbolic representations (occupancy maps) and real images. Computer vision techniques, for example, morphological operations, prepared originally to work with a matrix of pixels, can be effectively transferred to three dimensions. This requires some additional work and adjustments (such as the voxelization of models), but it is possible and has been successfully utilized. The same can be said about clustering, distance transform, watershed algorithms, etc. They worked well in the past and seem to work correctly nowadays. The only challenge remaining is to preprocess the data to present analogous structures properly but with an increased number of input dimensions.

Indoor Navigation was the task that combined the previously described methods to achieve a new output and give it a more implementation-oriented application. Regardless of the specific subtask, the elements of segmentation and classification processes were carried out the same way as in the already discussed examples and then combined to obtain a more complex result. An issue more broadly and more thoroughly studied than Simultaneous Localization and Mapping (SLAM) was not described directly in this article.

The **Alignment/Matching** task is performed mainly as a step in solving more complicated problems, such as navigation in the environment. It implements segmentation and classification as a means to reach a different goal. It is hardly done independently without further processing. For two-dimensional data, the task is dominated by a methodology based on a two-step approach. First, the images are analyzed and used to create graphs of the processed environment. Later, the graphs are compared and their similarity is evaluated. For three-dimensional data in the form of point clouds, alignment is part of the structure acquisition process itself, known as 'point cloud registration'.

General discussion. One of the most important results of our work is the following: researchers should be aware of the characteristics of their input data representation and not be afraid to change it. Replacement of the acquisition device, and therefore the source data structure, as well as conversion of the already obtained data to another format, can significantly simplify the processing needed to achieve satisfactory results. Furthermore, we should generally be open to flexible and dynamic data representation changes. The dimensionality of acquired measurements is not fixed. We can reproject it and adjust it to our needs. We observed that, in some cases, it is worth trying to obtain a more complicated input structure and cast it onto a simpler one. We could see this in the processing of floor plans generated from point clouds or graphs based on the occupancy maps. What is interesting is also the visible interdisciplinarity of solutions. Many of the algorithms successfully processing N-dimensional structures could process the N + 1-dimensional samples with just minor adjustments. Segmentation and classification processes are evolving from a problem themselves to a step in more complex systems.

6. Bibliometrics Analysis

Two subjective feelings accompanied us during the research. The first one concerned the number of articles—the closer to the current year, the more articles were found, and with a visibly growing trend. The second concerned the content of the publications—most of the articles seemed to have similarly referenced basic methodologies or processing.

To prove these statements, we conducted an additional, purely bibliometric analysis of found publications. The first claim was analyzed with the use of statistical evaluation. To visualize it, we created a year-to-year publication quantities chart, presented in Figure 16. The chart is based on 51 papers selected for full, more detailed analysis. A significant increase in the number of publications regarding the room segmentation and classification topic can be spotted in the last five years. The numbers almost tripled from 13 articles published in the first half of the analyzed period (years 2012–2016) to 37 in the second (years 2017–2021). An extreme outlier of the trend was the year 2020, when the year-to-year ratio reached the level of −50%. This date coincides with the peak of the COVID-19 pandemic outbreak. The quantity of articles from 2022 is also significantly smaller, but as this survey was conducted before the end of the year 2022 its value should be omitted in the statistics discussion as a not reliable measurement. In this way, the first statement is confirmed. In general, a growing interest in research on the topic of automated room segmentation and classification can be seen, which manifests itself in an increasing number of academic publications discussing it.

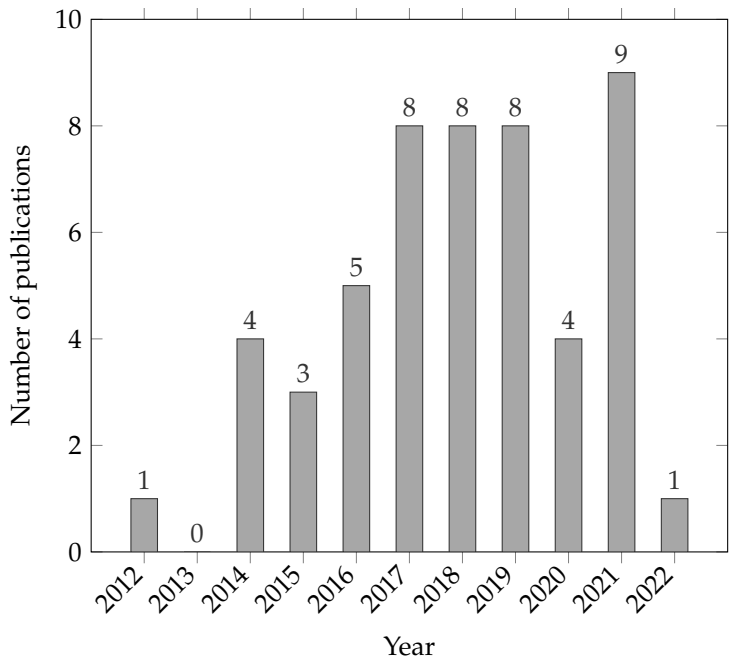

Figure 16. The number of found publications discussing the topic of room segmentation and classification in 2D images per year between 2012 and 2022.

To prove the second statement, we created a reference diagram of the links between the analyzed articles themselves. This part of the bibliometric analysis was based mainly on the widened set of publications, including papers pointed out as important during the survey extension. It constructed a collection of 89 articles, which were analyzed for their linking. By a "link," we mean a citation of an article from the analyzed collection by another article from the collection. To conduct the bibliometric analysis, we used two dedicated applications. VOSviewer was used to obtain the metadata regarding the citations and references [221], and Gephi was used for link visualization [222]. The prepared graph of links between found papers is shown in Figure 17.

To simplify interpretation, only five articles (with the highest number of links) are presented using their authors' names and year of publication. Including all labels would make the graph unreadable. The counts of links per document varied between 0 and 26. Most references (26) pointed to the article by de las Heras et al. from 2014 [87]. The following four most referenced works were (in decreasing order) Ahmed et al. from 2012 [86] with 22 links, Bormann et al. from 2016 [1] with 19, Ahmed et al. from 2011 [112] with 18, and finally, Dodge et al. from 2017 [94] with 14 links. For a more comprehensive

analysis and trend comparison, the links were also recalculated for the smaller set of articles. There would be only two updates. First, if narrowed to the main flow of the survey (51 articles), just one of the five publications with the most links would be replaced. The difference would be the article by Ahmed et al. from 2011 being not in the range of accepted publication dates, and Madugalla et al. from 2020 waqs provided instead [103]. The second change would relate to the order of results: the two first publications would switch places. We consider the second statement as confirmed, likewise the first one. Even after almost a decade, all the newfound articles were intensively referring to one another and with a visible set of well-known base methodologies.

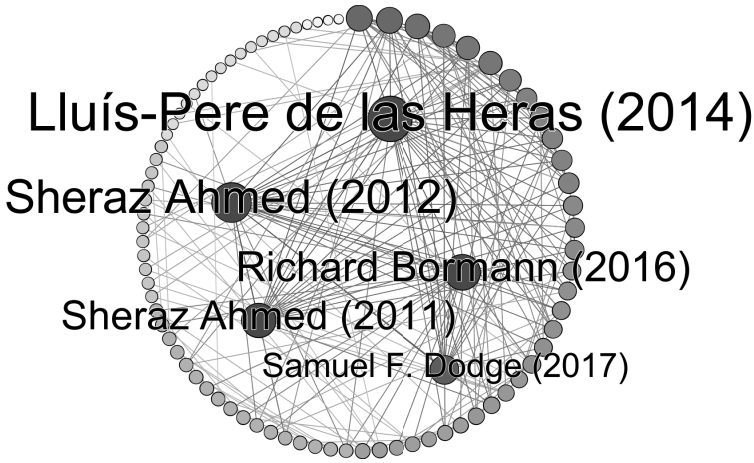

Figure 17. Diagram of link-references between articles of the extended bucket of papers processing 2D Images. In decreasing order, from top to bottom, papers with the most links: ([1,86,87,94,112]).

7. Challenges

The extensiveness of the conducted survey allowed us to detect some common problems and challenges that the researchers had to overcome in their studies. They differ from one analyzed solution to another but, at the same time, are a good indicator of what frequent difficulties may be encountered during the research, independently from the low-level accomplished task. We specify a set of five such challenges and describe them in this section. The most important one (in our opinion) seems to be aligned with the topic of this publication. It is the diversity of input data and their representation. During the research, we have pointed out many main types of data but it is still not enough. The real challenge here can be easily described with the use of building floor plans. They are not unified. We can find multiple different styles of plans. They can be monochromatic or colorful, simplified for easier understanding or filled with technical details for constructors, empty or filled with furniture pictograms—even the same building can have many different floor plan representations. The same can be said about other data types, e.g., spatial models being mesh models or point clouds. The second and third challenges are related and should be discussed together. The first problem is the amount of data that needs to be collected, stored, and processed. Nowadays, researchers have to expect millions of samples and terabytes of measurements. They need to have a plan regarding how to receive it fluently, save it, and prepare it for later use. The second topic is how to overcome these issues while working with very limited hardware resources. It works fine when more computing power and better equipment come together with the increasing volume of data. An example here can be the process of professional 3D modeling in digital twin reconstruction. Although a scan of a building delivers massive amounts of data and requires powerful computers, they are available and used to work with all the acquired measurements. But this is not always the case. On the one hand, the growing interest in smart solutions results in a rapidly increasing number of sensors used and providing data, but on the other hand such sensors are not powerful in the aspect of processing capabilities. The idea of edge computing seems to be an urgent topic for extensive research, as the

data volume increases rapidly and the sensors are forced to be small, cheap, and with long battery life, restricting their computational power. The fourth discovered challenge is dedicated to human participation in data gathering. As mentioned in the introduction to this research, the best scenario is achieved when the proposed solution requires little to no user involvement. As we noticed in the found articles, not every solution followed this assumption. In multiple cases, the data collection process required a researcher to spend many hours walking with a scanner or taking photos of the environment. As the operator's time is of increasing value for the whole process, newly created solutions will have to face the requirement for fully autonomous data acquisition and processing. This could be achieved with robots, drones, or smart sensors. The last discussed visible challenge is information loss. It may seem obvious, as the semantic meaning reconstruction is the main topic of this research, but it needs to be stated clearly—in some representations, specific data are irretrievably lost. An example here is a floor plan with no information on floor height. We could find multiple different ways to recreate information, but none of them can be accepted as an indisputable and always correct way of information retrieval. While preparing an implementation, researchers should be aware of the limitations present in the selected data type. In the case of floor plans, it would be enough to save the floor height as a piece of additional information in the image. However, if skipped at the beginning of processing, this information may never be reconstructed correctly. The described challenges presented a wide range of fields for further academic research and possible improvements of already existing solutions. They mostly followed the main changes visible in the current world—the increasing amount of data and its diversity with a simultaneous expected decrease in the operator's involvement in system maintenance.

8. Newest Trends and Place for Future Work

The extensiveness of the conducted research forced a specific range of publication dates to be accepted. As already presented, indoor environment analysis is a dynamically evolving field. Nevertheless, later in this section we try to indicate the main future directions of research.

One of the most important, inevitable trends is the popularization of IoT sensors and their use in everyday life. They can be integrated into various home appliances. Inspired by neuroscience, Zhu et al. [223] presented scene classification in the application of home service robots. Yang et al. [224] benefited from the diversity of available data representations and used a set of numerous formats of maps to create an autonomous navigation and landing system for an Unmanned Aerial Vehicle (UAV). Finally, Shaharuddin et al. [225] reviewed the role of IoT sensors in the aspect of fire hazard contingency in smart homes. **Internet-connected devices have become a standard in the industry, and we should place special attention on the utilization of the possibilities it generates.**

Another important trend is the increasing capability of large-volume data processing with gradually more sophisticated methods. Mahmoud et al. [226] proposed a well-performing framework for scan-to-BIM generation with semantic segmentation of input point clouds. They report values for the precision, recall, and F-score of indoor element reconstruction of 96–99%. A full-scale digital twin was proposed to be automatically generated and used in a later production system's simulation by Soomer et al. [227]. They processed huge amounts of data to plan and predict the manufacturing process, optimize it, and provide cost savings to clients. Images with an additional channel of depth were analyzed by Zheng et al. [228] with the use of transformer neural network architecture, which is gaining popularity. **All these methods were possible to implement thanks to the increasing computing power of typical workstations.**

Another aspect we would like to mention is the carbon emissions reduction and energy optimization of multiple types. This can be done at a micro- as well as a macro-scale. Single building modeling and its property analysis were analyzed regarding various aspects. Han et al. [229] presented the context of the thermal storage performance of a building; Pachano et al. [230] studied the self-consumption optimization of energy produced by pho-

tovoltaic systems; Deng et al. [231], Roumi et al. [232], and Sulaiman and Mustaffa [233] focused on the personal satisfaction (e.g., the thermal comfort) in connection with the building energy consumption. At the macro-scale, industrial analysis of energy quota trading was presented by Wei et al. [234]. **Society's ecological awareness is growing, leading to an increase in the number of analyses focusing on energy optimization.**

We left the most important trend for last: **the application of machine-learning in the analysis of indoor environment has become a standard approach**. It can be clearly seen with a search for the set of keywords 'machine-learning indoor environment' in the online repositories of MDPI, Elsevier's Science Direct, and IEEE Xplore. The summarized number of records found in all these three repositories reached 2265 for the year 2021; in 2022, it was already 2878 (+27% year-to-year), and in 2023 the number of entries increased by a further 20% and reached 3454 records.

All these trends are well-suited to be the direction of future research, and their results will definitely attract a significant amount of readers' attention.

9. Summary

Our literature review shows that the topic of room segmentation and its semantic meaning retrieval is a fresh area of research, dynamically developing new solutions relying on various data representations. In the fully automated approach to manufacturing processes, it is crucial to have a complete understanding of owned assets, especially in the form of digitalized documentation and virtual models that can be processed by an AI-based system. We have to know what data is already available or can be easily acquired, and how we can utilize it.

Performed analysis of the existing literature with the SLR methodology allowed us to identify three perspectives of dwelling on the existing works, called taxonomies. The first one (Section 3) answered the question dedicated to available input data structures. Four different types of data were found to be utilized: 2D images, 3D spatial data, graph structures, and feature sets. The second taxonomy (Section 4) presented the category of processing methods. We specified four such categories: three types of segmentation (with increasingly more precise semantic room understanding) and plain classification itself. The third taxonomy (Section 5) discussed the task accomplished by the analyzed implementations. We found eleven main types of assignments, which differed significantly from one another. Implementations stretched from tasks of 3D model reconstruction and floor plan segmentation through indoor navigation and path planning to change detection and many others.

Our subjective perception of the ongoing trends in the research was validated in the bibliometric analysis performed. Its results (Section 6) confirmed all the initial assumptions of a visibly growing trend in the number of publications that (1) discuss the topic of room segmentation and classification, (2) share similar basic processing, and (3) are strongly related to each other. The categorized knowledge also allows the reader to observe frequently discovered challenges, which seem to indicate the most common difficulties that researchers should be aware of and try to overcome (Section 7).

In summary, this research indicates that all environments can be digitalized and somehow facilitate the automation of manufacturing and controlling processes. The range of available data types and their possible applications is so vast that practically every enterprise is able to find a solution suitable for meeting its needs and capabilities. We have researched, described, and summarized recent solutions that could help them gather these data, understand them, and use them to better manage their resources. We did this with a particular focus on one specific resource management aspect—facilities' digitalization and their space organization. We wanted to know what type of input data can be processed, how, and for what reason. After reviewing such a broad collection of papers we can conclude that the diversity in the input data structures and the processing methodologies should be considered not as a problem but as an opportunity. We can choose the data representation that is best for our use case. We can adjust it to the hardware limitations,

business applications, or even our own convenience. Researchers should not restrain data reprojection, as it may significantly simplify the solution.

The limitations of this work are, at the same time, the indicators of potential future research. Because of a very wide range of covered applications and described solutions, their presentation neither goes into algorithmic details nor discusses implementation nuances. The expected continuation of this review would be to analyze its results once again and answer additional research questions, focusing on the exact methodologies presented in the found solutions. Instead of answering 'what' and 'why', it would also be of high value to elaborate on 'how' things could be achieved. As the areas of machine learning and the Internet of Things are changing every day, a review as wide as ours is expected to be continually extended with the newest innovations that appear. It would be advisable to continue such research regarding a narrower, more specific aspect of the presented taxonomies, placing special attention on only the newest ideas and trends.

Author Contributions: S.P., data curation, methodology, software, validation, writing—original draft, visualization, investigation; D.M., investigation, writing—review and editing, conceptualization, methodology, validation, supervision, funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Polish Ministry of Science and Higher Education as part of the Implementation Doctorate program at the Silesian University of Technology, Gliwice, Poland (contract No 32/014/SDW/005-46), project entitled "PMW" in the years 2021–2025 (contract no. 5169/H2020/2020/2), Statutory Research funds of the Department of Applied Informatics, Silesian University of Technology, Gliwice, Poland (grant No 02/100/BK_24/0035 and grant No BKM-2024), the pro-quality grant (02/100/RGJ23/0026) of the Rector of the Silesian University of Technology, Gliwice, Poland, and partially by the ReActive Too project that has received funding from the European Union's Horizon 2020 Research, Innovation, and Staff Exchange Programme under the Marie Skłodowska-Curie Action (Grant Agreement No 871163).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ACM	Association for Computing Machinery
AGG	Attributed Graph Grammar
AI	Artificial Intelligence
ANN	Artificial Neural Network
AR	Augmented Reality
BIM	Building Information Modeling
BLS	Backpack Laser Scanner
BNC	Bayesian Network Classifier
BoW	Bag Of Words
CAD	Computer Aided Design
CBIR	Content-Based Image Retrieval
CNN	Convolutional Neural Network
DBLP	Digital Bibliography and Library Project
DCEL	Doubly-Connected Edge List
DL	Deep Learning
DSFL	Discriminative and Shareable Feature Learning
DuDe	Dual Space Decomposition
EC	Exclusion Criteria
EDT	Euclidean Distance Transformation
ESN	Echo State Network
GAN	Generative Adversarial Network
GAT	Graph Attention Network
GCN	Graph Convolutional Network
GMM	Gaussian Mixture Model

GNN	Graph Neural Network
GVG	Generalized Voronoi Graph
IC	Inclusion Criteria
IIoT	Industrial Internet of Things
IMU	Inertial measurement unit
IoT	Internet of Things
ISODATA	Iterative Self-Organizing Data Analysis Technique Algorithm
MCL	Monte Carlo Localization
MCMC	Markov Chain Monte Carlo
MDL	Minimum Description Length
ML	Machine Learning
MLN	Markov Logic Networks
MLP	Multi-Layer Perception
MLS	Mobile Laser Scanner
OCR	Optical Character Recognition
OTC	Oriented Texture Curves
RAG	Region Adjacency Graph
RNN	Recurrent Neural Network
RQ	Research Question
RSS	Received Signal Strength
SD	Science Direct
SL	Springer Link
SLAM	Simultaneous Localization And Mapping
SLR	Systematic Literature Review
SLS	Static Laser Scanner
SVM	Support Vector Machine
UAV	Unmanned Aerial Vehicle
VR	Virtual Reality

References

1. Bormann, R.; Jordan, F.; Li, W.; Hampp, J.; Hagele, M. Room segmentation: Survey, implementation, and analysis. In Proceedings of the 2016 IEEE International Conference on Robotics and Automation (ICRA), Stockholm, Sweden, 16–21 May 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 1019–1026. [CrossRef]
2. Gimenez, L.; Hippolyte, J.L.; Robert, S.; Suard, F.; Zreik, K. Review: Reconstruction of 3D building information models from 2D scanned plans. *J. Build. Eng.* **2015**, *2*, 24–35. [CrossRef]
3. Kang, Z.; Yang, J.; Yang, Z.; Cheng, S. A review of techniques for 3D reconstruction of indoor environments. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 330. [CrossRef]
4. Mackenzie, H.; Dewey, A.; Drahota, A.; Kilburn, S.; Kalra, P.R.; Fogg, C.; Zachariah, D. Systematic reviews: What they are, why they are important, and how to get involved. *J. Clin. Prev. Cardiol.* **2012**, *1*, 193–202.
5. Gough, D.; Oliver, S.; Thomas, J. *An Introduction to Systematic Reviews*, 2nd ed.; SAGE: London, UK, 2017.
6. Grant, M.J.; Booth, A. A typology of reviews: An analysis of 14 review types and associated methodologies: A typology of reviews. *Maria J. Grant Andrew Booth* **2009**, *26*, 91–108. [CrossRef]
7. Xiao, Y.; Watson, M. Guidance on Conducting a Systematic Literature Review. *J. Plan. Educ. Res.* **2019**, *39*, 93–112. [CrossRef]
8. Santos, F.; Moreira, A.; Costa, P. Towards extraction of topological maps from 2D and 3D occupancy grids. In Proceedings of the Progress in Artificial Intelligence, Azores, Portugal, 9–12 September 2013; Springer: Berlin/Heidelberg, Germany, 2013; Volume 8154, pp. 307–318. ISBN 978-3-642-40668-3. [CrossRef]
9. Mura, C.; Mattausch, O.; Villanueva, A.J.; Gobbetti, E.; Pajarola, R. Robust Reconstruction of Interior Building Structures with Multiple Rooms under Clutter and Occlusions. In Proceedings of the 2013 International Conference on Computer-Aided Design and Computer Graphics, Hong Kong, China, 16–18 November 2013; pp. 52–59. [CrossRef]
10. Borrmann, D.; Nüchter, A.; Đakulović, M.; Maurović, I.; Petrović, I.; Osmanković, D.; Velagić, J. A mobile robot based system for fully automated thermal 3D mapping. *Adv. Eng. Inform.* **2014**, *28*, 425–440. [CrossRef]
11. Maurović, I.; ðakulović, M.; Petrović, I. Autonomous Exploration of Large Unknown Indoor Environments for Dense 3D Model Building. *IFAC Proc. Vol.* **2014**, *47*, 10188–10193. [CrossRef]
12. Mura, C.; Mattausch, O.; Villanueva, A.J.; Gobbetti, E.; Pajarola, R. Automatic room detection and reconstruction in cluttered indoor environments with complex room layouts. *Comput. Graph.* **2014**, *44*, 20–32. [CrossRef]
13. Ochmann, S.; Vock, R.; Wessel, R.; Tamke, M.; Klein, R. Automatic generation of structural building descriptions from 3D point cloud scans. In Proceedings of the GRAPP 2014—Proceedings of the 9th International Conference on Computer Graphics Theory and Applications, Lisbon, Portugal, 5–8 January 2014; SciTePress: Setúbal, Portugal, 2014; pp. 120–127. [CrossRef]

14. Ochmann, S.; Vock, R.; Wessel, R.; Klein, R. Towards the extraction of hierarchical building descriptions from 3D indoor scans. In Proceedings of the Eurographics Workshop on 3D Object Retrieval, EG 3DOR, Strasbourg, France, 6 April 2014; Veltkamp R., Tabia H., B.B.V.J.P., Eds.; Eurographics Association: Cambridge, MA, USA, 2014; ISSN 19970463. [CrossRef]
15. Swadzba, A.; Wachsmuth, S. A detailed analysis of a new 3D spatial feature vector for indoor scene classification. *Robot. Auton. Syst.* **2014**, *62*, 646–662. [CrossRef]
16. Macher, H.; Landes, T.; Grussenmeyer, P. Point clouds segmentation as base for as-built BIM creation. *ISPRS Ann. Photogramm. Remote. Sens. Spat. Inf. Sci.* **2015**, *2*, 191–197. [CrossRef]
17. Turner, E.; Cheng, P.; Zakhor, A. Fast, automated, scalable generation of textured 3D models of indoor environments. *IEEE J. Sel. Top. Signal Process.* **2015**, *9*, 409–421. [CrossRef]
18. Turner, E.; Zakhor, A. Automatic Indoor 3D Surface Reconstruction with Segmented Building and Object Elements. In Proceedings of the Proceedings—2015 International Conference on 3D Vision, 3DV 2015, Lyon, France, 19–22 October 2015; Brown M., Kosecka J., T.C., Eds.; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2015; pp. 362–370. [CrossRef]
19. Ikehata, S.; Yang, H.; Furukawa, Y. Structured Indoor Modeling. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 1323–1331. ISSN 2380-7504. [CrossRef]
20. Armeni, I.; Sener, O.; Zamir, A.; Jiang, H.; Brilakis, I.; Fischer, M.; Savarese, S. 3D semantic parsing of large-scale indoor spaces. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; IEEE Computer Society: Washington, DC, USA, 2016; pp. 1534–1543. ISSN 10636919. [CrossRef]
21. Manfredi, G.; Devin, S.; Devy, M.; Sidobre, D. Autonomous Apartment Exploration, Modelling and Segmentation for Service Robotics. *IFAC-PapersOnLine* **2016**, *49*, 120–125. [CrossRef]
22. Martínez-Gómez, J.; Morell, V.; Cazorla, M.; García-Varea, I. Semantic localization in the PCL library. *Robot. Auton. Syst.* **2016**, *75*, 641–648. [CrossRef]
23. Ochmann, S.; Vock, R.; Wessel, R.; Klein, R. Automatic reconstruction of parametric building models from indoor point clouds. *Comput. Graph.* **2016**, *54*, 94–103. [CrossRef]
24. Liu, Q.; Li, R.; Hu, H.; Gu, D. Using semantic maps for room recognition to aid visually impaired people. In Proceedings of the 2016 22nd International Conference on Automation and Computing (ICAC), Colchester, UK, 7–8 September 2016; pp. 89–94. [CrossRef]
25. Jung, J.; Stachniss, C.; Kim, C. Automatic Room Segmentation of 3D Laser Data Using Morphological Processing. *SPRS Int. J. Geo-Inf.* **2017**, *6*, 206. [CrossRef]
26. Macher, H.; Landes, T.; Grussenmeyer, P. From point clouds to building information models: 3D semi-automatic reconstruction of indoors of existing buildings. *Appl. Sci.* **2017**, *7*, 1030. [CrossRef]
27. Murali, S.; Speciale, P.; Oswald, M.; Pollefeys, M. Indoor Scan2BIM: Building information models of house interiors. In Proceedings of the IEEE International Conference on Intelligent Robots and Systems, Vancouver, BC, Canada, 24–28 September 2017; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2017; pp. 6126–6133. ISSN 21530858. [CrossRef]
28. Nikoohemat, S.; Peter, M.; Oude Elberink, S.; Vosselman, G. Exploiting Indoor Mobile Laser Scanner Trajectories for Semantic Interpretation of Point Clouds. In Proceedings of the ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Wuhan, China, 18–22 September 2017; Copernicus GmbH: Göttingen, Germany, 2017; Volume 4, pp. 355–362, ISSN 21949042. [CrossRef]
29. Wang, R.; Xie, L.; Chen, D. Modeling indoor spaces using decomposition and reconstruction of structural elements. *Photogramm. Eng. Remote. Sens.* **2017**, *83*, 827–841. [CrossRef]
30. Xie, L.; Wang, R. Automatic indoor building reconstruction from mobile laser scanning data. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **2017**, *42*, 417–422. [CrossRef]
31. Ambruş, R.; Claici, S.; Wendt, A. Automatic Room Segmentation From Unstructured 3-D Data of Indoor Environments. *IEEE Robot. Autom. Lett.* **2017**, *2*, 749–756. [CrossRef]
32. Bobkov, D.; Kiechle, M.; Hilsenbeck, S.; Steinbach, E. Room segmentation in 3D point clouds using anisotropic potential fields. In Proceedings of the 2017 IEEE International Conference on Multimedia and Expo (ICME), Hong Kong, China, 10–14 July 2017; pp. 727–732. ISSN 1945–788X. [CrossRef]
33. Brucker, M.; Durner, M.; Ambrus, R.; Marton, Z.; Wendt, A.; Jensfelt, P.; Arras, K.; Triebel, R. Semantic Labeling of Indoor Environments from 3D RGB Maps. In Proceedings of the IEEE International Conference on Robotics and Automation, Brisbane, QLD, Australia, 21–25 May 2018; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA; pp. 1871–1878. ISSN 10504729. [CrossRef]
34. Elseicy, A.; Nikoohemat, S.; Peter, M.; Elberink, S. Space subdivision of indoor mobile laser scanning data based on the scanner trajectory. *Remote Sens.* **2018**, *10*, 1815. [CrossRef]
35. Jung, J.; Stachniss, C.; Ju, S.; Heo, J. Automated 3D volumetric reconstruction of multiple-room building interiors for as-built BIM. *Adv. Eng. Inform.* **2018**, *38*, 811–825. [CrossRef]
36. Li, L.; Su, F.; Yang, F.; Zhu, H.; Li, D.; Zuo, X.; Li, F.; Liu, Y.; Ying, S. Reconstruction of three-dimensional (3D) indoor interiors with multiple stories via comprehensive segmentation. *Remote Sens.* **2018**, *10*, 1281. [CrossRef]
37. Magri, L.; Fusiello, A. Reconstruction of interior walls from point cloud data with min-hashed J-Linkage. In Proceedings of the Proceedings—2018 International Conference on 3D Vision, 3DV 2018, Verona, Italy, 5–8 September 2018; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2018; pp. 131–139. [CrossRef]

38. Nikoohemat, S.; Peter, M.; Elberink, S.; Vosselman, G. Semantic interpretation of mobile laser scanner point clouds in Indoor Scenes using trajectories. *Remote Sens.* **2018**, *10*, 1754. [CrossRef]
39. Zheng, Y.; Peter, M.; Zhong, R.; Elberink, S.; Zhou, Q. Space subdivision in indoor mobile laser scanning point clouds based on scanline analysis. *Sensors* **2018**, *18*, 1838. [CrossRef] [PubMed]
40. Sharma, O.; Pandey, J.; Akhtar, H.; Rathee, G. Navigation in AR based on digital replicas. *Vis. Comput.* **2018**, *34*, 925–936. [CrossRef]
41. Cui, Y.; Li, Q.; Dong, Z. Structural 3D reconstruction of indoor space for 5G signal simulation with mobile laser scanning point clouds. *Sensors* **2018**, *11*, 2262. [CrossRef]
42. Cui, Y.; Li, Q.; Yang, B.; Xiao, W.; Chen, C.; Dong, Z. Automatic 3-D Reconstruction of Indoor Environment with Mobile Laser Scanning Point Clouds. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 3117–3130. [CrossRef]
43. Koeva, M.; Nikoohemat, S.; Elberink, S.; Morales, J.; Lemmen, C.; Zevenbergen, J. Towards 3D indoor cadastre based on change detection from point clouds. *Remote Sens.* **2019**, *11*, 1972. [CrossRef]
44. Maset, E.; Magri, L.; Fusiello, A. Improving automatic reconstruction of interior walls from point cloud data. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **2019**, *42*, 849–855. [CrossRef]
45. Nikoohemat, S.; Diakité, A.; Zlatanova, S.; Vosselman, G. INDOOR 3D MODELING and FLEXIBLE SPACE SUBDIVISION from POINT CLOUDS. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.* **2019**, *4*, 285–292. [CrossRef]
46. Ochmann, S.; Vock, R.; Klein, R. Automatic reconstruction of fully volumetric 3D building models from oriented point clouds. *ISPRS J. Photogramm. Remote Sens.* **2019**, *151*, 251–262. [CrossRef]
47. Previtali, M.; Barazzetti, L.; Roncoroni, F. Automated Detection and Layout Regularization of Similar Features in Indoor Point Cloud. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **2019**, *42*, 631–638. [CrossRef]
48. Shi, W.; Ahmed, W.; Li, N.; Fan, W.; Xiang, H.; Wang, M. Semantic geometric modelling of unstructured indoor point cloud. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 9. [CrossRef]
49. Tang, S.; Zhang, Y.; Li, Y.; Yuan, Z.; Wang, Y.; Zhang, X.; Li, X.; Zhang, Y.; Guo, R.; Wang, W. Fast and automatic reconstruction of semantically rich 3D indoor maps from low-quality RGB-D sequences. *Sensors* **2019**, *19*, 533. [CrossRef] [PubMed]
50. Wang, P.; Cheng, J.; Feng, W. An approach for construct semantic map with scene classification and object semantic segmentation. In Proceedings of the 2018 IEEE International Conference on Real-Time Computing and Robotics, RCAR 2018, Kandima, Maldives, 1–5 August 2018; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2018; pp. 270–275. [CrossRef]
51. Yang, F.; Li, L.; Su, F.; Li, D.; Zhu, H.; Ying, S.; Zuo, X.; Tang, L. Semantic decomposition and recognition of indoor spaces with structural constraints for 3D indoor modelling. *Autom. Constr.* **2019**, *106*, 102913. [CrossRef]
52. Yang, F.; Zhou, G.; Su, F.; Zuo, X.; Tang, L.; Liang, Y.; Zhu, H.; Li, L. Automatic indoor reconstruction from point clouds in multi-room environments with curved walls. *Sensors* **2019**, *19*, 3798. [CrossRef] [PubMed]
53. He, Z.; Hou, J.; Schwertfeger, S. Furniture Free Mapping using 3D Lidars. In Proceedings of the 2019 IEEE International Conference on Robotics and Biomimetics (ROBIO), Dali, China, 6–8 December 2019; pp. 583–589. [CrossRef]
54. Nikoohemat, S.; Diakité, A.; Zlatanova, S.; Vosselman, G. Indoor 3D reconstruction from point clouds for optimal routing in complex buildings to support disaster management. *Autom. Constr.* **2020**, *113*, 103109. [CrossRef]
55. Otero, R.; Frías, E.; Lagüela, S.; Arias, P. Automatic gbXML modeling from LiDAR data for energy studies. *Remote Sens.* **2020**, *12*, 2679. [CrossRef]
56. Rusli, I.; Trilaksono, B.R.; Adiprawita, W. RoomSLAM: Simultaneous Localization and Mapping With Objects and Indoor Layout Structure. *IEEE Access* **2020**, *8*, 196992–197004. [CrossRef]
57. Ryu, M.; Oh, S.; Kim, M.; Cho, H.; Son, C.; Kim, T. Algorithm for generating 3d geometric representation based on indoor point cloud data. *Appl. Sci.* **2020**, *10*, 8073. [CrossRef]
58. Phalak, A.; Badrinarayanan, V.; Rabinovich, A. Scan2Plan: Efficient Floorplan Generation from 3D Scans of Indoor Scenes. *arXiv* **2020**. http://arxiv.org/abs/2003.07356.
59. Frías, E.; Balado, J.; Díaz-Vilariño, L.; Lorenzo, H. Point Cloud Room Segmentation Based on Indoor Spaces and 3D Mathematical Morphology. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **2020**, *44*, 49–55. [CrossRef]
60. Ai, M.; Li, Z.; Shan, J. Topologically consistent reconstruction for complex indoor structures from point clouds. *Remote Sens.* **2021**, *13*, 3844. [CrossRef]
61. Cai, Y.; Fan, L. An efficient approach to automatic construction of 3d watertight geometry of buildings using point clouds. *Remote Sens.* **2021**, *13*, 1947. [CrossRef]
62. Fang, H.; Lafarge, F.; Pan, C.; Huang, H. Floorplan generation from 3D point clouds: A space partitioning approach. *ISPRS J. Photogramm. Remote Sens.* **2021**, *175*, 44–55. [CrossRef]
63. He, Z.; Sun, H.; Hou, J.; Ha, Y.; Schwertfeger, S. Hierarchical topometric representation of 3D robotic maps. *Auton. Robot.* **2021**, *45*, 755–771. [CrossRef]
64. Hübner, P.; Weinmann, M.; Wursthorn, S.; Hinz, S. Automatic voxel-based 3D indoor reconstruction and room partitioning from triangle meshes. *ISPRS J. Photogramm. Remote Sens.* **2021**, *181*, 254–278. [CrossRef]

65. Wang, Y.; Ramezani, M.; Mattamala, M.; Fallon, M. Scalable and elastic LiDAR reconstruction in complex environments through spatial analysis. In Proceedings of the 2021 10th European Conference on Mobile Robots, ECMR 2021—Proceedings, Bonn, Germany, 31 August–3 September 2021; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2021. [CrossRef]
66. Yang, F.; Che, M.; Zuo, X.; Li, L.; Zhang, J.; Zhang, C. Volumetric Representation and Sphere Packing of Indoor Space for Three-Dimensional Room Segmentation. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 739. [CrossRef]
67. Yang, J.; Kang, Z.; Zeng, L.; Hope Akwensi, P.; Sester, M. Semantics-guided reconstruction of indoor navigation elements from 3D colorized points. *ISPRS J. Photogramm. Remote Sens.* **2021**, *173*, 238–261. [CrossRef]
68. Cai, R.; Li, H.; Xie, J.; Jin, X. Accurate floorplan reconstruction using geometric priors. *Comput. Graph.* **2022**, *102*, 360–369. [CrossRef]
69. Chang, A.; Dai, A.; Funkhouser, T.; Halber, M.; Nießner, M.; Savva, M.; Song, S.; Zeng, A.; Zhang, Y. Matterport3D: Learning from RGB-D Data in Indoor Environments. *arXiv* **2017**. Available online: http://arxiv.org/abs/1709.06158, (accessed on 28 April 2024).
70. Blöchliger, F.; Fehr, M.; Dymczyk, M.; Schneider, T.; Siegwart, R. Topomap: Topological Mapping and Navigation Based on Visual SLAM Maps. *arXiv* **2018**. Available online: http://arxiv.org/abs/1709.05533, (accessed on 28 April 2024).
71. Chen, J.; Liu, C.; Wu, J.; Furukawa, Y. Floor-SP: Inverse CAD for Floorplans by Sequential Room-Wise Shortest Path. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 2661–2670. [CrossRef]
72. Carrera V.J.L.; Zhao, Z.; Braun, T.; Li, Z.; Neto, A. A real-time robust indoor tracking system in smartphones. *Comput. Commun.* **2018**, *117*, 104–115. [CrossRef]
73. Matez-Bandera, J.; Monroy, J.; Gonzalez-Jimenez, J. Efficient semantic place categorization by a robot through active line-of-sight selection. *Knowl.-Based Syst.* **2022**, *240*, 108022. [CrossRef]
74. Liu, N.; Lin, B.; Yuan, L.; Lv, G.; Yu, Z.; Zhou, L. An Interactive Indoor 3D Reconstruction Method Based on Conformal Geometry Algebra. *Adv. Appl. Clifford Algebras* **2018**, *28*, 73. [CrossRef]
75. Weinmann, M.; Wursthorn, S.; Weinmann, M.; Hübner, P. Efficient 3D Mapping and Modelling of Indoor Scenes with the Microsoft HoloLens: A Survey. *PFG* **2021**, *89*, 319–333. [CrossRef]
76. Franz, S.; Irmler, R.; Rüppel, U. Real-time collaborative reconstruction of digital building models with mobile devices. *Adv. Eng. Inform.* **2018**, *38*, 569–580. [CrossRef]
77. Gao, R.; Zhao, M.; Ye, T.; Ye, F.; Wang, Y.; Bian, K.; Wang, T.; Li, X. Jigsaw: Indoor Floor Plan Reconstruction via Mobile Crowdsensing. In Proceedings of the 20th Annual International Conference on Mobile Computing and Networking, MobiCom'14, Maui, HI, USA, 7–11 September 2014; Association for Computing Machinery: New York, NY, USA, 2014; pp. 249–260. [CrossRef]
78. Liu, C.; Wu, J.; Furukawa, Y. FloorNet: A unified framework for floorplan reconstruction from 3D scans. In Proceedings of the Computer Vision—ECCV 2018, Munich, Germany, 8–14 September 2018; Springer: Cham, Switzerland, 2018; Volume 11210, pp. 203–219. ISBN 9783030012304. [CrossRef]
79. Liu, H.; Yang, Y.L.; AlHalawani, S.; Mitra, N.J. Constraint-aware interior layout exploration for pre-cast concrete-based buildings. *Vis. Comput.* **2013**, *29*, 663–673. [CrossRef]
80. Luperto, M.; D'Emilio, L.; Amigoni, F. A generative spectral model for semantic mapping of buildings. In Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, 28 September–2 October 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 4451–4458. [CrossRef]
81. Luperto, M.; Riva, A.; Amigoni, F. Semantic classification by reasoning on the whole structure of buildings using statistical relational learning techniques. In Proceedings of the Proceedings—IEEE International Conference on Robotics and Automation, Singapore, 29 May–3 June 2017; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2017; pp. 2562–2568. ISSN 10504729. [CrossRef]
82. Luperto, M.; Amigoni, F. Predicting the global structure of indoor environments: A constructive machine learning approach. *Auton. Robot.* **2019**, *43*, 813–835. [CrossRef]
83. Nauata, N.; Chang, K.H.; Cheng, C.Y.; Mori, G.; Furukawa, Y. House-GAN: Relational Generative Adversarial Networks for Graph-Constrained House Layout Generation. In Proceedings of the Computer Vision—ECCV 2020, Glasgow, UK, 23–28 August 2020; Vedaldi, A., Bischof, H., Brox, T., Frahm, J.M., Eds.; Springer International Publishing: Cham, Switzerland, 2020; Volume 12346, pp. 162–177. [CrossRef]
84. Paudel, A.; Dhakal, R.; Bhattarai, S. Room Classification on Floor Plan Graphs using Graph Neural Networks. *arXiv* **2021**. http://arxiv.org/abs/2108.05947.
85. Wang, Z.; Sacks, R.; Yeung, T. Exploring graph neural networks for semantic enrichment: Room-type classification. *Autom. Constr.* **2022**, *134*, 104039. [CrossRef]
86. Ahmed, S.; Liwicki, M.; Weber, M.; Dengel, A. Automatic Room Detection and Room Labeling from Architectural Floor Plans. In Proceedings of the 2012 10th IAPR International Workshop on Document Analysis Systems, Gold Coast, QLD, Australia, 27–29 March 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 339–343. [CrossRef]
87. Heras, L.P.d.l.; Ahmed, S.; Liwicki, M.; Valveny, E.; Sánchez, G. Statistical segmentation and structural recognition for floor plan interpretation - Notation invariant structural element recognition. *IJDAR* **2014**, *17*, 221–237. [CrossRef]
88. Liu, Z.; Von Wichert, G. Extracting semantic indoor maps from occupancy grids. *Robot. Auton. Syst.* **2014**, *62*, 663–674. [CrossRef]

89. Camozzato, D.; Dihl, L.; Silveira, I.; Marson, F.; Musse, S. Procedural floor plan generation from building sketches. *Vis. Comput.* **2015**, *31*, 753–763. [CrossRef]
90. Goncu, C.; Madugalla, A.; Marinai, S.; Marriott, K. Accessible On-Line Floor Plans. In Proceedings of the 24th International Conference on World Wide Web, International World Wide Web Conferences Steering Committee, WWW'15, Florence, Italy, 18–22 May 2015; pp. 388–398. [CrossRef]
91. de las Heras, L.P.; Terrades, O.R.; Lladós, J. Attributed Graph Grammar for floor plan analysis. In Proceedings of the 2015 13th International Conference on Document Analysis and Recognition (ICDAR), Tunis, Tunisia, 23–26 August 2015; pp. 726–730. [CrossRef]
92. Sharma, D.; Chattopadhyay, C.; Harit, G. A unified framework for semantic matching of architectural floorplans. In Proceedings of the 2016 23rd International Conference on Pattern Recognition (ICPR), Cancun, Mexico, 4–8 December 2016; pp. 2422–2427. [CrossRef]
93. Zhang, X.; Wong, A.K.S.; Lea, C.T. Automatic Floor Plan Analysis for Adaptive Indoor Wi-Fi Positioning System. In Proceedings of the 2016 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 15–17 December 2016; pp. 869–874. [CrossRef]
94. Dodge, S.; Xu, J.; Stenger, B. Parsing floor plan images. In Proceedings of the 2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA), Nagoya, Japan, 8–12 May 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 358–361. [CrossRef]
95. Madugalla, A.; Marriott, K.; Marinai, S. Partitioning Open Plan Areas in Floor Plans. In Proceedings of the 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), Kyoto, Japan, 9–15 November 2017; Volume 1, pp. 47–52, ISSN 2379–2140. [CrossRef]
96. Goyal, S.; Chattopadhyay, C.; Bhatnagar, G. Plan2Text: A framework for describing building floor plan images from first person perspective. In Proceedings of the Proceedings—2018 IEEE 14th International Colloquium on Signal Processing and its Application, CSPA 2018, Penang, Malaysia, 9–10 March 2018; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2018; pp. 35–40. [CrossRef]
97. Sharma, D.; Chattopadhyay, C. High-level feature aggregation for fine-grained architectural floor plan retrieval. *IET Comput. Vis.* **2018**, *12*, 702–709. [CrossRef]
98. Yamasaki, T.; Zhang, J.; Takada, Y. Apartment Structure Estimation Using Fully Convolutional Networks and Graph Model. In Proceedings of the 2018 ACM Workshop on Multimedia for Real Estate Tech, Association for Computing Machinery, RETech'18, Yokohama, Japan, 11 June 2018; pp. 1–6. [CrossRef]
99. Goyal, S.; Bhavsar, S.; Patel, S.; Chattopadhyay, C.; Bhatnagar, G. Sugaman: Describing floor plans for visually impaired by annotation learning and proximity-based grammar. *IET Image Process.* **2019**, *13*, 2623–2635. [CrossRef]
100. Kalervo, A.; Ylioinas, J.; Häikiö, M.; Karhu, A.; Kannala, J. CubiCasa5K: A Dataset and an Improved Multi-Task Model for Floorplan Image Analysis. *arXiv* **2019**. Available online: http://arxiv.org/abs/1904.01920 (accessed on 28 April 2024).
101. Zeng, Z.; Li, X.; Yu, Y.K.; Fu, C.W. Deep Floor Plan Recognition Using a Multi-Task Network with Room-Boundary-Guided Attention. *arXiv* **2019**. Available online: http://arxiv.org/abs/1908.11025 (accessed on 28 April 2024).
102. Mewada, H.K.; Patel, A.V.; Chaudhari, J.P.; Mahant, K.K.; Vala, A. Automatic room information retrieval and classification from floor plan using linear regression model. *IJDAR* **2020**, *23*, 253–266. [CrossRef]
103. Madugalla, A.; Marriott, K.; Marinai, S.; Capobianco, S.; Goncu, C. Creating Accessible Online Floor Plans for Visually Impaired Readers. *ACM Trans. Access. Comput.* **2020**, *13*, 1–37. [CrossRef]
104. Dong, S.; Wang, W.; Li, W.; Zou, K. Vectorization of floor plans based on EdgeGAN. *Information* **2021**, *12*, 206. [CrossRef]
105. Foroughi, F.; Wang, J.; Nemati, A.; Chen, Z.; Pei, H. MapSegNet: A Fully Automated Model Based on the Encoder-Decoder Architecture for Indoor Map Segmentation. *IEEE Access* **2021**, *9*, 101530–101542. [CrossRef]
106. Gan, Y.; Wang, S.Y.; Huang, C.E.; Hsieh, Y.C.; Wang, H.Y.; Lin, W.H.; Chong, S.N.; Liong, S.T. How Many Bedrooms Do You Need? A Real-Estate Recommender System from Architectural Floor Plan Images. *Sci. Program.* **2021**, *2021*, 9914557. [CrossRef]
107. Goyal, S.; Chattopadhyay, C.; Bhatnagar, G. Knowledge-driven description synthesis for floor plan interpretation. *IJDAR* **2021**, *24*, 19–32. [CrossRef]
108. Lu, Z.; Wang, T.; Guo, J.; Meng, W.; Xiao, J.; Zhang, W.; Zhang, X. Data-driven floor plan understanding in rural residential buildings via deep recognition. *Inf. Sci.* **2021**, *567*, 58–74. [CrossRef]
109. Murugan, G.; Moyal, V.; Nandankar, P.; Pandithurai, O.; Pimo, E.S.J. A novel CNN method for the accurate spatial data recovery from digital images. *Mater. Today Proc.* **2021**, *80*, 1706–1712. [CrossRef]
110. Park, S.; Kim, H. 3dplannet: Generating 3D models from 2d floor plan images using ensemble methods. *Electronics* **2021**, *10*, 2729. [CrossRef]
111. Lv, X.; Zhao, S.; Yu, X.; Zhao, B. Residential floor plan recognition and reconstruction. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 16712–16721. ISSN 2575–7075. [CrossRef]
112. Ahmed, S.; Liwicki, M.; Weber, M.; Dengel, A. Improved Automatic Analysis of Architectural Floor Plans. In Proceedings of the 2011 International Conference on Document Analysis and Recognition, ICDAR 2011, Beijing, China, 18–21 September 2011; IEEE Computer Society: Washington, DC, USA, 2011; pp. 864–869. [CrossRef]
113. Liu, Z.; Chen, D.; Von Wichert, G. 2D Semantic Mapping on Occupancy Grids. In Proceedings of the German Conference on Robotics, Munich, Germany, 21–22 May 2012; p. 6.

114. Ahmed, S.; Weber, M.; Liwicki, M.; Langenhan, C.; Dengel, A.; Petzold, F. Automatic analysis and sketch-based retrieval of architectural floor plans. *Pattern Recognit. Lett.* **2014**, *35*, 91–100. [CrossRef]
115. Paladugu, A.; Tian, Q.; Maguluri, H.; Li, B. Towards building an automated system for describing indoor floor maps for individuals with visual impairment. *Cyber-Phys. Syst.* **2016**, *1*, 132–159. [CrossRef]
116. Gimenez, L.; Robert, S.; Suard, F.; Zreik, K. Automatic reconstruction of 3D building models from scanned 2D floor plans. *Autom. Constr.* **2016**, *63*, 48–56. [CrossRef]
117. Liu, C.; Wu, J.; Kohli, P.; Furukawa, Y. Raster-to-Vector: Revisiting Floorplan Transformation. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2017; pp. 2214–2222. ISSN 15505499. [CrossRef]
118. Sandelin, F.; Sjöberg, K. *Semantic and Instance Segmentation of Room Features in Floor Plans using Mask R-CNN*; Uppsala University, Department of Information Technology: Uppsala, Sweden, 2019.
119. Goyal, S.; Mistry, V.; Chattopadhyay, C.; Bhatnagar, G. BRIDGE: Building Plan Repository for Image Description Generation, and Evaluation. In Proceedings of the 2019 International Conference on Document Analysis and Recognition (ICDAR), Sydney, NSW, Australia, 20–25 September 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1071–1076. [CrossRef]
120. Jang, H.; Yu, K.; Yang, J. Indoor reconstruction from floorplan images with a deep learning approach. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 65. [CrossRef]
121. Kim, S.; Park, S.; Kim, H.; Yu, K. Deep Floor Plan Analysis for Complicated Drawings Based on Style Transfer. *J. Comput. Civ. Eng.* **2020**, *35*
[CrossRef]
122. Surikov, I.; Nakhatovich, M.; Belyaev, S.; Savchuk, D. Floor plan recognition and vectorization using combination unet, faster-rcnn, statistical component analysis and ramer-douglas-peucker. In Proceedings of the Computing Science, Communication and Security, Gujarat, India, 26–27 March 2020; Springer: Singapore, 2020; Volume 1235, pp. 16–28. ISBN 9789811566479. [CrossRef]
123. Zhang, Y.; He, Y.; Zhu, S.; Di, X. The Direction-Aware, Learnable, Additive Kernels and the Adversarial Network for Deep Floor Plan Recognition. *arXiv* **2020**. Available online: http://arxiv.org/abs/2001.11194 (accessed on 28 April 2024).
124. Liu, Z.; von Wichert, G. A Generalizable Knowledge Framework for Semantic Indoor Mapping Based on Markov Logic Networks and Data Driven MCMC. *arXiv* **2020**. Available online: http://arxiv.org/abs/2002.08402 (accessed on 28 April 2024). [CrossRef]
125. Kim, H.; Kim, S.; Yu, K. Automatic Extraction of Indoor Spatial Information from Floor Plan Image: A Patch-Based Deep Learning Methodology Application on Large-Scale Complex Buildings. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 828. [CrossRef]
126. Song, J.; Yu, K. Framework for indoor elements classification via inductive learning on floor plan graphs. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 97. [CrossRef]
127. Hellbach, S.; Himstedt, M.; Bahrmann, F.; Riedel, M.; Villmann, T.; Böhme, H.J. Some Room for GLVQ: Semantic Labeling of Occupancy Grid Maps. In Proceedings of the Advances in Self-Organizing Maps and Learning Vector Quantization—Proceedings of the 10th International Workshop, WSOM 2014, Mittweida, Germany, 2–4 July 2014; Villmann, T., Schleif, F.M., Kaden, M., Lange, M., Eds.; Advances in Intelligent Systems and Computing; Springer: Berlin/Heidelberg, Germany, 2014; Volume 295, pp. 133–143. [CrossRef]
128. Luperto, M.; Quattrini Li, A.; Amigoni, F. A system for building semantic maps of indoor environments exploiting the concept of building typology. In *RoboCup 2013: Robot World Cup XVII*; Springer: Berlin/Heidelberg, Germany, 2014 Volume 8371, pp. 504–515. ISBN 9783662444672. [CrossRef]
129. Fermin-Leon, L.; Neira, J.; Castellanos, J. TIGRE: Topological graph based robotic exploration. In Proceedings of the 2017 European Conference on Mobile Robots, ECMR 2017, Paris, France, 6–8 September 2017; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2017. [CrossRef]
130. Kakuma, D.; Tsuichihara, S.; Ricardez, G.; Takamatsu, J.; Ogasawara, T. Alignment of Occupancy Grid and Floor Maps Using Graph Matching. In Proceedings of the Proceedings—IEEE 11th International Conference on Semantic Computing, ICSC 2017, San Diego, CA, USA, 30 January–1 February 2017; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2017; pp. 57–60. [CrossRef]
131. Kleiner, A.; Baravalle, R.; Kolling, A.; Pilotti, P.; Munich, M. A solution to room-by-room coverage for autonomous cleaning robots. In Proceedings of the IEEE International Conference on Intelligent Robots and Systems, Vancouver, BC, Canada, 24–28 September 2017; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2017; pp. 5346–5352. ISSN 21530858. [CrossRef]
132. Fermin-Leon, L.; Neira, J.; Castellanos, J.A. Incremental contour-based topological segmentation for robot exploration. In Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore, 29 May–3 June 2017; pp. 2554–2561. [CrossRef]
133. Hang, M.; Lin, M.; Li, S.; Chen, Z.; Ding, R. A multi-strategy path planner based on space accessibility. In Proceedings of the 2017 IEEE International Conference on Robotics and Biomimetics, ROBIO 2017, Macau, Macao, 5–8 December 2017; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2017; pp. 2154–2161. [CrossRef]
134. Liu, B.; Zuo, L.; Zhang, C.H.; Liu, Y. An approach to graph-based grid map segmentation for robot global localization. In Proceedings of the 2018 IEEE International Conference on Mechatronics and Automation, ICMA 2018, Changchun, China, 5–8 August 2018; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2018; pp. 1812–1817. [CrossRef]

135. Mielle, M.; Magnusson, M.; Lilienthal, A. A method to segment maps from different modalities using free space layout maoris: Map of ripples segmentation. In Proceedings of the Proceedings—IEEE International Conference on Robotics and Automation, Brisbane, QLD, Australia, 21–25 May 2018; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2017; pp. 4993–4999. ISSN 10504729. [CrossRef]
136. Hiller, M.; Qiu, C.; Particke, F.; Hofmann, C.; Thielecke, J. Learning Topometric Semantic Maps from Occupancy Grids. In Proceedings of the IEEE International Conference on Intelligent Robots and Systems, Macau, China, 3–8 November 2019; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2017; pp. 4190–4197. ISSN 21530858. [CrossRef]
137. Luperto, M.; Arcerito, V.; Amigoni, F. Predicting the layout of partially observed rooms from grid maps. In Proceedings of the Proceedings—IEEE International Conference on Robotics and Automation, Montreal, QC, Canada, 20–24 May 2019; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2017; pp. 6898–6904. ISSN: 10504729, [CrossRef]
138. Schwertfeger, S.; Yu, T. Room Detection for Topological Maps. *arXiv* **2019**. Available online: http://arxiv.org/abs/1912.01279 (accessed on 28 April 2024).
139. Hou, J.; Yuan, Y.; Schwertfeger, S. Area Graph: Generation of Topological Maps using the Voronoi Diagram. In Proceedings of the 2019 19th International Conference on Advanced Robotics (ICAR), Belo Horizonte, Brazil, 2–6 December 2019; pp. 509–515. [CrossRef]
140. Tien, M.; Park, Y.; Jung, K.H.; Kim, S.Y.; Kye, J.E. Performance evaluation on the accuracy of the semantic map of an autonomous robot equipped with P2P communication module. *Peer-to-Peer Netw. Appl.* **2020**, *13*, 704–716. [CrossRef]
141. Zheng, T.; Duan, Z.; Wang, J.; Lu, G.; Li, S.; Yu, Z. Research on Distance Transform and Neural Network Lidar Information Sampling Classification-Based Semantic Segmentation of 2D Indoor Room Maps. *Sensors* **2020**, *21*, 1365. [CrossRef]
142. Luperto, M.; Kucner, T.P.; Tassi, A.; Magnusson, M.; Amigoni, F. Robust Structure Identification and Room Segmentation of Cluttered Indoor Environments from Occupancy Grid Maps. *arXiv* **2022**. Available online: http://arxiv.org/abs/2203.03519 (accessed on 28 April 2024).
143. Shi, L.; Kodagoda, S.; Dissanayake, G. Application of semi-supervised learning with Voronoi Graph for place classification. In Proceedings of the IEEE International Conference on Intelligent Robots and Systems, Vilamoura-Algarve, Portugal, 7–12 October 2012; pp. 2991–2996. ISSN 21530858. [CrossRef]
144. Sjoo, K. Semantic map segmentation using function-based energy maximization. In Proceedings of the 2012 IEEE International Conference on Robotics and Automation, Saint Paul, MN, USA, 14–18 May 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 4066–4073. [CrossRef]
145. Capobianco, R.; Gemignani, G.; Bloisi, D.; Nardi, D.; Iocchi, L. Automatic extraction of structural representations of environments. In *Intelligent Autonomous Systems 13*; Springer: Cham, Switzerland, 2015; Volume 302, pp. 721–733. ISBN 9783319083377. [CrossRef]
146. Liu, M.; Colas, F.; Oth, L.; Siegwart, R. Incremental topological segmentation for semi-structured environments using discretized GVG. *Auton. Robot.* **2015**, *38*, 143–160. [CrossRef]
147. Goeddel, R.; Olson, E. Learning semantic place labels from occupancy grids using CNNs. In Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Daejeon, Republic of Korea, 9–14 October 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 3999–4004. [CrossRef]
148. Hou, J.; Kuang, H.; Schwertfeger, S. Fast 2D map matching based on area graphs. In Proceedings of the IEEE International Conference on Robotics and Biomimetics, ROBIO 2019, Dali, China, 6–8 December 2019; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2019; pp. 1723–1729. [CrossRef]
149. Rubio, F.; Martínez-Gómez, J.; Flores, M.J.; Puerta, J.M. Comparison between Bayesian network classifiers and SVMs for semantic localization. *Expert Syst. Appl.* **2016**, *64*, 434–443. [CrossRef]
150. Ursic, P.; Mandeljc, R.; Leonardis, A.; Kristan, M. Part-based room categorization for household service robots. In Proceedings of the Proceedings—IEEE International Conference on Robotics and Automation, Stockholm, Sweden, 16–21 May 2016; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2019; pp. 2287–2294. ISSN 10504729. [CrossRef]
151. Fleer, D. Human-Like Room Segmentation for Domestic Cleaning Robots. *Robotics* **2017**, *6*, 35. [CrossRef]
152. Young, J.; Basile, V.; Suchi, M.; Kunze, L.; Hawes, N.; Vincze, M.; Caputo, B. Making Sense of Indoor Spaces Using Semantic Web Mining and Situated Robot Perception. In Proceedings of the The Semantic Web: ESWC 2017 Satellite Events, Portorož, Slovenia, 28 May–1 June 2017; Springer: Cham, Switzerland, 2017; Volume 10577, pp. 299–313. [CrossRef]
153. Pintore, G.; Ganovelli, F.; Pintus, R.; Scopigno, R.; Gobbetti, E. 3D floor plan recovery from overlapping spherical images. *Comp. Visual Media* **2018**, *4*, 367–383. [CrossRef]
154. Pintore, G.; Ganovelli, F.; Pintus, R.; Scopigno, R.; Gobbetti, E. Recovering 3D Indoor Floor Plans by Exploiting Low-cost Spherical Photography. In *Pacific Graphics Short Papers*; The Eurographics Association: Eindhoven, The Netherlands, 2018. [CrossRef]
155. Othman, K.; Rad, A. An indoor room classification system for social robots via integration of CNN and ECOC. *Appl. Sci.* **2019**, *9*, 470. [CrossRef]
156. Balaska, V.; Bampis, L.; Boudourides, M.; Gasteratos, A. Unsupervised semantic clustering and localization for mobile robotics tasks. *Robot. Auton. Syst.* **2020**, *131*, 103567. [CrossRef]
157. Sadeghi, F.; Tappen, M.F. Latent Pyramidal Regions for Recognizing Scenes. In Proceedings of the Computer Vision—ECCV 2012, Florence, Italy, 7–13 October 2012; Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2021; pp. 228–241. [CrossRef]

158. Erkent, O.; Bozma, I. Place representation in topological maps based on bubble space. In Proceedings of the 2012 IEEE International Conference on Robotics and Automation, Saint Paul, MN, USA, 14–18 May 2012; pp. 3497–3502. ISSN 1050-4729. [CrossRef]
159. Ranganathan, A. PLISS: Labeling places using online changepoint detection. *Auton Robot* **2012**, *32*, 351–368. [CrossRef]
160. Parizi, S.N.; Oberlin, J.G.; Felzenszwalb, P.F. Reconfigurable models for scene recognition. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 2775–2782. [CrossRef]
161. Sadovnik, A.; Chen, T. Hierarchical object groups for scene classification. In Proceedings of the 2012 19th IEEE International Conference on Image Processing, Orlando, FL, USA, 30 September–3 October 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 1881–1884. [CrossRef]
162. Mozos, O.; Mizutani, H.; Kurazume, R.; Hasegawa, T. Categorization of Indoor Places Using the Kinect Sensor. *Sensors* **2012**, *12*, 6695–711. [CrossRef] [PubMed]
163. Juneja, M.; Vedaldi, A.; Jawahar, C.; Zisserman, A. Blocks That Shout: Distinctive Parts for Scene Classification. In Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; IEEE: Piscataway, NJ, USA, 2013; pp. 923–930. [CrossRef]
164. Margolin, R.; Zelnik-Manor, L.; Tal, A. OTC: A Novel Local Descriptor for Scene Classification. In Proceedings of the Computer Vision—ECCV 2014, Zurich, Switzerland, 6–12 September 2014; Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2014; pp. 377–391. [CrossRef]
165. Zuo, Z.; Wang, G.; Shuai, B.; Zhao, L.; Yang, Q.; Jiang, X. Learning Discriminative and Shareable Features for Scene Classification. In Proceedings of the Computer Vision—ECCV 2014, Zurich, Switzerland, 6–12 September 2014; Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2014; pp. 552–568. [CrossRef]
166. Jie, Z.; Yan, S. Robust Scene Classification with Cross-Level LLC Coding on CNN Features. In Proceedings of the Computer Vision—ACCV 2014, Singapore, 1–5 November 2014; Cremers, D., Reid, I., Saito, H., Yang, M.H., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2014; pp. 376–390. [CrossRef]
167. Mesnil, G.; Rifai, S.; Bordes, A.; Glorot, X.; Bengio, Y.; Vincent, P. Unsupervised Learning of Semantics of Object Detections for Scene Categorization. In Proceedings of the Pattern Recognition Applications and Methods, Lisbon, Portugal, 10–12 January 2015; Fred, A., De Marsico, M., Eds.; Advances in Intelligent Systems and Computing; Springer International Publishing: Cham, Switzerland, 2014; pp. 209–224. [CrossRef]
168. Zhou, B.; Lapedriza, A.; Xiao, J.; Torralba, A.; Oliva, A. Learning Deep Features for Scene Recognition using Places Database. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 1–9.
169. Dixit, M.; Chen, S.; Gao, D.; Rasiwasia, N.; Vasconcelos, N. Scene classification with semantic Fisher vectors. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 2974–2983. [CrossRef]
170. Pintore, G.; Garro, V.; Ganovelli, F.; Gobbetti, E.; Agus, M. Omnidirectional image capture on mobile devices for fast automatic generation of 2.5D indoor maps. In Proceedings of the 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Placid, NY, USA, 7–10 March 2016; pp. 1–9. [CrossRef]
171. Cruz, E.; Rangel, J.C.; Gomez-Donoso, F.; Bauer, Z.; Cazorla, M.; Garcia-Rodriguez, J. Finding the Place: How to Train and Use Convolutional Neural Networks for a Dynamically Learning Robot. In Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil, 8–13 July 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1–8. [CrossRef]
172. Simonsen, C.; Thiesson, F.; Philipsen, M.; Moeslund, T. GENERALIZING FLOOR PLANS USING GRAPH NEURAL NETWORKS. In Proceedings of the Proceedings—International Conference on Image Processing, ICIP, Anchorage, AK, USA, 19–22 September 2021; IEEE Computer Society: Washington, DC, USA; pp. 654–658, ISSN 15224880, [CrossRef]
173. Wei, Q.; Wei, Q.; Liu, Y.; Guan, Q.; Liu, D. Data-driven room classification for office buildings based on echo state network. In Proceedings of the the 27th Chinese Control and Decision Conference (2015 CCDC), Qingdao, China, 23–25 May 2015; pp. 2602–2607. ISSN: 1948-9447, [CrossRef]
174. Shi, G.; Zhao, B.; Li, C.; Wei, Q.; Liu, D. An echo state network based approach to room classification of office buildings. *Neurocomputing* **2019**, *333*, 319–328. [CrossRef]
175. Uršič, P.; Kristan, M.; Skočaj, D.; Leonardis, A. Room classification using a hierarchical representation of space. In Proceedings of the 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, Vilamoura-Algarve, Portugal, 7–12 October 2012; pp. 1371–1378. ISSN 2153-0866. [CrossRef]
176. Turner, E.; Zakhor, A. Floor plan generation and room labeling of indoor environments from laser range data. In Proceedings of the 2014 International Conference on Computer Graphics Theory and Applications (GRAPP), Lisbon, Portugal, 5–8 January 2014; pp. 1–12.
177. Turner, E.; Zakhor, A. Multistory floor plan generation and room labeling of building interiors from laser range data. In Proceedings of the Computer Vision, Imaging and Computer Graphics—Theory and Applications, Berlin, Germany, 11–14 March 2015; Springer: Berlin/Heidelberg, Germany, 2015; Volume 550, pp. 29–44. ISBN 9783319251165. [CrossRef]
178. Ursic, P.; Leonardis, A.; Skocaj, D.; Kristan, M. Hierarchical spatial model for 2D range data based room categorization. In Proceedings of the Proceedings—IEEE International Conference on Robotics and Automation, Stockholm, Sweden, 16–21 May 2016; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2016; pp. 4514–4521. ISSN 10504729. [CrossRef]

179. He, X.; Liu, H.; Huang, W. Room categorization using local receptive fields-based extreme learning machine. In Proceedings of the 2017 2nd International Conference on Advanced Robotics and Mechatronics, ICARM 2017, Hefei and Tai'an, China, 27–31 August 2017; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2016; pp. 620–625. [CrossRef]
180. Wu, H.; Tian, G.h.; Li, Y.; Zhou, F.y.; Duan, P. Spatial semantic hybrid map building and application of mobile service robot. *Robot. Auton. Syst.* **2014**, *62*, 923–941. [CrossRef]
181. Hardegger, M.; Roggen, D.; Tröster, G. 3D ActionSLAM: Wearable Person Tracking in Multi-Floor Environments. *Pers. Ubiquit. Comput.* **2015**, *19*, 123–141. [CrossRef]
182. Rojas Castro, D.; Revel, A.; Ménard, M. Document image analysis by a mobile robot for autonomous indoor navigation. In Proceedings of the 2015 13th International Conference on Document Analysis and Recognition (ICDAR), Tunis, Tunisia, 23–26 August 2015; pp. 156–160. [CrossRef]
183. Loch-Dehbi, S.; Dehbi, Y.; Gröger, G.; Plümer, L. Prediction of Building Floorplans Using Logical and Stochastic Reasoning Based on Sparse Observations. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.* **2016**, *4*, 265–270. [CrossRef]
184. Dehbi, Y.; Loch-Dehbi, S.; Plümer, L. Parameter Estimation and Model Selection for Indoor Environments Based on Sparse Observations. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.* **2017**, *4*, 303–310. [CrossRef]
185. Loch-Dehbi, S.; Dehbi, Y.; Pl mer, L. Estimation of 3D indoor models with constraint propagation and stochastic reasoning in the absence of indoor measurements. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 90. [CrossRef]
186. Dehbi, Y.; Gojayeva, N.; Pickert, A.; Haunert, J.H.; Plümer, L. Room shapes and functional uses predicted from sparse data. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.* **2018**, *4*, 33–40. [CrossRef]
187. Shahbandi, S.G.; Magnusson, M.; Iagnemma, K. Nonlinear Optimization of Multimodal Two-Dimensional Map Alignment With Application to Prior Knowledge Transfer. *IEEE Robot. Autom. Lett.* **2018**, *3*, 2040–2047. [CrossRef]
188. Hu, X.; Fan, H.; Noskov, A.; Zipf, A.; Wang, Z.; Shang, J. Feasibility of using grammars to infer room semantics. *Remote Sens.* **2019**, *11*, 1535. [CrossRef]
189. Zhou, R.; Lu, X.; Zhao, H.S.; Fu, Y.; Tang, M.J. Automatic Construction of Floor Plan with Smartphone Sensorsb. *J. Electron. Sci. Technol.* **2019**, *17*, 13–25. [CrossRef]
190. Pronobis, A.; Jensfelt, P. Large-scale semantic mapping and reasoning with heterogeneous modalities. In Proceedings of the 2012 IEEE International Conference on Robotics and Automation, Saint Paul, MN, USA, 4–18 May 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 3515–3522. [CrossRef]
191. Kostavelis, I.; Charalampous, K.; Gasteratos, A. Online Spatiotemporal-Coherent Semantic Maps for Advanced Robot Navigation. In Proceedings of the 5th Workshop on Planning, Perception and Navigation for Intelligent Vehicles, in Conjunction with the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Tokyo, Japan, 3–8 November 2013.
192. Hemachandra, S.; Walter, M.R.; Tellex, S.; Teller, S. Learning spatial-semantic representations from natural language descriptions and scene classifications. In Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, China, 31 May–7 June 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 2623–2630. [CrossRef]
193. Sünderhauf, N.; Dayoub, F.; McMahon, S.; Talbot, B.; Schulz, R.; Corke, P.; Wyeth, G.; Upcroft, B.; Milford, M. Place Categorization and Semantic Mapping on a Mobile Robot. *arXiv* **2015**. Available online: http://arxiv.org/abs/1507.02428 (accessed on 28 April 2024).
194. Kostavelis, I.; Charalampous, K.; Gasteratos, A.; Tsotsos, J.K. Robot navigation via spatial and temporal coherent semantic maps. *Eng. Appl. Artif. Intell.* **2016**, *48*, 173–187. [CrossRef]
195. Kostavelis, I.; Gasteratos, A. Semantic maps from multiple visual cues. *Expert Syst. Appl.* **2017**, *68*, 45–57. [CrossRef]
196. Liu, M.; Chen, R.; Li, D.; Chen, Y.; Guo, G.; Cao, Z.; Pan, Y. Scene Recognition for Indoor Localization Using a Multi-Sensor Fusion Approach. *Sensors* **2017**, *17*, 2847. [CrossRef]
197. Luo, R.C.; Chiou, M. Hierarchical Semantic Mapping Using Convolutional Neural Networks for Intelligent Service Robotics. *IEEE Access* **2018**, *6*, 61287–61294. [CrossRef]
198. Jin, C.; Elibol, A.; Zhu, P.; Chong, N.Y. Semantic Mapping Based on Image Feature Fusion in Indoor Environments. In Proceedings of the 2021 21st International Conference on Control, Automation and Systems (ICCAS), Jeju, Republic of Korea, 12–15 October 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 693–698. [CrossRef]
199. Schäfer, J. Practical concerns of implementing machine learning algorithms for W-LAN location fingerprinting. In Proceedings of the 2014 6th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), St. Petersburg, Russia, 6–8 October 2014; pp. 310–317. ISSN 2157-023X, [CrossRef]
200. Laska, M.; Blankenbach, J.; Klamma, R. Adaptive indoor area localization for perpetual crowdsourced data collection. *Sensors* **2020**, *20*, 1443. [CrossRef]
201. Peters, N.; Lei, H.; Friedland, G. Name That Room: Room Identification Using Acoustic Features in a Recording. In Proceedings of the 20th ACM International Conference on Multimedia. Association for Computing Machinery, MM'12, Nara, Japan, 29 October–2 November 2012; pp. 841–844. [CrossRef]
202. Song, Q.; Gu, C.; Tan, R. Deep Room Recognition Using Inaudible Echos. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2018**, *2*, 135. [CrossRef]

203. Au-Yeung, J.; Banavar, M.K.; Vanitha, M. Room Classification using Acoustic Signals. In Proceedings of the 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), Vellore, India, 24–25 February 2020; pp. 1–4. [CrossRef]
204. Resuli, N.; Skubic, M.; Kovaleski, S. Learning Room Structure and Activity Patterns Using RF Sensing for In-Home Monitoring of Older Adults. In Proceedings of the 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Seoul, Republic of Korea, 16–19 December 2020; pp. 2054–2061. [CrossRef]
205. Dziwis, D.; Zimmermann, S.; Lübeck, T.; Arend, J.M.; Bau, D.; Pörschmann, C. Machine Learning-Based Room Classification for Selecting Binaural Room Impulse Responses in Augmented Reality Applications. In Proceedings of the 2021 Immersive and 3D Audio: From Architecture to Automotive (I3DA), Bologna, Italy, 8–10 September 2021; pp. 1–8. [CrossRef]
206. Walter, M.; Hemachandra, S.; Homberg, B.; Tellex, S.; Teller, S. Learning Semantic Maps from Natural Language Descriptions. In Proceedings of the Robotics: Science and Systems IX; Robotics: Science and Systems Foundation, Berlin, Germany, 24–28 June 2013. [CrossRef]
207. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; IEEE: Piscataway, NJ, USA, 2026; pp. 770–778. [CrossRef]
208. Wang, Y.; Funk, N.; Ramezani, M.; Papatheodorou, S.; Popovic, M.; Camurri, M.; Leutenegger, S.; Fallon, M. Elastic and Efficient LiDAR Reconstruction for Large-Scale Exploration Tasks. In Proceedings of the 2021 IEEE International Conference on Robotics and Automation (ICRA), Xi'an, China, 30 May–5 June 2021; IEEE: Piscataway, NJ, USA, 2016; pp. 5035–5041. [CrossRef]
209. Carrera, J.L.; Li, Z.; Zhao, Z.; Braun, T.; Neto, A. A Real-Time Indoor Tracking System in Smartphones. In Proceedings of the Proceedings of the 19th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems. Association for Computing Machinery, MSWiM'16, Malta, Malta, 13–17 November 2016; pp. 292–301. [CrossRef]
210. Coughlan, J.; Yuille, A. Manhattan World: Compass direction from a single image by Bayesian inference. In Proceedings of the Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 20–27 September 1999; IEEE: Piscataway, NJ, USA, 2016; Volume 2, pp. 941–947 [CrossRef]
211. Frasconi, P.; Costa, F.; De Raedt, L.; De Grave, K. kLog: A Language for Logical and Relational Learning with Kernels. *Artif. Intell.* **2014**, *217*, 117–143. [CrossRef]
212. Hamilton, W.L.; Ying, R.; Leskovec, J. Inductive Representation Learning on Large Graphs. *arXiv* **2018**. [CrossRef]
213. Du, J.; Zhang, S.; Wu, G.; Moura, J.M.F.; Kar, S. Topology Adaptive Graph Convolutional Networks. *arXiv* **2018**. [CrossRef]
214. Kipf, T.N.; Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. *arXiv* **2017**. [CrossRef]
215. Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; Bengio, Y. Graph Attention Networks. *arXiv* **2018**. [CrossRef]
216. Tan, M.; Le, Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *arXiv* **2018**. Available online: http://arxiv.org/abs/1905.11946 (accessed on 28 April 2024).
217. Lin, K.S. Adaptive WiFi positioning system with unsupervised map construction. *Electron. Comput. Eng.* **2015**, b1514560. [CrossRef]
218. Ball, G.H.; Hall, D.J. *Isodata, a Novel Method of Data Analysis and Pattern Classification*; Stanford Research Institute: Menlo Park, CA, USA, 1965.
219. Roth, S.D. Ray casting for modeling solids. *Comput. Graph. Image Process.* **1982**, *18*, 109–144. [CrossRef]
220. Blondel, V.D.; Guillaume, J.L.; Lambiotte, R.; Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, *2008*, P10008. [CrossRef]
221. van Eck, N.J.; Waltman, L. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics* **2010**, *84*, 523–538. [CrossRef] [PubMed]
222. Bastian, M.; Heymann, S.; Jacomy, M. Gephi: An Open Source Software for Exploring and Manipulating Networks. In Proceedings of the International AAAI Conference on Web and Social Media, San Jose, CA, USA, 17–20 May 2009; Volume 3, pp. 361–362.
223. Zhu, B.; Fan, X.; Gao, X.; Xu, G.; Xie, J. A heterogeneous attention fusion mechanism for the cross-environment scene classification of the home service robot. *Robot. Auton. Syst.* **2024**, *173*, 104619. [CrossRef]
224. Yang, L.; Ye, J.; Zhang, Y.; Wang, L.; Qiu, C. A semantic SLAM-based method for navigation and landing of UAVs in indoor environments. *Knowl.-Based Syst.* **2024**, *293*, 111693. [CrossRef]
225. Shaharuddin, S.; Abdul Maulud, K.N.; Syed Abdul Rahman, S.A.F.; Che Ani, A.I.; Pradhan, B. The role of IoT sensor in smart building context for indoor fire hazard scenario: A systematic review of interdisciplinary articles. *Internet Things* **2023**, *22*, 100803. [CrossRef]
226. Mahmoud, M.; Chen, W.; Yang, Y.; Li, Y. Automated BIM generation for large-scale indoor complex environments based on deep learning. *Autom. Constr.* **2024**, *162*, 105376. [CrossRef]
227. Sommer, M.; Stjepandić, J.; Stobrawa, S.; Soden, M.V. Automated generation of digital twin for a built environment using scan and object detection as input for production planning. *J. Ind. Inf. Integr.* **2023**, *33*, 100462. [CrossRef]
228. Zheng, Y.; Xu, Y.; Shu, S.; Sarem, M. Indoor semantic segmentation based on Swin-Transformer. *J. Vis. Commun. Image Represent.* **2024**, *98*, 103991. [CrossRef]
229. Han, Y.; Zhou, Z.; Li, W.; Feng, J.; Wang, C. Exploring building component thermal storage performance for optimizing indoor thermal environment—A case study in Beijing. *Energy Build.* **2024**, *304*, 113834. [CrossRef]

230. Pachano, J.E.; Fernández-Vigil Iglesias, M.; Peppas, A.; Fernández Bandera, C. Enhancing self-consumption for decarbonization: An optimization strategy based on a calibrated building energy model. *Energy Build.* **2023**, *298*, 113576. [CrossRef]
231. Deng, M.; Fu, B.; Menassa, C.C.; Kamat, V.R. Learning-Based personal models for joint optimization of thermal comfort and energy consumption in flexible workplaces. *Energy Build.* **2023**, *298*, 113438. [CrossRef]
232. Roumi, S.; Zhang, F.; Stewart, R.A.; Santamouris, M. Indoor environment quality effects on occupant satisfaction and energy consumption: Empirical evidence from subtropical offices. *Energy Build.* **2024**, *303*, 113784. [CrossRef]
233. Sulaiman, M.H.; Mustaffa, Z. Using the evolutionary mating algorithm for optimizing the user comfort and energy consumption in smart building. *J. Build. Eng.* **2023**, *76*, 107139. [CrossRef]
234. Wei, Y.; Du, M.; Huang, Z. The effects of energy quota trading on total factor productivity and economic potential in industrial sector: Evidence from China. *J. Clean. Prod.* **2024**, *445*, 141227. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.