Sensor Fusion of 2D-LiDAR and 360-degree Camera Data for Room Layout Reconstruction

Fynn Pieper
Institute Systems Engineering for Future Mobility
German Aerospace Center (DLR)
Oldenburg, Germany
fynn.pieper@dlr.de

***Abstract*—Precise floor plans are the foundation for ensuring safety, accessibility, and efficiency within critical infrastructures. To meet this need, we introduce a method that seamlessly merges the accurate depth measurements from a 2D LiDAR with the nuanced data interpretation capabilities of artificial neural networks applied to RGB images. This allows our approach to capture fully furnished rooms with remarkable spatial accuracy and reliability from a single measurement. Central to our methodology is a novel two-stage filtering process. First, we harness geometric analysis to exclude extraneous LiDAR points, then refine this data using semantic segmentation, ensuring only measurements congruent with walls are retained. Walls are inferred through a systematic Hough transform and further validated using an innovative confidence metric. Using RGB data, we augment the wall estimation for room edge detection, allowing us to determine the azimuth position of room corners from the image, even when obscured by furniture. Subsequent data fusion yields layouts with an average 2D IoU of 90.47 % without relying on camera height assumption. By carefully leveraging the unique strengths of each data type, our approach significantly advances the field of room layout reconstruction (RLR). The achieved results hint at a considerable potential for real-world applications.**

***Index Terms*—critical infrastructure, sensor fusion, 360-degree camera, 2D LiDAR, room layout reconstruction**

I. Introduction

Floor plans have long served as indispensable tools in the architectural domain, visualizing the spatial layout of buildings in a top-down projection [1]. Historically, the creation and updating of these plans involved manual measurements and sketches, a labor-intensive and often error-prone process. Over time, with buildings frequently undergoing restructuring, expansions, or other alterations, many original floor plans inevitably lose their accuracy. This poses challenges, since maintaining a clear overview of a building's shape, dimensions, and the arrangement and location of its rooms is crucial. Clarity is particularly paramount for facilities critical to society such as schools, hospitals and detention centers [2].

An accurate floor plan plays a vital role in ensuring and enhancing the safety of people within the facility, as it clearly illustrates the locations of exits, fire safety equipment, and other safety features. In emergency scenarios, these plans enable first responders to rapidly pinpoint and access specific areas within the facility, helping to save lives and minimize damage. Technological advancements have propelled the research in accurate indoor room layout reconstruction (RLR), emphasizing both technical and societal importance.

II. Related Work

Generating layouts from established 3D spatial structures is central to contemporary architectural and spatial studies [3]. Different sensor modalities, notably LiDAR and cameras, have been at the forefront of this exploration [4]. While the precision and depth perception of LiDAR-based techniques are well-recognized, camera-based methods may offer higher resolution of data, but may cause spatial accuracy to degrade [5]. This section delves into the contemporary methodologies, examining the strengths and shortcomings associated with each sensor modality.

A. LiDAR Scanner-based Methods

Light Detection and Ranging (LiDAR) is a remote sensing method that uses light in the form of a pulsed laser to measure distances to a target. Generally, LiDAR point clouds provide high spatial accuracy and reliability, making them ideal for scenarios where precise spatial relationships and structural integrity are vital. For this reason, LiDAR-based techniques excel in applications requiring detailed spatial mapping and precise measurement, such as topographic mapping [6], forestry [7], urban planning [8], and, as detailed here, RLR [9]. The typical approach to LiDAR-based RLR starts out with a single scan, subdividing the point cloud into clusters [10]. Each cluster represents normally contains candidate walls but may include ceiling and floor when a 3D-LiDAR is used [11]. Then, corner points and line segments are deduced, indicating the potential structure of the room. However, transparent or moving surfaces and sensor tilting present unique challenges for accurate LiDAR-based RLR. Further, LiDAR scanners' difficulty in differentiating between clutter and walls can compromise RLR results, especially in furnished rooms where walls might be obscured by various objects [12], [13].

B. Perspective Camera-based Methods

Perspective cameras typically have a fixed field of view (FOV) of less than 120 degrees and can scan large areas of a room at a high resolution. A challenge arises when reducing a 3D scene to a 2D RGB image, as it entails a loss of depth information. While depth information is directly applicable to RLR, RGB data must be placed in a larger context to be used. As a result, two primary strategies have evolved to address the RLR: geometry-based and learning-based methods [14].

Geometry-based methods focus on identifying distinct image clues that correspond to geometric features such as corners, walls, floors or ceilings [15]. As a simplifying framework, numerous studies employ the Manhattan World assumption [16]–[19], which is suggesting that indoor scenes predominantly align with a Cartesian coordinate system. This is evident in the way most walls, ceilings, and floors are oriented along orthogonal planes and helps to relate clues to corresponding geometric features. Building on this idea, Hoiem et al. [20] categorized geometric classes by learning appearance-based models. Extending this model to indoor layouts, Lee [21] and Zhang [22] used Geometric Context (GC) to include prior knowledge and Orientation Maps (OM), depicting local beliefs in plane orientation. Beyond GC and OM, Del Pero et al. [15] further consider the room edges. Still, in most of the state-of-the-art, progress on beyond-Manhattan layouts is dismissed as hardly obtainable [19].

While geometry-based methods focus on salient room features, more recent techniques adopt machine learning to uncover clues concealed from human understanding. Learning-based methods typically utilize artificial neural networks (ANNs) for supervised [16], [23] and semi-supervised [24], [25] learning. ANNs allow both local and global features to be analyzed for more robust reconstructions even with local occlusions or faint room edges [26]. Contextual information can be encoded using classifiers or probabilistic graphical models for automatic classification [27], [28]. While being powerful in practice, learning-based methods have unpredictable applicability to new environments due to their heavy dependence on training data.

C. Panoramic Camera-based Methods

Panoramic cameras with 360° horizontal and 180° vertical FOV can capture entire scenes in an equirectangular projection (ERP) format. Some methods [18], [29], [30] have bridged the gap to panoramic images by projecting multiple perspective images onto a sphere and applying existing techniques. More recent approaches introduced the assessment of object positions [31] and generalization beyond cuboid layouts [32]. Concurrently, powerful ANNs have improved the recognition of both low- and high-level features, enhancing the interpretation of detailed features and global scene understanding [33].

Self-contained solutions, so-called end-to-end approaches, have seen significant progress within the past five years. LayoutNet [18] pioneered the end-to-end approach by applying a convolutional neural network to perform RLR directly on the ERP, improving both accuracy and robustness through redundant information. This approach also introduces a pre-alignment procedure, standardizing the orientation of the input image to simplify the learning task. DuLa-Net [33] extended this idea by including a projected ceiling view with aligned ERP, thereby enhancing estimation accuracy. Meanwhile, HorizonNet [34] introduced a faster solution, encoding room edges into one-dimensional vectors, thus minimizing the parameters to be learned and reducing computation time significantly. AtlantaNet [35] presented a unique innovation by being the first end-to-end approach to reconstruct room layouts beyond the Manhattan World. Finally, HoHoNet [36] and PanelNet [37] propose holistic approaches to RLR, integrating semantic segmentation and depth estimation. While the first encodes input data in the frequency domain to increase processing speed, the latter employs a consecutive vertical panel projection to analyze local and global context.

D. Sensor Fusion Methods

Despite the limitations of 2D-LiDAR and RGB data, the fusion of both remains an underexplored yet promising approach for RLR. Weiss and Zell's [38] work utilized a 2D-LiDAR and a panning perspective camera to create a 3D corridor model, albeit with significant LiDAR-based inaccuracies. Li et al. [39] mounted both devices on a panning device for detailed reconstructions but faced calibration issues. An alternative method by [40] prioritized RGB data analysis before LiDAR, aiding boundary extraction [41]. Remarkably, all methods integrating cameras and 2D-LiDAR scanners fail to capitalize on the potential of sensor fusion, thus being significantly inferior to RLR approaches based exclusively on panorama data. While the diverse nature of the sensor modalities impedes a direct comparison and risks overshadowing individual contributions [42], Table I collates the performance of all methods across various room layouts. LiDAR measurements remain crucial for scale retrieval, whereas panorama-based methods excel in shape identification.

In this study, we recognize the lack of development of sensor fusion based methods and contribute a pioneering method of sensor fusion, combining 2D LiDAR and panoramic RGB data. Unlike prior work, our approach leverages the strengths and mitigates the limitations of each sensor to enhance spatial accuracy in room layout reconstruction. This method, to our understanding, is the first attempt to systematically integrate these two sensor modalities for RLR.

TABLE I
CHARACTERISTICS OF ROOM LAYOUTS ACHIEVABLE BY THE COMBINED METHODS FROM THE STATE-OF-THE-ART WITH REGARDS TO DIFFERENT SENSOR MODALITIES, COMPARED TO OUR PROPOSED APPROACH.

	LiDAR	**Perspec. Camera**	**Panor. Camera**	**Sensor Fusion**	**Proposed Method**
Layouts:					
(Half) Cuboids	+	**O**	+	+	+
L-shaped rooms	+	**O**	+	**O**	+
Manhattan conform	**O**	-	+	+	+
Angled walls	**O**	-	**O**	-	+
Rounded walls	-	-	**O**	-	-
Special Cases:					
Clutter-resistant	-	**O**	+	**O**	+
Occlusion-handling	-	**O**	+	-	-
Context-awareness	-	**O**	+	-	+
Distortion-resistant	+	+	**O**	+	+
Dimensions:					
To scale	+	-*	-*	+	+

+ : Good results obtainable, **O** : Mediocre results, **-** : Not possible.
* Requires the assumption of a known camera height.

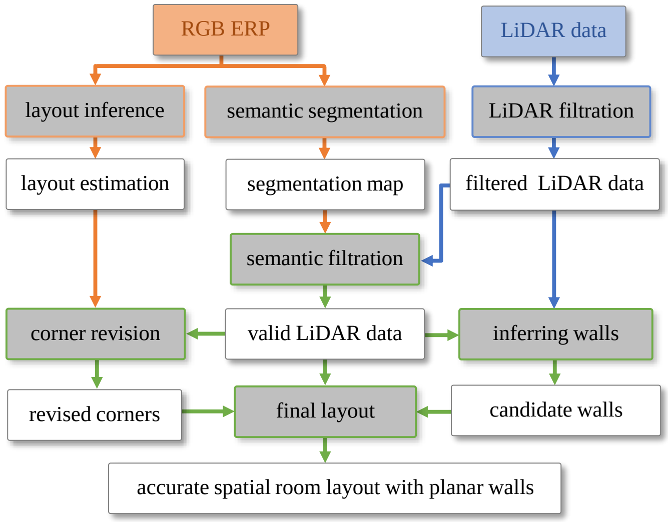


Fig. 1. Schematic representation of the data preprocessing and sensor fusion process. The RGB data, abstract in nature, are processed via an artificial neural network (ANN) to decode spatial information, indicated by the color orange. The LiDAR data, which are blue, undergo a two-stage filtering process to eliminate unreliable or irrelevant measurements. The fusion of the data, marked by green, combines the processed RGB and filtered LiDAR data, taking geometric considerations into account, producing the final layout.

III. Method Design

Following our previous discussion, the next section explores the conceptual underpinnings of our sensor fusion strategy. We outline our approach, highlighting the crucial role of preprocessing sensor data to maximize the potential of each sensor type. This deep dive into the preparatory phase underscores the need to tailor the process to the distinct characteristics of RGB and LiDAR data, ensuring a precise, resilient, and true-to-scale RLR. Rooted in maximizing each sensor's strengths, our method begins with the strategic preprocessing of sensor data, as illustrated in Fig. 1. This strategy allows for a more targeted basis for our sensor fusion than a direct fusion approach. Recognizing that the spatial information encoded in RGB data is largely abstract and requires decoding, we apply an artificial neural network (ANN) to extract quantitative information representing the room's physical features.

By leveraging the parameterized ANN from [36], the corner directions and a semantic segmentation of the room are acquired. A critical component of our method involves a two-stage filtering of the LiDAR data. Starting with a geometric assessment and a subsequent semantic evaluation based on the RGB's semantic segmentation, unreliable or non-relevant measurements are eliminated. Building on the remaining measurements, we deduce candidate walls and assess their validity with a novel metric. Our key contribution lies in the subsequent fusion of candidate walls, validated LiDAR measurments and revised azimuth position of the room corners. By combining data from the panoramic camera with that of the LiDAR scanner and carefully analyzing spatial geometry, our method aims to provide a precise estimation of the room layout of any polygonal shape.

A. RGB Data Abstraction

Recognizing the proficiency of current end-to-end methods in detecting ambiguous clues directly within the image plane, our approach incorporates the ANN from [36] only for room corner position, dropping its layout estimation. Still, HoHoNet establishes a strong baseline to compare against. The method's selection was influenced by its demonstrated efficiency in layout estimation and its provision of semantic segmentation, which is integral for our subsequent verification and filtration of LiDAR data. The implementation of [36] determines the layout corner positions, the wall-wall boundary probabilities for each image column, and wall-ceiling and wall-floor boundaries within the RGB image. Our method uses only the horizontal positions of the wall-wall boundaries, as these directly indicate the azimuth directions of the room corners. Specifically, we chose to sidestep the depth information due to its inherent ambiguity.

However, due to the horizontal offset of the corner estimation via Principal Component Analysis (PCA), we advocate for a corner revision based on the probability of wall-wall boundaries and the wall-ceiling and wall-floor boundaries. This recalibration enhances the accuracy of corner direction identification. Sometimes this may cause an overestimation in corner counts, which will be rectified in later stages of our process, while missing corners cannot be retrieved.

B. LiDAR Data Filtering

In fully observable, unobstructed environments, 2D LiDAR data suffice for RLR as measurements typically align with room boundaries. However, this approach fails when confronted with open scenes or reflective surfaces, resulting in inaccurate spatial enlargements. To mitigate these misperceptions, our method involves a two-stage filtering process to remove misleading data points prior to layout inference as detailed in Fig. 1.

Utilizing geometric analysis, we initially filter out data points that lie outside the target room, often a consequence of openings like doors and windows which manifest as sudden depth increments (see Fig.2).
Recognizing the difference between these sudden changes and the gradual depth increments typical of larger rooms is essential for preserving pertinent measurements.

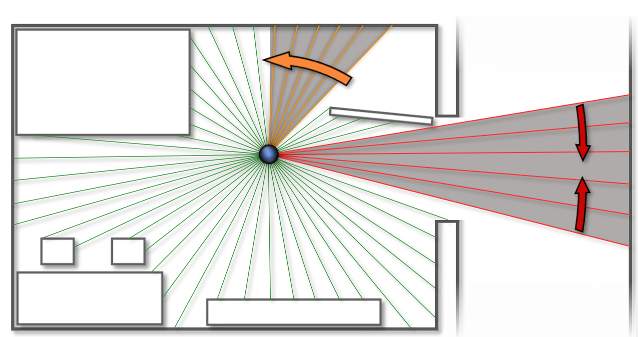


Fig. 2. First stage of filtering LiDAR measurements to exclude data from outside the evaluated room. For the red measurements a sudden increase and decrease in depth is detected, while only one of which is found for the orange measurements. Consequently, only the red measurements are filtered.

Our algorithm detects abrupt depth changes, viewing them as hints of areas extending beyond the intended evaluation space. As detailed in Fig. 2, the algorithm pinpoints outward LiDAR distance spikes and removes intermediate data points if another jump is located within a predetermined angular threshold. The chosen thresholds are adhering to architectural standards [44] with angular and depth limits set at 40 degrees and 1 m, respectively.

Subsequently, the retained LiDAR data undergoes a refinement process via semantic segmentation. This step aims to preserve only wall-related measurements, discarding those associated with unwanted clutter. The RGB-based semantic segmentation (see A.2) is used to produce a binary mask, differentiating between wall and non-wall regions. Maintaining the angular resolution, any clutter-induced measurements are excluded. Notably, this refinement, even in the face of training data variations, consistently delivers commendable outcomes.

While the semantic evaluation proofs efficient, it alone may not adequately differentiate between the desired walls and those from adjacent rooms, thus emphasizing the importance of the initial filtration stage. The resulting validated depth measurements, with clutter and extraneous walls filtered out, are crucial for robust room size determination.

C. Candidate Wall Estimation

By mapping the validated LiDAR data in a top-down projection, prominent lines coinciding with the LiDAR data are identified, indicating potential walls. This involves the Hough transform technique, renowned for its robustness to measurement dropouts, rendering it optimal for detecting lines in the filtered data. This transform is applied to the binary map of validated depth data, shown in 3, resulting in lines with an attributed confidence value. However, as these lines are infinite, segmentation is required. Multiple segmentation options are evaluated; however, issues such as misalignment due to noise or furniture lead to fragmented wall segments. Thus, the preferred approach postpones segmentation, leaving straight lines intact for a later stage, despite the risk of redundant wall detection.

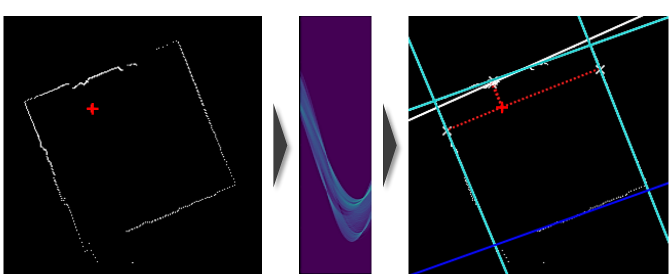


Fig. 3. Candidate wall estimation process: The Hough transform is applied to the filtered depth data for wall estimation. It presents a binary map of the LiDAR data with a red marker signifying the scanner's location (left). The Hough Transform identifies potential walls encoded as lines in the Hough space (middle), which can be transfered into the original projection, outlining candidate walls marked in cyan (right). Subsequently, redundant reconstructions are filtered out (white line), and existing walls are used to perform an improved determination of further walls (blue line).

To resolve this, a novel metric is introduced, identifying the most suitable lines by considering a probability score based on Hough transform confidence and weighted line consensus based on the Manhattan assumption. Nevertheless, due to a conservative detection process some potential wall candidates may remain undetected. A second systematic line detection is proposed, factoring in the assumption of at least four walls per room and focusing on data points not used in the initial wall reconstruction. A second Hough transform is then applied, significantly enhancing the identification of candidate walls. However, in cases of extensive prior data filtration or high amounts of clutter, supplementing RGB data becomes crucial for identifying missing walls.

D. Layout Reconstruction

The final stage of our approach involves reconstructing the floor plan from candidate walls, validated depth data, and revised corner estimation. The image-based corner estimation is instrumental in estimating the directions of room corners in the LiDAR data. The method involves projecting candidate walls and corner directions in a top-down view, connecting corners and indicating corner position relative to the camera. Before finalizing the layout, data alignment with the HoHoNet reconstruction is necessary for comparative analysis, using candidate walls as spatial reference and orientation adjustment. Intersections of candidate walls within the LiDAR's detectable range are infered directly. However, excess corners may occur due to multiple walls exhibiting similar orientations, and must be removed, retaining only the innermost corners. Subsequently, a verification procedure is proposed, in which the found corners are matched with those identified from the RGB data.

However, this approach does not guarantee corner estimation for all walls, especially in cluttered or open spaces where walls may not have been detected. To circumvent this issue, the remaining corners are identified using the corner directions and the filtered LiDAR data directly. If a candidate wall already suggests a wall's presence, it can guide corner finding as shown in Fig. 4. This corner can be located by discovering the intersection point between the candidate wall and the corner direction from the RGB estimation or the validated LiDAR point that corresponds with the corner direction. The point selection is guided by choosing a point further from the camera, as it tends to accurately represent the actual layout.

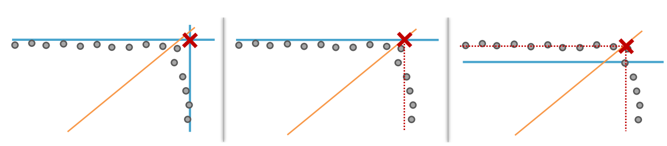


Fig. 4. Illustration of corner selection process. Conclusive corners (red) are determined by the intersection of candidate walls (cyan). Otherwise, the corner is deduced by finding the candidate wall's intersection with the corner direction (orange). However, if the candidate wall is closer to the camera than the LiDAR data (grey circles) suggest, or not present at all, the corner is determined by the LiDAR data point closest to the corner direction.

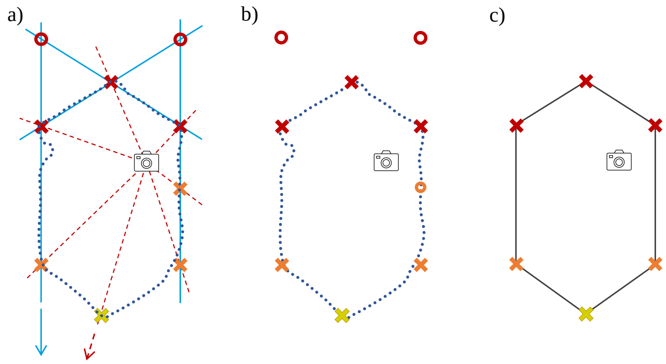


Fig. 5. Exemplary layout inference from candidate walls (blue), filtered LiDAR data (blue points) and corner directions (dashed in red). In the first step, the intersections of the candidate walls are marked (red). Then, the intersections between exactly one candidate wall and the predicted corner direction are found (orange). In the third step, ambiguous corner locations are found by directly matching the respective corner direction to the closest valid LiDAR point (yellow). If one of the orange or yellow corners is redundant and does not contribute to the layout (orange or yellow circles), they will be filtered out. Similarly, intersections that are either too far from the camera (arrows) or do not match the corner directions (red circles) will be filtered out. All the detected candidate corners, together with the original LiDAR points, are depicted in b). In c) the remaining corners are used to reconstruct the layout of the room.

This consideration is essential given that proximate LiDAR data can be skewed by furniture. However, if a LiDAR point exists behind the wall under evaluation, it is favored for a more precise approximation. Care must be taken in considering only feasible LiDAR points in the detected corner direction to avoid potential interference from points in the negative direction.

While this way of finding ambiguous corners might not always produce the optimal result, as LiDAR data points directly adjacent to the corner direction line might have already been filtered out, practical tests corroborate its efficiency in most cases. Redundant corners, which typically originating from the revised corner estimation, are removed for layout reconstruction. The procedure is detailed in Fig. 5 and accommodates diverse room shapes, not limited to Manhattan World assumptions. However, given that most real-world rooms align with these assumptions, retrospective probability adjustments can be applied to remaining corners.

E. Data Processing and Format Conversion

The method converts Matterport3D depthmaps or LiDAR readings into a consistent format for large-scale room evaluations, avoiding manual labeling. Depth maps from Matterport3D are recorded at half the room's height to minimize distortion and reformatted into a single vector, matching the 2D-LiDAR data format. For realistic representation, the data are noise-corrupted before usage.

IV. Evaluation

The proposed method's performance is evaluated through a systematic scheme involving both qualitative and quantitative analyses, comparing it against the state-of-the-art method, HoHoNet. This comparative analysis gives insights into the advantages and limitations of the sensor fusion procedure. The evaluation uses the Matterport3D dataset, which is ideal due to its multimodality and compatibility with HoHoNet's training. Consistent with the training of [36], we employed the Matterport3D dataset [43] due to its rich RGB, depth, and semantic data, as well as its diverse collection of furnished spaces. To ensure a fair evaluation, ground truth annotations (GTAs) from Zou et al. [42] are used. However, given that these GTAs have been scaled to match the camera height assumption of 1.6 m, the resized "original GTAs" are used for the proposed method. Additional "optimized GTAs" are manually collected from 3D point clouds and panorama data for the proposed method's evaluation. The evaluation is conducted on 24 representative rooms from the Matterport3D dataset and then on actual premises, providing a real application scenario.

A. Qualitative Evaluation

The qualitative evaluation centers around subjective criteria to compare the performance of the presented method with the baseline and optimized GTAs. The analysis of six representative results reveals a consistent superiority of the proposed method in identifying and approximating room shapes, compared to the baseline. Exemplary, Fig. 6.1 and 6.2 demonstrate accurate room shape identification, even with complex geometries, by the proposed method while the baseline struggles with room shape reconstruction. While the quantitative metrics are elaborated later, they are displayed for reference.

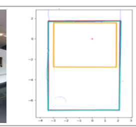
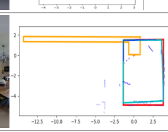
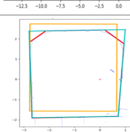
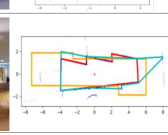
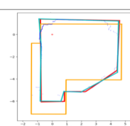
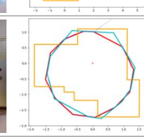
	RGB	Layout Reconstruction	2D IoU	Corner RMSE	Layout RMSE	MRE
1			**98.55** 44.52	**0.03** 2.12	**0.04** 2.61	**0.004** 0.238
2			**92.38** 0.05	**0.19** ---	**0.12** 2.96	**0.027** ---
3			**96.75** 80.86	--- ---	**0.14** 0.44	--- ---
4			**74.04** 51.11	--- ---	**0.56** 1.16	--- ---
5			**96.42** 71.28	**0.13** ---	**0.19** 0.82	**0.018** ---
6			**92.47** 62.99	--- ---	**0.11** 0.38	--- ---

Fig. 6. Qualitative evaluation of the proposed method (cyan), HoHoNet (orange) and optimized ground truth (red). For reference, the quantitative evaluation metrics are also shown, if the corner count was estimated correctly. The proposed method outperforms the baseline on both Manhattan-conform and arbitrarily shaped layouts.

TABLE II

Comparative numeric results on the Matterport3D dataset, indicating that the proposed method significantly outperforms HoHoNet across all metrics. Arrows denote optimal direction for each metric; means and standard deviations are reported.

	2D IoU [%] ↑	Corner RMSE [m] ↓	Layout RMSE [m] ↓	MRE [%] ↓
Proposed Method (optimized GTA)	**90.47** $\pm$ 7.46	**0.160** $\pm$ 0.121	**0.358** $\pm$ 0.240	**0.057** $\pm$ 0.067
Proposed Method (original GTA)	80.38 $\pm$ 17.41	0.361 $\pm$ 0.411	0.439 $\pm$ 0.286	0.133 $\pm$ 0.142
HoHoNet (optimized GTA)	60.35 $\pm$ 17.83	0.682 $\pm$ 0.581	1.110 $\pm$ 0.897	0.169 $\pm$ 0.142
HoHoNet (original GTA)	67.75 $\pm$ 22.66	0.404 $\pm$ 0.192	0.893 $\pm$ 0.756	0.104 $\pm$ 0.088
HoHoNet (as reported by [36])	82.32	—	0.394	0.104

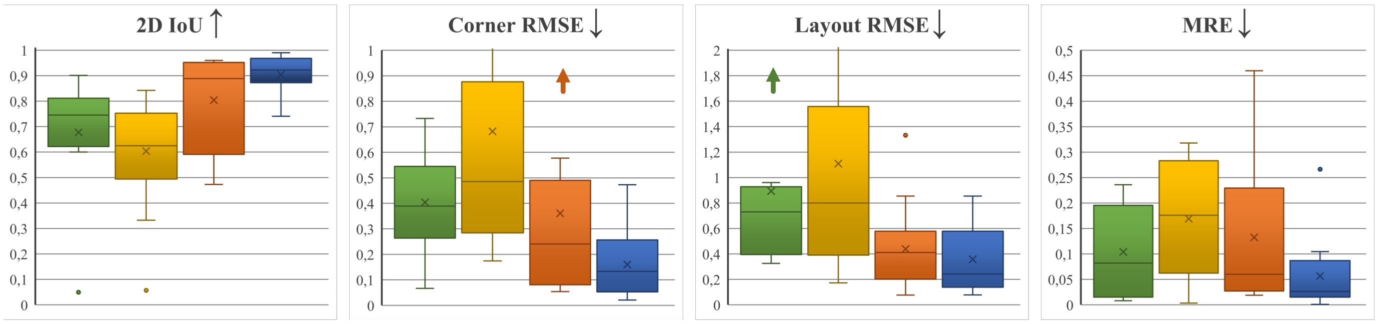


Fig. 7. Box plot evaluation of the four evaluation metrics. The arrow behind each metric indicates the desired optimum. Small arrows above the plots indicate outliers. Each distribution is depicting the spread of the results of the proposed method as well as HoHoNet for each the original and the optimised ground truth. The 2D IoU and the Layout RMSE were calculated for all of the 24 layouts, while the Corner RMSE and the MRE were calculated for 14 layouts by the proposed method and 10 by HoHoNet, as these represent the layouts for which the number of corners was predicted correctly. Green and yellow plots are HoHoNet with original ground truth and optimized ground truth, respectively, orange and blue represent the proposed method with original and optimized ground truth, respectively.

In non-Manhattan room layouts, as shown in Figures 6.3 to 6.6, the proposed method generally predicts the room shape better. Specifically, the proposed method correctly identifies non-Manhattan aligned room shapes (Fig. 6.3), captures dimensions more accurately despite overestimating certain areas (Fig. 6.4), accurately approximates room corners (Fig. 6.5), and recognizes complex room shapes such as convex polygons (Fig. 6.6) - refer to A.1 for 3D results. Meanwhile, the baseline method shows substantial limitations, often falling back to Manhattan shape predictions and struggling with accurate room dimension estimation. The research indicates that the proposed method surpasses the baseline method in accurately detecting wall positions, consistently identifying corners or walls, and reliably determining the underlying room shape, including angled walls and room contours. In contrast, the baseline method struggles with accurate room dimension identification and often misestimates wall placement. The key weakness of HoHoNet becomes apparent when evaluating non-Manhattan-shaped rooms, where significant model deviations are likely to occur.

B. Quantitative Evaluation

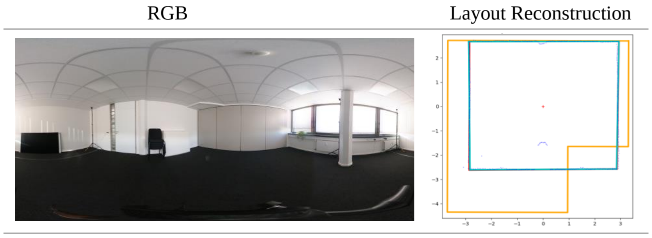
Additionally, an objective evaluation was carried out to compare the performance of the proposed method and the baseline method using accepted metrics. These metrics include the 2D Intersection over Union (2D IoU), Corner Root-Mean-Squared Error (Corner RMSE), Layout Root-Mean-Squared Error (Layout RMSE), and Mean-Relative-Error (MRE).

The 2D IoU measures overlap between two 2D shapes, hence gauging the consensus between method results and GTAs. A 2D IoU value closer to 1 signifies improved congruence. Corner RMSE and Layout RMSE measure the difference between predicted corners and layout, respectively, and the actual ground truth and are measured in meters. Lower values in these metrics indicate better performance. MRE provides a means to account for distance variations, weighing corner estimation errors by their distance from the camera, with lower values indicating better corner estimations.

As can be seen from Table II, Fig. 7 and in detail in Fig. 6, the proposed method on average outperforms the baseline method across all four metrics, indicating superior accuracy, especially for non-Manhattan layouts and rooms with substantial clutter or intricate wall patterns. Predictably, each method achieves the best performance corresponding to the ground truth it was optimized for. Detailed results for individual rooms are included in A.3.

C. Results on with Sensor Data

To validate the previous results, we evaluated our method using data from our own office, captured with a consumer-grade sensor setup: a RICOH THETA Z1 Panorama Camera paired with a DTOF LiDAR LD06. In these tests, our method consistently and accurately identified walls, corners, and overall room layout, aligning closely with the established ground truth, depicted in Fig. 8. Despite the presence of complex elements like windows and a free-standing pillar, it correctly interpreted these as clutter, offering precision in spatial dimensions estimation. Quantitative metrics confirm the high fidelity of these estimations within a minimal tolerance range.



	2D IoU	Corner RMSE	Layout RMSE	MRE
Ours	**99.37**	**0.02**	**0.01**	**0.002**
HoHoNet	63.73	---	2.61	---

Fig. 8. Layout results on DLR office dataset, exemplifying the real application scenario. The proposed method (cyan) closely matches the manually determined ground truth (red), while the baseline (yellow) fails to identify the underlying room shape.

In contrast, HoHoNet significantly struggled with this task by incorrectly inferring an L-shaped room layout, which is reflected by the relatively low 2D IoU and Layout RMSE. The primary failure point appears to be a misinterpretation of the free-standing pillar as an extra corner, which not only distorted the layout estimation but also led to considerable overestimation of room dimensions. The contrast in performance underscores our proposed method's superiority in terms of accuracy and adaptability to other datasets. These results, although drawn from a single layout, indicate the method's efficacy for its intended use case.

In summary, the proposed method substantially surpasses the baseline on the Matterport3D dataset, achieving an average 2D IoU of over 90% and showing marked improvement in GTA-alignment. The method excels in estimating non-Manhattan spaces, resulting in the elimination of the Manhattan assumption. The reduced variability in results, as evidenced by tighter box plot ranges and decreased standard deviation across all metrics, highlights the method's consistency in diverse spaces.

V. Discussion

We've found our new method to be effective. Combining panoramic images with 2D-LiDAR data to assess spaces in multiple ways, it provides a comprehensive perspective. Merging RGB and LiDAR data results in better estimations of wall positions and room shapes compared to relying solely on RGB, especially in spaces filled with clutter. The RGB-based corner estimation compensates for data limitations and allows to determine corners even when few or no candidate walls were found. A key improvement with our method is how it can filter out corners that line up directly with two others, boosting our corner accuracy by 50% over the baseline. Additionally, the ability to factor in non-Manhattan room shapes emphasizes its practicality in RLR. The two-step LiDAR data filtration contributes significantly to both reconstruction accuracy and reliability. Even though the quality of the reconstruction might vary with data density and accuracy, the method remains robust in challenging situations and maintains a 2D IoU consistently over 70%, even on round layouts.

A. Limitations of the Proposed Method

Although our method works well for most room shapes, it approximates curved-wall rooms as polygons. Another issue is that our method might perform less reliable because it was trained on the Stanford 2D-3D-S dataset, not the Matterport3D dataset we used. This discrepancy may lead to subpar filtration of LiDAR data. Even slight inaccuracies in segmentation can compromise the filtering process by rejecting valid measurements or accepting invalid ones.

Additionally, there's a considerable dependency on the precise filtration of LiDAR measurements. Performance could suffer if valid LiDAR points get misclassified as invalid, particularly in cluttered environments or close object proximities. While such instances are rare and their impact can often be minimized with thorough data collection, these remain as considerations for potential users. It should be noted that in the rooms evaluated, such effects were not observed.

B. Comparison to State-of-the-Art

The existing state-of-the-art can be categorized into LiDAR-based methods, renowned for their precise room dimension capture, and RGB-based methods, adept at reconstructing cluttered rooms due to their ANN utilization for data interpretation. The presented method effectively bridges this divide by combining both techniques, pioneering a sensor fusion approach to map intricate rooms with enhanced spatial accuracy and reliability. By capitalizing on each sensors inherent advantages, it eliminates several assumptions needed for RGB-based reconstruction.

Compared to the best-performing method from the state-of-the-art, HoHoNet, the proposed method shows significant improvements throughout all critical metrics, validating the superiority of sensor fusion methods in layout reconstruction.

VI. Conclusion and Prospects

This study presents a new approach to Room Layout Reconstruction (RLR) that uses both LiDAR and RGB data to retrospectively map out unknown furnished rooms. Our method demonstrated a strong improvement in layout estimation, achieving a 2D IoU of over 90%, compared prevailing methodologies. However, the context of specific metrics and conditions under which this improvement was noted is crucial for readers to understand its significance. Key to our approach's success is the ability to overcome sensor-specific challenges: utilizing a two-stage point cloud filtering process for precise LiDAR measurements and leveraging RGB data to deduce azimuth positions of elusive corners, thereby reducing ambiguities inherent in RGB depth perception. While our method excels in capturing conventional Manhattan layouts, it's ability to accurately represent arbitrary room shapes without the presumption of a known camera height underscores its broader applicability in various architectural and design scenarios.

Looking ahead, our priority is to enhance the semantic segmentation by supplementing the same training data as for the corner estimation. Such enhancements not only promise better LiDAR data filtration but also more reliable wall identification. The technique is set to be adapted to address more intricate room layouts, expanding its range of applicability - from circular to semi-open spaces, and those that challenge the camera's line of sight. By addressing these challenges, we aim to further establish our method as a versatile tool for various architectural scenarios.

VII. Conflict of Interests

The authors declare no conflict of interest, that could have influenced the results of this work.

References

[1] O. Brian, et al. "Toward automated modeling of floor plans." Proceedings of the symposium on 3D data processing, visualization and transmission. Vol. 2. 2010.

[2] Ching, Francis DK. "Architecture form, space & order." (2020).

[3] X. Yin, P. Wonka, and A. Razdan. "Generating 3d building models from architectural drawings: A survey." IEEE computer graphics and applications 29.1 (2008): 20-30.

[4] G. Vosselman, and H.-G. Maas. Airborne and terrestrial laser scanning. CRC Press (Taylor & Francis), 2010.

[5] P. J. Besl, and N. D. McKay. "Method for registration of 3-D shapes." Sensor fusion IV: control paradigms and data structures. Vol. 1611. Spie, 1992.

[6] Glenn, Nancy F., et al. "Analysis of LiDAR-derived topographic information for characterizing and differentiating landslide morphology and activity." Geomorphology 73.1-2 (2006): 131-148.

[7] Lim, Kevin, et al. "LiDAR remote sensing of forest structure." Progress in physical geography 27.1 (2003): 88-106.

[8] Roh, Hyunchul, et al. "Accurate mobile urban mapping via digital map-based SLAM." Sensors 16.8 (2016): 1315.

[9] Jin, Yusheng. "AutoRooms: Automatic Room Segmentation Based on Wall Constraints from Point Clouds." 2023 4th International Conference on Computer Vision, Image and Deep Learning (CVIDL). IEEE, 2023.

[10] R. Hoffman and A. K. Jain, "Segmentation and Classification of Range Images," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-9, no. 5, pp. 608-620, Sept. 1987.

[11] C. Wang et al. "Semantic line framework-based indoor building modeling using backpacked laser scanning point cloud". In: ISPRS Journal of Photogrammetry and Remote Sensing 143 (2018), pp. 150–166.

[12] H. Yoshisada et al. "Indoor Map Generation from Multiple LIDAR Point Clouds". In: 2018 IEEE International Conference on Smart Computing (SMARTCOMP).IEEE, 2018, pp. 73–80.

[13] S. Karam, V. Lehtola, and G. Vosselman. "Strategies to integrate imu and lidar slam for indoor mapping". In: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences V-1-2020 (2020), pp. 223–230.

[14] C. Yan et al. "3D Room Layout Estimation From a Single RGB Image". In: IEEE Transactions on Multimedia 22.11 (2020), pp. 3014–3024.

[15] L. Del Pero et al. "Bayesian geometric modeling of indoor scenes". In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2012, pp. 2719–2726.

[16] C. Lee et al. RoomNet: End-to-End Room Layout Estimation. 2017.

[17] W. Zhang and Y. Liu. "Room Layout Estimation by Learning Depth Maps of Planes from 2D Layout Labels". In: 2021 27th International Conference on Mechatronics and Machine Vision in Practice (M2VIP). IEEE, 2021, pp. 777–782.

[18] C. Zou et al. "LayoutNet: Reconstructing the 3D Room Layout From a Single RGB Image". In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018.

[19] S. Ramalingam et al. "Manhattan Junction Catalogue for Spatial Reasoning of Indoor Scenes". In: 2013 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2013, pp. 3065–3072.

[20] D. Hoiem, A. A. Efros, and M. Hebert. "Geometric context from a single image". In: Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1. IEEE, 2005, 654–661 Vol. 1.

[21] D. C. Lee, M. Hebert, and T. Kanade. "Geometric reasoning for single image structure recovery". In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2009, pp. 2136–2143.

[22] Y. Zhang et al. "PanoContext: A Whole-Room 3D Context Model for Panoramic Scene Understanding". In: Computer vision - ECCV 2014. Ed. by David J. Fleet. Vol. 8694. LNCS sublibrary: SL 6 - Image processing, computer vision, pattern recognition, and graphics. Cham: Springer, 2014, pp. 668–686.

[23] C. Zhang et al. DeepPanoContext: Panoramic 3D Scene Understanding with Holistic Scene Context Graph and Relation-based Optimization. 2021.

[24] K. Zhou, K. Yang, and K. Wang. "Panoramic depth estimation via supervised and unsupervised learning in indoor scenes". In: Appl. Opt. 60.26 (2021), pp. 8188–8197.

[25] P. Vu Tran. SSLayout360: Semi-Supervised Indoor Layout Estimation from 360-Degree Panorama. 2021.

[26] X. Wang, D. F. Fouhey, and A. Gupta. Designing Deep Networks for Surface Normal Estimation. 2014.

[27] H. Zhao et al. "Physics Inspired Optimization on Semantic Transfer Features: An Alternative Method for Room Layout Estimation". In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017, pp. 870–878.

[28] Z. Kang et al. "A Review of Techniques for 3D Reconstruction of Indoor Environments". In: ISPRS International Journal of Geo Information 9.5 (2020), p. 330.

[29] H. Yang and Hui Zhang, eds. Efficient 3D Room Shape Recovery from a Single Panorama. Piscataway, NJ: IEEE, 2016.

[30] Y. Yang et al. "Automatic 3D Indoor Scene Modeling from Single Panorama". In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2018, pp. 3926–3934.

[31] Y. Zhang et al. "PanoContext: A Whole-Room 3D Context Model for Panoramic Scene Understanding". In: Computer vision - ECCV 2014. Ed. by David J. Fleet. Vol. 8694. LNCS sublibrary: SL 6 - Image processing, computer vision, pattern recognition, and graphics. Cham: Springer, 2014, pp. 668–686.

[32] J. Xu et al. "Pano2CAD: Room Layout from a Single Panorama Image". In: 2017 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2017, pp. 354–362.

[33] Shang-Ta Yang et al. "DuLa-Net: A Dual-Projection Network for Estimating Room Layouts From a Single RGB Panorama". In: IEEE/CVF (CVPR). IEEE, 2019, pp. 3358–3367.

[34] C. Sun et al. HorizonNet: Learning Room Layout with 1D Representation and Pano Stretch Data Augmentation. 2019.

[35] G. Pintore, M. Agus, and E. Gobbetti. "AtlantaNet: Inferring the 3D Indoor Layout from a Single 360° Image Beyond the Manhattan World Assumption". In: Computer Vision – ECCV 2020. Ed. by Andrea Vedaldi et al. Vol. 12353. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2020, pp. 432–448.

[36] C. Sun, M. Sun, and H.-T. Chen. "HoHoNet: 360 Indoor Holistic Understanding with Latent Horizontal Features". 2020.

[37] H. Yu, et al. "PanelNet: Understanding 360 Indoor Environment via Panel Representation." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.

[38] C. Weiss and A. Zell. "Automatic Generation of Indoor VR-Models by a Mobile Robot with a Laser Range Finder and a Color Camera". In: Autonome Mobile Systeme 2005. Springer-Verlag, 2006, pp. 107–113.

[39] Department of Electrical and Computer Engineering Oakland University. "2D LiDAR and camera fusion in 3D modeling of indoor environment". IEEE, 2015.

[40] J. Li and R. L. Stevenson. "Indoor Layout Estimation by 2D LiDAR and Camera Fusion". 2020.

[41] J. Li and R. L. Stevenson. "2D LiDAR and Camera Fusion Using Motion Cues for Indoor Layout Estimation". 2021.

[42] C. Zou et al. "Manhattan Room Layout Reconstruction from a Single 360 image: A Comparative Study of State-of-the-art Methods". 2019.

[43] A. Chang et al. "Matterport3D: Learning from RGB-D Data in Indoor Environments". In: 2017 International Conference on 3D Vision (3DV). IEEE, 2017, pp. 667–676.

[44] German Building Standards Committee. Planning principles of barrier-free building: DIN 18040-1:2010-10.

APPENDIX

Fig. A.1. Comparative Analysis of the 3D Layout Reconstructions: Evaluating the Precision of HoHoNet and the Proposed Method in Detecting Hallway Ends, Identifying Wall Projections, and Correctly Assigning Complex Structures in Diverse Layouts. Both methods identify the elongated hallway in Fig. 7.4, yet struggle to accurately map ridged wall shapes, resulting in wall parts being projected onto the floor. The proposed method outperforms HoHoNet by correctly identifying hallway ends and preventing overestimations. In the second layout, taken from Fig. 7.5, it accurately assigns the furnace to an angled wall, unlike HoHoNet's Manhattan wall projection, and avoids unnatural room warping. For the rounded layout, which can also be seen in Fig. 7.6, the proposed method surpasses HoHoNet's rigid Manhattan layout, reconstructing a polygon closely mirroring the round shape of the actual layouts. Overall, it demonstrates significant improvements over the baseline.

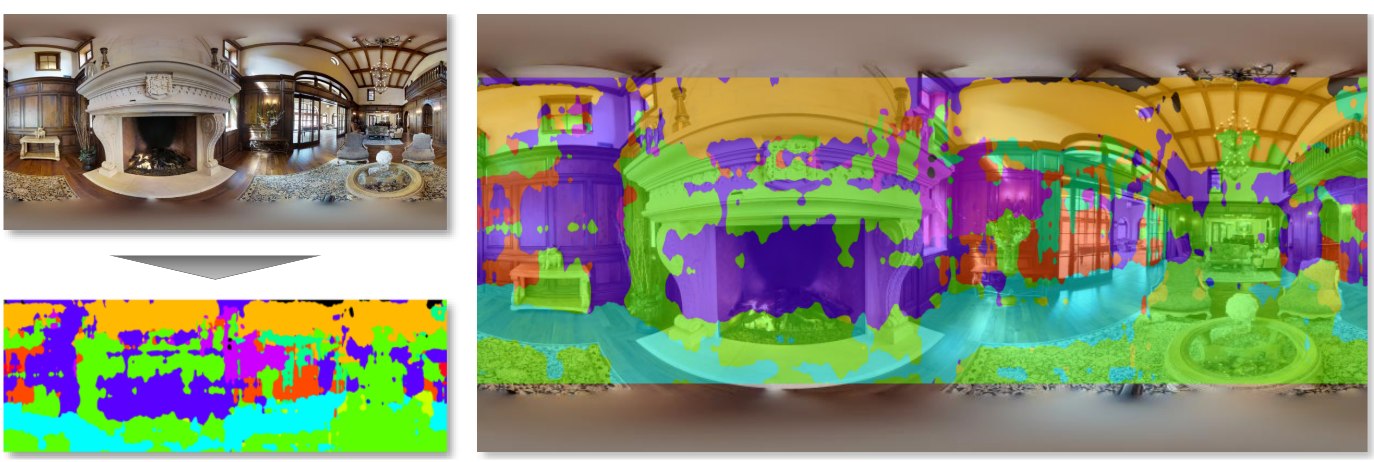


Fig. A.2. Illustration of the semantic segmentation from the original RGB data as proposed by HoHoNet on the Matterport3D dataset. The left side depicts the original RGB and the shows the predicted semantic segmentation map. Lastly, both images are depicted superimposed. Effectively, the purple areas (walls) and light red areas (boards) are used for the estimation of walls as these represent places where a reliable LiDAR measurement is detected.

Matterport3D - Layout ID:	HoHoNet - LayoutAnnotator GT				HoHoNet - Hand labled GT				Proposed Method- LayoutAnnotator GT				Proposed Method - Hand labled GT			
	2D IoU	RMSE	RMSE 2D	MRE	2D IoU	RMSE	RMSE 2D	MRE	2D IoU	RMSE	RMSE 2D	MRE	2D IoU	RMSE	RMSE 2D	MRE
52280dcf1af54cd9806e8574f13468b3	0.90099	0.06662	0.32602	0.01016	0.54608	0.46511	0.17294	0.27274	0.47309	0.57754	0.54216	0.45986	0.87871	0.20912	0.19609	0.04928
3542707c604c4c27835928ac9878c631					0.80858		0.43832				0.13653		0.96758		0.13653	
02ee4a5177f844c0867d3e174a1080e2					0.62993		0.38269				0.10784		0.92472		0.10784	
9ea0131d87a2428fb2f6cdb6167fa1bc					0.71727		0.57693				0.82728		0.77721		0.82728	
6e95a11ced8b44b39fc3d795b7b32721					0.3327		2.72805				0.66191		0.91807		0.66191	
3ecb492d5019411785bab293ff4f74ee	0.77983	0.31612	0.3489	0.03041	0.78099	0.30461	0.37524	0.00362	0.96011	0.05388	0.0965	0.03018	0.98987	0.02067	0.10537	0.00089
0f42320ef887416fb8ed9b91b6641cfc	0.71351	0.54464	0.60823	0.18502	0.70044	0.58546	0.35886	0.19874	0.95044	0.09914	0.45543	0.02411	0.97861	0.05282	0.21584	0.01217
4dbaf3f1a80b401dad9fc7a8bcb297d4	0.62936	0.73298	0.96051	0.1327	0.62049		1.22546		0.5735	1.42686	1.33262	0.2857	0.893		0.62023	
0b302846f0994ec9851862b1d317d7f2					0.51114		1.16155				0.55668		0.74038		0.55668	
0fc9937d10d249b8b34264cd53cbba5a	0.69177	0.39095	0.8297	0.19884	0.705	0.34748	0.75426	0.15311	0.90876	0.1247	0.07642	0.05711	0.97138	0.07538	0.07834	0.00133
2f3e7d98347047829630f0114694a532	0.82661	0.24652	0.91747	0.00807	0.84232	0.22366	0.87542	0.00841	0.95519	0.06206	0.23722	0.01902	0.96546	0.07015	0.20756	0.01734
a6c8016b0fa94ebaa51c2808c6fec8ba					0.49185		2.96135				0.58472		0.92021		0.58472	
1eae5d4a0c0146c2b884d6b96d9b0e16					0.51494		1.66924				0.25367		0.89462		0.25367	
c7b097de55f242fa8bdd795d3945637d					0.47962		1.06425				0.52832		0.78592	0.47314	0.52832	0.08406
4cc8ffa4a220460f974971ef2da47296	0.60038	0.54515	0.63129	0.23602	0.48677	0.7067	0.7902	0.31799	0.83718	0.26092	0.31733	0.08534	0.96766	0.05255	0.10569	0.02538
4cf4c84ed4ae48b2b4076fb0e742d5f1	0.8061	0.38768	0.84024	0.03158	0.76453	0.50544	0.67171	0.08027	0.86821	0.24063	0.40571	0.05997	0.97863	0.05036	0.23057	0.01641
1c598b0601734b5da0ae03b359079dca	0.04943		3.06146		0.05673		3.13688		0.91546		0.19076		0.92389	0.18559	0.12097	0.02753
1e3a672fa1d24d668866455162e5b58a					0.82776		0.33276				0.85479		0.78221		0.85479	
1dbf4725d8054e37b42fafae2e9d2fb2	0.77728		0.41073		0.59233		0.19522		0.59656	0.40251	0.35512	0.17321	0.89947	0.14122	0.14583	0.07365
40eda297975f4648bbc301ff98940fcf					0.79629	0.17434	0.9501	0.10386			0.60956		0.79364	0.25346	0.60956	0.10448
649a58719cd7455ebe18675c9719dfa7					0.4352	2.12247	2.67771	0.23843			0.4168		0.98553		0.4168	
1da6bc4a3acc48868c4ea1904c26c92f					0.7128		0.81048				0.14998		0.96424	0.1247	0.18866	0.01757
00c6dc6b90a947a48c378f08b061234e					0.50033	1.38536	1.90208	0.31356			0.51478		0.94092	0.26408	0.51478	0.09578
6cc97300dcd2469c82449756fef5cffd					0.62925		0.41697				0.32618		0.87027	0.27112	0.32618	0.2664
average performance:	0.67753	0.40383	0.89346	0.1041	0.60347	0.68206	1.10953	0.16907	0.80385	0.36092	0.4391	0.13272	0.90468	0.16031	0.35809	0.05659
standard deviation:	0.22658	0.19193	0.7559	0.08838	0.17827	0.58137	0.89742	0.11183	0.17406	0.41059	0.28628	0.14164	0.07462	0.12081	0.24012	0.06739

Fig. A.3. Comparative analysis of the Proposed Method and HoHoNet on individual rooms using common metrics. For each method and ground truth the metrics are reported with an average value and standard deviation. The original ground truth is derived from LayoutAnnotator [42] annotations, while the optimized ground truth is manually derived from the 3D point cloud.