



Optimizing Architectural Layouts and Spatial Configurations in Polar Environments Using Image Segmentation and Semantic Mapping Techniques

Dezhong Shang^{ID}, Guangtian Zou^{*ID}

School of Architecture and Design, Key Laboratory of Cold Region Urban and Rural Human Settlement Environment Science and Technology, Ministry of Industry and Information Technology, Harbin Institute of Technology, Harbin 150001, China

Corresponding Author Email: zougta@hit.edu.cn

Copyright: ©2025 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ts.420412>

ABSTRACT

Received: 8 December 2024

Revised: 17 June 2025

Accepted: 3 July 2025

Available online: 14 August 2025

Keywords:

polar environment, image segmentation, architectural layout optimization, SLAM, grid map, spatial configuration algorithm

The extreme climatic and environmental conditions of polar regions impose stringent constraints on architectural safety, functionality, and adaptability. With the growing prevalence of scientific expeditions and resource exploration in these remote territories, the demand for resilient, efficient, and adaptive architectural solutions has increased substantially. However, conventional design methodologies have been found inadequate in addressing the multifaceted challenges posed by low temperatures, harsh illumination conditions, and limited spatial flexibility. In particular, standard image segmentation algorithms often underperform in polar indoor environments due to dynamic lighting variations, high reflectivity of ice and snow surfaces, and structural ambiguities. Additionally, existing optimization frameworks for architectural layouts frequently neglect the thermal inefficiencies induced by extreme cold, as well as the distinctive functional zoning requirements of polar buildings, such as isolation zones, decontamination chambers, and modular emergency units. To address these limitations, an integrated architectural optimization approach has been developed, combining image segmentation, semantic mapping, and spatial configuration modelling tailored for polar contexts. First, a planar image matching technique has been proposed, leveraging angular and distance-based features to extract object orientations and spatial relationships, thereby enhancing scene recognition robustness under variable visual conditions. Second, a semantic simultaneous localization and mapping (semantic SLAM) framework has been adapted for indoor architectural segmentation, enabling real-time integration of semantic information into the SLAM pipeline for high-precision spatial modelling and environmental interpretation. Third, a grid map-based optimization model has been constructed to quantify spatial attributes and incorporate environmental variables—such as thermal conductivity, wind flow, and material performance—into layout decision-making. Functional zoning constraints specific to polar operations have also been embedded within the optimization objective functions to ensure mission-specific spatial configurations. The innovations presented lie in the mitigation of polar-specific visual interference in image processing, the enhancement of architectural segmentation through semantic-augmented SLAM, and the development of an environmentally responsive spatial optimization framework. These contributions are expected to provide foundational support for intelligent, data-driven architectural design in extreme environments, while offering methodological advancements to the broader fields of remote architecture, robotics, and environmental informatics.

1. INTRODUCTION

Polar regions are characterized by extreme low temperatures, strong winds, and ice and snow coverage throughout the year [1-3], and also exhibit special natural phenomena such as polar day and polar night. These factors impose very high requirements on the safety, functionality, and adaptability of local buildings. With the increasing frequency of polar scientific investigations and resource exploration activities [4, 5], the demand for polar buildings is continuously increasing. How to achieve optimized architectural layout and efficient spatial configuration under extreme environments [6, 7] has become a key issue to ensure

the smooth progress of polar activities. Traditional architectural design and layout methods are difficult to fully adapt to the particularity of polar environments, urgently requiring advanced image segmentation and algorithm technologies to provide new solutions for the intelligent design of polar buildings.

Relevant research is of great significance for improving the utilization efficiency of polar buildings, reducing operational costs, and ensuring personnel safety. From a practical application perspective, optimized architectural layout and reasonable spatial configuration [8, 9] can improve the resistance of polar buildings to extreme climates, reduce energy consumption, and provide a more comfortable and safe

working and living environment for researchers and expedition members, thereby enhancing the sustainability and efficiency of polar activities. From the disciplinary development perspective, this research can promote the interdisciplinary integration of computer vision, artificial intelligence, and architectural science in polar environments, expand the application boundaries of related technologies, and provide new theoretical and methodological support for architectural design and optimization in extreme environments.

Although existing studies have explored architectural layout optimization and image segmentation fields, there are still obvious deficiencies in addressing the particularity of polar environments. For example, some image segmentation algorithms based on conventional environments [10-14] show significantly reduced segmentation accuracy when processing images of indoor polar buildings affected by drastic lighting changes and snow and ice reflections, making it difficult to accurately identify functional areas and structural details inside buildings; some architectural layout optimization models [15-18] do not fully consider the effects of low temperatures on building material properties and heat transfer efficiency, resulting in poor applicability of optimization results in actual polar environments; meanwhile, existing spatial configuration methods [19, 20] often overlook the layout needs of special functional areas in polar buildings, failing to meet the special functional requirements of polar activities.

This thesis conducts in-depth research on architectural layout optimization and spatial configuration under polar environments, mainly including three parts. First, a planar image matching method for indoor architectural scenes based on angles and distances is proposed, which extracts angular features and spatial distance relationships of objects in images to construct a robust matching model, improving the accuracy and robustness of image matching in complex indoor polar building environments. Second, a semantic SLAM method for indoor architectural scene image segmentation is developed, integrating semantic information into the SLAM process to achieve real-time semantic segmentation and 3D map construction of polar building interiors, providing fine semantic information for spatial analysis. Third, an architectural layout optimization and spatial configuration framework based on grid maps under polar environments is constructed. By combining polar environmental parameters and architectural functional requirements, the building space is quantitatively represented using grid maps, and layout optimization objective functions and constraints are formulated to achieve efficient spatial configuration of polar buildings. The value of this research lies in that the proposed algorithms can specifically solve the special problems of architectural image processing and layout optimization under polar environments, provide technical support for intelligent design of polar buildings, and enrich interdisciplinary research achievements at the intersection of extreme environment architecture and algorithms.

2. PLANAR MATCHING FOR INDOOR POLAR SCENES USING ANGLES AND DISTANCES

Only accurate planar matching can ensure the accuracy of spatial structure cognition, thereby supporting reasonable layout planning. Due to polar day and polar night, indoor polar

environments experience drastic lighting changes, and snow and ice reflections easily cause image grayscale distortion. Traditional matching methods relying on texture or color are prone to failure. Therefore, this paper chooses to perform image planar matching based on angles and distances in indoor scenes of polar buildings to cope with the special interference of polar environments on image data. The angles and center distances of planes, as geometric features, are less affected by lighting, reflections, and other environmental factors, and can maintain stability under extreme conditions. Meanwhile, polar buildings are mostly modular structures, where the angles and spatial distance relationships of planes such as walls and floors have strong regularity. Using these as matching criteria can effectively reduce noise interference and provide reliable planar topological relationships for the 3D maps constructed by subsequent semantic SLAM. This forms the basis for optimizing the layout of polar buildings.

The angle- and distance-based planar matching method for indoor architectural scenes proposed in this paper is implemented with a closed-loop process of "extraction - comparison - association/creation" centered on a global plane database. First, planar features of the current frame are extracted from RGB-D images, focusing on calculating the normal vector angles and center 3D coordinates of each plane. Then, these features are fully compared with all planes in the global plane map, rather than being limited to keyframes only, to avoid matching omissions caused by keyframe selection bias in polar environments. During comparison, a double-threshold judgment is used: when the angle difference between two planes is $\leq 8^\circ$ and the center distance ≤ 0.1 m, they are determined to be the same plane. At this time, new geometric constraints are added to the local bundle adjustment to strengthen the consistency of the global map. If no global plane meeting the thresholds is found, a new entry for the plane is created in the global database. This design filters noise through strict geometric constraints, while improving matching coverage by full comparison, ensuring the completeness and accuracy of planar associations in complex indoor polar environments.

3. SEMANTIC SLAM METHOD FOR INDOOR ARCHITECTURAL SCENE IMAGE SEGMENTATION

In polar indoor environments, drastic lighting changes and snow and ice reflection interference easily cause cumulative errors in traditional SLAM without loop closure detection and global optimization, making it difficult to accurately locate camera poses and thus affecting the spatial consistency of image segmentation. Semantic information, as a stable high-level feature in the environment, can provide additional short-range constraints for SLAM. For example, the regularity in spatial distribution of semantic categories such as walls, doors, and windows can assist in correcting pose deviations and improve the robustness of SLAM under extreme conditions. Meanwhile, the functional zoning of polar buildings, such as low-temperature laboratories and material storage areas, requires very high accuracy of semantic labels. Only through semantic SLAM, which deeply integrates image segmentation with spatial localization, can dual accurate cognition of "geometric structure + semantic categories" of indoor scenes be realized, providing foundational data with both spatial coordinates and functional attributes for subsequent grid map-based layout optimization. This need cannot be met by pure

image segmentation or traditional SLAM alone.

The semantic SLAM method for indoor architectural scene image segmentation proposed in this paper is implemented around two core objectives: "semantic constraints to enhance SLAM accuracy" and "global consistent semantic map construction." First, semantic information such as categories of walls, floors, and equipment is extracted from the current frame using image semantic segmentation technology and integrated as short-range constraints into the SLAM optimization process. By constructing a semantic cost function based on an observation probability model, the matching degree between semantic labels and spatial positions is quantified and incorporated into the nonlinear optimization framework of SLAM. This reduces pose estimation errors through semantic consistency constraints when loop closure detection is lacking. Second, for polar indoor scenes, a binocular camera collects color and depth images, and combining the precise camera poses output by semantic SLAM, the semantic segmentation results are fused with 3D point clouds to generate an initial dense 3D grid map. Since polar environments may cause local semantic label noise, such as category misjudgment triggered by snow and ice reflections, a conditional random field model is introduced to perform global optimization on grid labels by modeling semantic correlations between adjacent grids, correcting isolated noise points and ultimately obtaining a globally consistent semantic grid map.

3.1 Image semantic segmentation based on the improved DeepLabv3

Indoor polar buildings contain multi-scale objects, such as large walls and small experimental equipment, doors, and windows. Influenced by drastic lighting changes and snow and ice reflections, object edges tend to be blurred. Therefore, this paper chooses to introduce an improved version of DeepLabv3, which retains the multi-scale information capture ability of atrous convolution and combines the detail recovery ability of an efficient decoder in the encoder-decoder structure. This approach can utilize the pyramid pooling of the encoder to extract features at different scales and sharpen segmentation edges through the decoder, providing more accurate semantic labels for semantic SLAM. This is the premise for effectively integrating semantic constraints into SLAM optimization. At the same time, this paper chooses to introduce an improved Xception model, which reduces computation while improving accuracy. Its lightweight design adapts to the potentially limited computing power of hardware devices under low temperatures in polar scenarios, ensuring that semantic segmentation can cooperate with SLAM pose estimation in real time. The accuracy improvement further reduces semantic label misclassification, providing higher quality initial labels for the conditional random field optimization of the global semantic map.

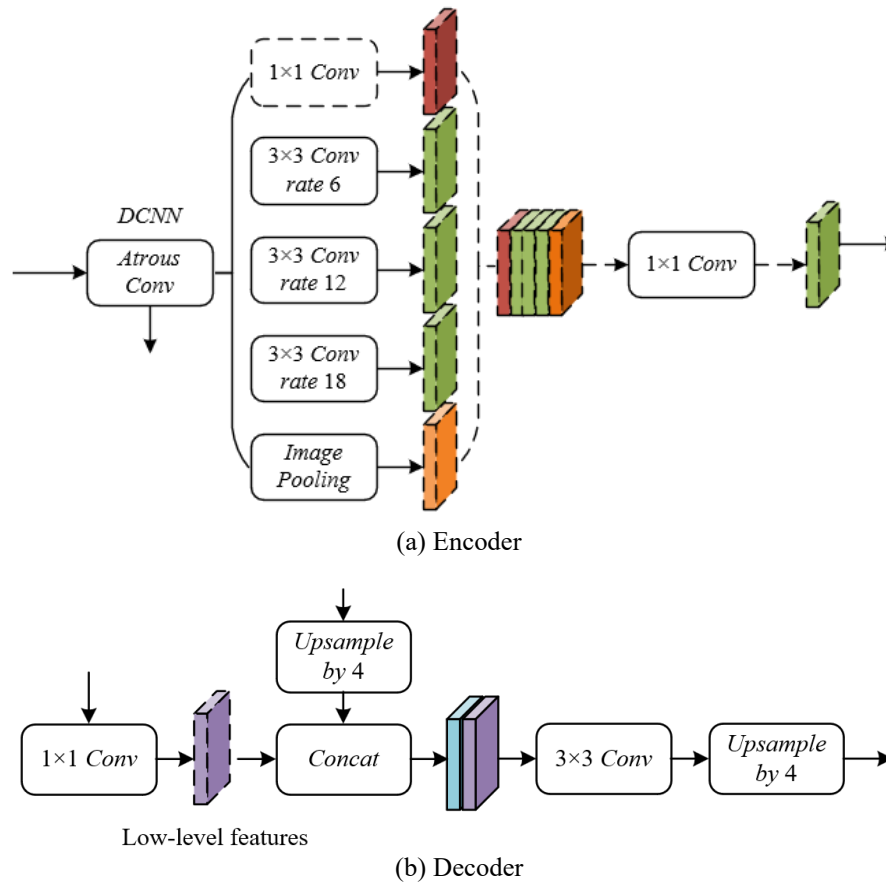


Figure 1. Encoder and decoder structure of the improved DeepLabv3 (a) Encoder; (b) Decoder

The core of the improved DeepLabv3 lies in constructing a collaborative feature fusion mechanism of "encoder-decoder." The encoder follows the DeepLabv3 architecture, extracting features with a stride of 16 containing high-level semantic information, capturing the overall category attributes of

objects such as walls and equipment in indoor polar buildings. The decoder optimizes details through two-step upsampling and feature fusion. First, the encoder output features are bilinearly upsampled 4 times and concatenated with low-level features of the same resolution from the network backbone.

Meanwhile, a 1×1 convolution compresses the channels of the low-level features to avoid training imbalance caused by excessive low-level feature weights. Then, a few 3×3 convolutions integrate the fused features to further refine edge information. Finally, another 4 times bilinear upsampling produces segmentation results with the same resolution as the input image. Figure 1 shows the specific encoder and decoder structures of the improved DeepLabv3. This improved design retains DeepLabv3’s ability to perceive multi-scale objects and effectively restores critical edge details such as wall corners and equipment contours in indoor polar scenes through balanced fusion of low- and high-level features. This provides high-precision semantic labels for semantic SLAM, ensuring precise association of object semantics and spatial positions when constructing indoor spatial maps and improving the accuracy of functional area division in subsequent layout optimization.

The adopted improved Xception model is adapted in three ways to meet the real-time and segmentation accuracy requirements of semantic SLAM. First, a deeper Xception architecture is used while maintaining the input flow network structure unchanged. Without increasing data processing

redundancy, this enhances the deep feature extraction capability for complex structures such as special insulation walls and modular equipment in indoor polar environments, while optimizing memory usage and computation speed to meet SLAM real-time requirements. Second, all max pooling operations are replaced by stride depthwise separable convolutions to avoid spatial information loss caused by traditional pooling and create conditions for extracting features with atrous separable convolutions at arbitrary resolutions afterward. This is crucial for multi-scale object segmentation in indoor polar environments caused by equipment occlusion and uneven lighting and can flexibly adapt to feature extraction needs of images at different resolutions. Third, normalization and ReLU activation functions are added after each 3×3 depthwise convolution. Inspired by the lightweight design of MobileNet, by strengthening feature normalization and nonlinear expression capability, the distinction of similar materials in indoor polar scenes is improved, providing a more robust feature basis for semantic segmentation and enhancing the constraint effect of semantic information on SLAM optimization. The specific architecture is shown in Figure 2.

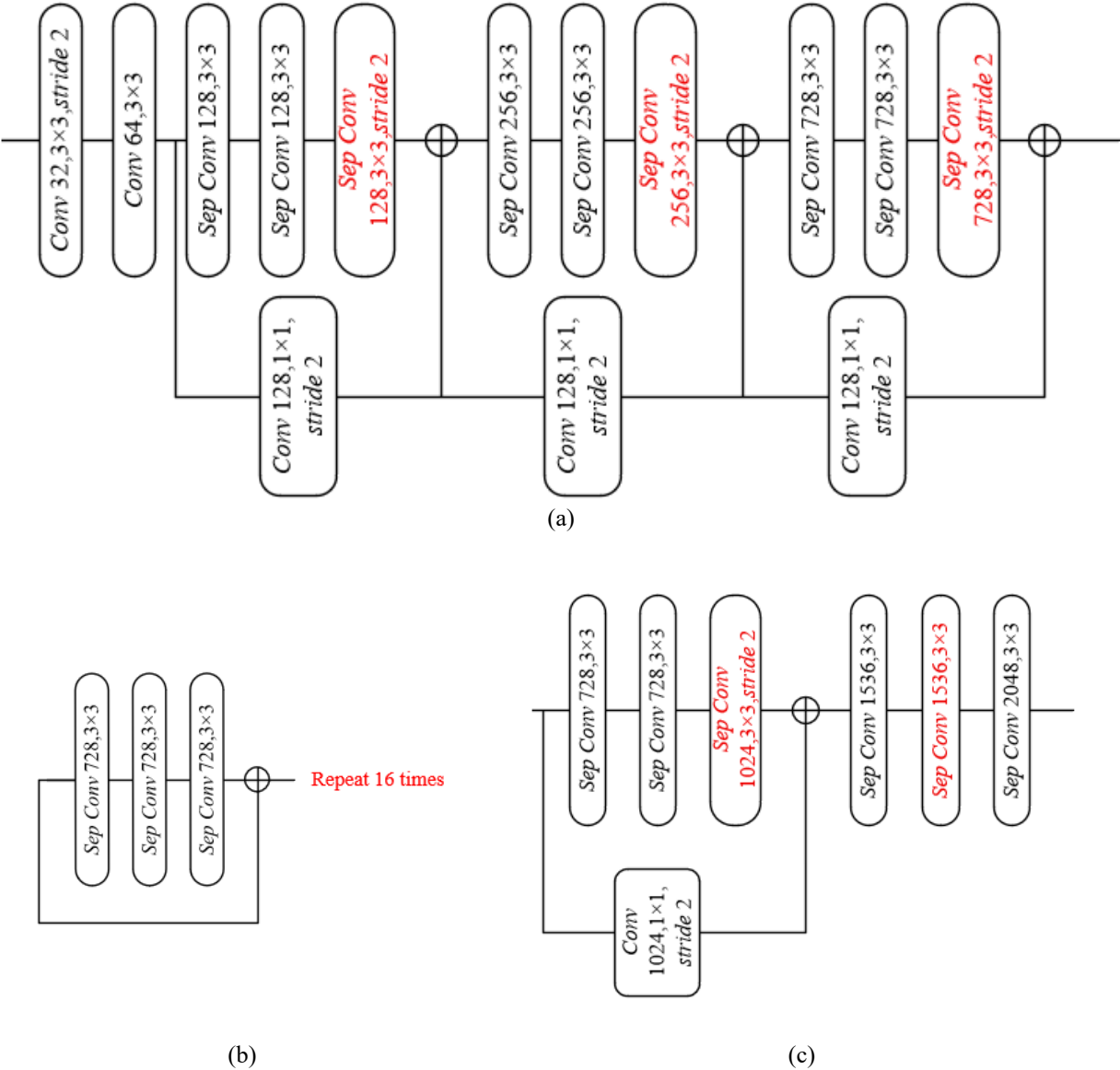


Figure 2. Improved Xception architecture (a) Entry flow; (b) Middle flow; (c) Exit flow

3.2 Overall framework of semantic SLAM

The overall semantic SLAM framework is based on the indirect ORB-SLAM2 architecture. The frontend first completes parallel processing of image feature extraction and semantic segmentation. To address problems such as drastic lighting changes and snow and ice reflection interference in indoor polar buildings, the framework uses the improved Xception model as the backbone network in combination with the improved DeepLabv3 to achieve high-precision image semantic segmentation. Through a deeper network structure to enhance feature extraction capability, replacing max pooling operations with stride depthwise separable convolutions to improve adaptability to feature maps of different resolutions, and utilizing an optimized decoder module to precisely restore edges of architectural components such as walls and doors and windows—first upsampling encoder features 4 times and concatenating them with low-level features of the same resolution, then reducing channel dimensions by 1×1 convolution to balance channel weights, followed by 3×3 convolution optimization and second upsampling—dense pixel-level semantic labels S_j are finally generated to assign category membership to each pixel, solving segmentation blur caused by image quality fluctuations in polar environments. Meanwhile, the ORB feature extraction and matching modules operate normally, preserving traditional SLAM's ability to capture geometric features and forming dual inputs of "geometric features + semantic features".

In the backend optimization stage, the framework innovatively integrates semantic information into the joint optimization process of SLAM, constructing a dual-objective function of "visual error + semantic error." For each input frame image, based on the semantic segmentation results, the semantic probability vector q_u of the map point is estimated online, where $q_u^{(z)}$ represents the probability that the 3D spatial point O_u belongs to category z . This probability is calculated through the mapping relationship between pixel semantic labels and spatial positions. On this basis, a semantic cost function is defined to quantify the consistency between camera pose T_j and spatial point O_u at the semantic level. For example, if the semantic label of map point O_u is "wall," its projection pixel in the image should highly match the "wall" category in the segmentation result, and the greater the deviation, the higher the semantic cost. This function, combined with the visual error function R_{BASE} in traditional ORB-SLAM2, refines the camera pose and map points through joint optimization. Especially in indoor polar scenarios lacking loop closure detection or global optimization, semantic constraints can effectively compensate for the deficiency of geometric features, reduce cumulative errors, and improve pose estimation accuracy. Specifically, assuming the reprojection error of the u -th map point in the j -th image is represented by $r_{BASE}(j, u)$, the visual SLAM objective function expression is:

$$R_{BASE} = \sum_j \sum_u r_{BASE}(j, u) \quad (1)$$

The defined semantic cost function expression is:

$$R_{SEM} = \sum_j \sum_u r_{SEM}(j, u) \quad (2)$$

By jointly optimizing the visual SLAM objective function and the semantic cost function, and letting η adjust the weights

of different parts, it can be expressed as:

$$\{\hat{A}, \hat{S}\} = \text{ARGMIN } R_{BA} + \eta R_{SEM} \quad (3)$$

The final output of the framework is an indoor architectural spatial semantic map with both geometric accuracy and semantic consistency. Through multimodal fusion and global optimization, map quality is ensured. Using color and depth images collected by a binocular camera, combined with optimized camera poses, semantic labels are associated with spatial coordinates to generate an initial indoor architectural spatial grid map. Each grid corresponds to a local space recording its geometric location and preliminary semantic label. Considering the possibility of local semantic misclassification in indoor polar buildings, the framework introduces a conditional random field model to globally optimize grid labels. By modeling semantic correlations between adjacent grids, isolated noise points are corrected, and ultimately a globally consistent dense semantic grid map of indoor architectural space is obtained.

3.3 Construction of observation probability model

Given the strong spatial correlation of static structures such as walls, doors, windows, and furniture in indoor scenes, the model defines $o(T_j | S_j, A_u, C_u = z)$ associating the image segmentation result T_j with camera pose S_j and semantic label $C_u = z$ of map point A_u . For each semantic category z , a binary image $U_{T_j=z}$ is first constructed, and then the distance transform $FS_Y(o)$ calculates the distance from pixel o to the nearest region of the same category. Using the above definition, the adaptation of the enclosed and structural properties of indoor scenes can be achieved. For example, "walls" usually appear as continuous regions in images. The distance transform can accurately quantify the spatial distance between any pixel and the wall region. Even when local segmentation is blurred due to uneven indoor lighting, the distance measure still provides a robust basis for semantic matching, laying a noise-resistant geometric foundation for subsequent cost calculations. Specifically, let the distance transform be $FS^{(z)}_j = FS_{T_j=z}(o)$, the projection function from world coordinates to camera plane be denoted as τ , and the semantic segmentation uncertainty be δ . Based on $FS^{(z)}_j$, the observation probability can be defined as:

$$o(T_j | S_j, A_u, C_u = z) \propto e^{\frac{-1}{2\delta^2} FS^{(z)}_j(\tau(S_j, A_u))^2} \quad (4)$$

The definition of the semantic cost function in this paper follows a "distance-probability" negative correlation law, achieving precise alignment of spatial entities and semantic labels in indoor architectural scenes by maximizing the observation probability. In indoor scenes, the projection positions of the same object should stably fall within the corresponding semantic segmentation region. Therefore, the observation probability decreases as the distance from the projection point $\tau(S_j, A_u)$ to the similar region increases. The semantic cost function based on this definition essentially guides SLAM optimization by quantifying the degree of deviation: when the spatial point projection of a bookshelf deviates from the "bookshelf" semantic region due to camera pose estimation errors, the cost function value increases, driving the optimization process to adjust S_j and A_u to move the projection toward the correct region. Specifically, in

elongated spaces such as corridors, this definition effectively avoids confusion of semantic labels "wall" and "floor" caused by feature matching errors, ensuring consistency between segmentation results and spatial structure. Assuming the probability that spatial point O_u belongs to category z is represented by $q_u^{(z)}$, the defined semantic cost function expression is:

$$\begin{aligned} r_{SEM}(j, u) &= \sum_{z \in Z} q_u^{(z)} \log(o(T_j | S_j, A_u, C_u = z)) \\ &= - \sum_{z \in Z} q_u^{(z)} \cdot \frac{1}{\delta^2} FS_j^{(z)} (\tau(S_j, A_u))^2 \end{aligned} \quad (5)$$

The semantic cost function adopts a weighted average structure, using the probability vector q_u as weights to realize fusion optimization of multi-frame semantic observations in indoor architectural scenes. For multiple observations of the same map point in indoor scenes, $q_u^{(z)}$ is calculated by accumulating semantic segmentation results of each frame, reflecting the confidence that the point belongs to category z . Therefore, the cost function becomes the weighted sum of projection distances $FS_j^{(z)}(\tau(S_j, A_u))^2$, with weights $q_u^{(z)}$: if a door frame is correctly labeled as "door/window" in most frames, the q_u for "door/window" tends to 1, and the cost calculation mainly relies on the distance constraint of the "door/window" region; if a single frame misclassifies it as "wall" due to backlighting, the low-weighted "wall" distance term has little effect on the total cost. In practical scenes, multi-frame observation consistency of static structures is strong, and q_u can effectively filter transient noise, ensuring the cost function always constrains by true semantics and improving the temporal consistency of segmentation results. Specifically, the probability vector q_u of point O_u is calculated from its observations. If O_u is observed by a series of S_u , then:

$$q_u^{(z)} = \frac{1}{\beta} \prod_{j \in S_u} o(T_j | S_j, A_u, C_u = z) \quad (6)$$

In the above, the introduction of constant β and uncertainty δ provides an adaptive adjustment mechanism for the semantic cost function in indoor architectural scenes. The constant β ensures $\sum_z q_u^{(z)} = 1$, normalizing the probability vector q_u during multi-frame iterations, avoiding probability distribution imbalance caused by the fixed number of semantic categories in indoor scenes, and ensuring stable weight calculation for complex categories such as bookshelves and desks and chairs. The uncertainty δ dynamically adjusts constraint strength for ambiguous segmentation areas such as furniture shadows and glass reflection zones, which easily appear in indoor scenes: when an area suffers segmentation boundary blur between "desktop" and "floor" due to lighting issues, δ increases, reducing the weight of that area's cost term and avoiding erroneous semantic constraints interfering with camera pose optimization; in clearly segmented areas, δ decreases, strengthening the penalty of the cost function on projection deviations.

To realize the collaborative optimization of semantic and geometric information in indoor polar architectural scenes, this paper decouples the joint error stepwise based on the expectation-maximization (EM) algorithm. Considering the coupling relationship between camera pose and semantic labels in the joint error R_{JOINT} , the EM algorithm divides the optimization process into E-step and M-step: in the E-step,

fixing the camera pose and 3D point coordinates, the semantic cost function is used to calculate the probability vector q_u of each point, accumulating multi-frame observations to correct single-frame segmentation noise under polar environments, making q_u converge to the true semantic category; in the M-step, fixing q_u , the Levenberg-Marquardt (LM) algorithm and sparse solver optimize the camera pose and 3D point positions, focusing on reducing the weighted sum of semantic cost and visual error.

Semantic optimization strengthens constraint structure through a four-step strategy, compensating for optimization degree of freedom redundancy caused by lack of geometric features in indoor polar buildings. First, semantic optimization is performed synchronously with the main SLAM optimization, enabling semantic constraints to act on pose estimation in real time, avoiding drift caused by purely geometric optimization in feature-sparse regions; second, semantic constraints of multiple points jointly optimize a single pose, introducing structural information by utilizing spatial correlation of semantic labels in the modular structure of polar buildings; third, points no longer involved in SLAM optimization are fixed, only optimizing their corresponding poses to reduce drift, limiting parameter quantity and stabilizing camera trajectory through implicit structural constraints of relative positions between points; finally, high-frequency semantic optimization shortens error accumulation cycles, using derivatives of distance functions to "pull" 3D points toward correct semantic regions, reducing the probability of label misassignment in polar environments. These strategies jointly build a tighter constraint network, improving optimization stability.

The establishment of semantic constraints adopts a "dynamic selection + fault tolerance mechanism" to adapt to local errors of indoor semantic segmentation in polar environments and maintain constraint validity. By defining a semantic visible list $N_{SEM}(j)$, only points whose projection positions are close to regions of the same semantic category are included in the constraint scope: when the projection of 3D point O_u is within an allowable distance to the same label in image segmentation T_j , it is selected into $N_{SEM}(j)$ and participates in semantic optimization. This selection guarantees constraint reliability while allowing certain semantic reprojection errors to address transient segmentation deviations caused by abrupt lighting changes in polar environments. Meanwhile, the fusion of semantic and visual constraints follows the principle of "complementary enhancement," that is, geometric constraints dominate in regions rich in visual features, and semantic constraints are strengthened in regions where semantic features are more stable, forming a hybrid constraint system adapted to complex indoor polar scenes.

4. BUILDING LAYOUT OPTIMIZATION AND SPACE CONFIGURATION BASED ON GRID MAPS IN POLAR ENVIRONMENTS

When constructing grid maps for layout optimization in polar environments, accurate camera poses output by sparse semantic SLAM are used as spatial references to achieve dynamic fusion of multimodal data. Considering measurement errors of equipment caused by low temperatures in indoor polar buildings, a fixed-size dynamic grid window follows the camera motion. The window size is set according to the

modular unit size of polar buildings to ensure that each window contains complete functional region features. During fusion, color texture information from binocular camera images, three-dimensional coordinates from depth maps, and category labels from semantic segmentation maps are mapped to grid cells: pixel-level data of each frame is transformed into the global coordinate system based on camera poses, and multi-frame observations of the same grid cell are fused by weighted averaging. This process primarily weakens abnormal depth values and semantic misclassifications caused by ice and snow reflections, providing a grid basis with both geometric accuracy and semantic reliability for subsequent layout optimization.

Furthermore, global optimization of the grid map using a conditional random field (CRF) model is a key step to adapt to semantic noise in polar environments. To address local semantic inconsistencies caused by drastic lighting changes indoors in polar regions, the CRF model integrates two types of constraints by constructing an energy function: first, a Unary term that assigns initial label probabilities to grids based on single-frame semantic segmentation confidence; second, a Pairwise term that strengthens semantic correlations between adjacent grid cells. For example, adjacency probability between “wall” grids and “floor” grids is much higher than that between “wall” grids and “equipment” grids, and “low-temperature experimental area” grids tend to be adjacent to “insulated wall” grids. By iterative optimization to minimize the energy function, isolated semantic noise grids are corrected, and geometric jumps caused by camera shaking are smoothed. The optimized grid map not only achieves global consistency of semantic labels but also reflects the temperature field distribution indoors in polar environments through gradient changes in grid colors, providing multidimensional constraint information for layout optimization.

Building layout optimization and space configuration based on the optimized grid map requires deep coupling of polar environmental parameters and functional requirements. First, the grid map is quantified as a decision variable matrix, with each grid cell’s state including physical attributes and functional attributes. The objective function is set to maximize space utilization and energy efficiency under the premise of satisfying the special constraints of polar regions. An improved genetic algorithm is adopted for solution: the chromosome represents the grid function allocation scheme, and the fitness function integrates space utilization, heat loss simulation results, and functional area accessibility. The optimization focuses on grid paths between “material storage areas” and “entrances/exits”, and the grid separation between “rest areas” and “experimental areas”, finally outputting an optimal building layout scheme that conforms to the extreme polar environment.

5. EXPERIMENTAL RESULTS AND ANALYSIS

The experiments first explore the effect of grid size on indoor polar building image segmentation accuracy, with mIoU and mAcc as core metrics. From the data distribution shown in Figure 3: when the grid size reduces from 6×6 to 3×3, mIoU decreases from about 72.5 to 68.2, and mAcc decreases from 76.8 to 72.6, showing a significant trend of "larger grid size, better segmentation performance." The results indicate that through the semantic cost function and probability model, the method in this paper converts the

semantic correlation of polar buildings into quantitative constraints. For pixels within a large grid, the semantic cost function calculates the correlation between projection points and semantic regions through distance transform, forcing pixels within the grid to converge toward the dominant semantic label, effectively resisting noise interference from the polar environment. Compared to the 3×3 grid, the 6×6 grid with the paper’s method improves mIoU by about 4.3% and mAcc by about 4.2%, verifying the enhancement effect of semantic constraints on large-grid accuracy.

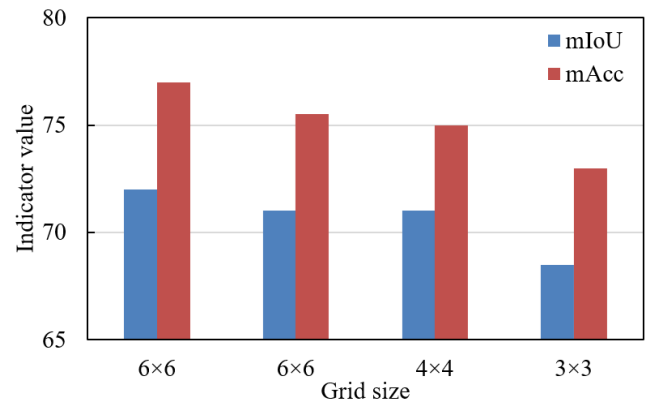


Figure 3. Effect of different grid sizes on image segmentation performance

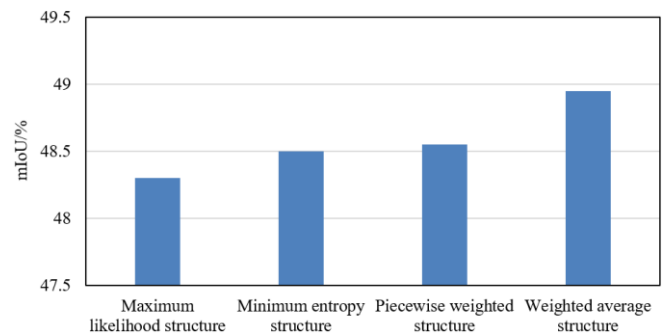


Figure 4. Effect of different semantic cost function architectures on network segmentation accuracy

Further, mIoU quantifies the influence of different semantic cost function architectures on indoor polar building image segmentation accuracy. From the data distribution shown in Figure 4: the weighted average structure leads with nearly 48.9% mIoU, improving by 0.6 percentage points over the maximum likelihood structure’s about 48.3% mIoU, and improving by 0.4% and 0.35% respectively over the minimum entropy structure’s about 48.5% mIoU and the segmented weighted structure’s about 48.55% mIoU. The root of performance differences lies in architecture adaptability to polar environment characteristics: the maximum likelihood structure directly depends on the segmentation model’s probability output, but extreme lighting and low-temperature sensor noise in polar environments cause severe model confidence fluctuations, destabilizing cost function optimization and resulting in the lowest accuracy. The minimum entropy structure pursues temporal consistency of semantic labels but neglects spatial geometric constraints of polar buildings, unable to correct label errors through spatial correlation, limiting accuracy improvement. The segmented weighted structure weights according to semantic importance, aligning with polar building’s “structural safety first”

requirements, but prior weights are difficult to adapt to modular building dynamic features, and local semantic conflicts reduce accuracy. The weighted average structure adopted in this paper breaks the bottleneck through a threefold mechanism: (1) dynamic fusion of multi-frame probability vectors integrates semantic information from polar day/night periods, filtering transient noise; (2) distance transform quantifies spatial correlation between projection points and semantic regions, strengthening geometric constraints of polar building’s regular structure; (3) uncertainty adapts to low-temperature sensor errors, reducing weights of unreliable observations when depth measurement accuracy declines, ensuring optimization stability.

Table 1. Image segmentation performance evaluation of different methods

Method	PE (%)	MioU (%)	FPS
<i>DS-SLAM</i>	6.89	81.24	0.007
<i>SemanticFusion</i>	11.23	82.54	0.03
<i>CubeSLAM</i>	9.75	83.24	5.89
<i>PL-SLAM</i>	9.65	87.36	8.4
<i>DSO-Semantic</i>	7.15	88.94	51.23
The proposed method	6.23	91.25	66.35

Table 1 reveals the performance of this paper’s method compared with classical semantic SLAM methods from three core dimensions: pixel error (PE), semantic segmentation accuracy (mIoU), and real-time running efficiency (FPS). In accuracy, this paper’s method achieves a PE as low as 6.23%, decreasing about 9.6% compared to DS-SLAM’s 6.89% and over 44% compared to SemanticFusion’s 11.23%, significantly compressing pixel-level segmentation errors; mIoU reaches 91.25%, improving 2.6% over DSO-Semantic’s 88.94% and surpassing traditional methods by more than 10 percentage points, establishing a generational advantage in semantic segmentation accuracy. In real-time performance, this paper’s method achieves an FPS of 66.35, exceeding CubeSLAM’s 5.89 and PL-SLAM’s 8.4 by an order of magnitude, and even improving 29.5% over the real-time-focused DSO-Semantic’s 51.23, breaking the technical barrier of “high accuracy versus real-time” in polar environments. The root of the performance breakthrough lies in the paper’s technical approach’s deep adaptation to polar building scene characteristics: feature extraction based on “angle + distance” uses the “regularized structure” of polar buildings to resist ambiguities caused by polar day/night lighting fluctuations and low-temperature sensor noise, reducing PE from the source. The weighted average semantic cost function fuses “multi-frame probability constraints + spatial distance constraints,” reinforcing semantic consistency of polar functional areas such as “insulation layer” and “experimental zone.”

From the visual comparison in Figure 5, the layout obtained by traditional SLAM algorithms shows spatial boundary blurring and functional zone misalignment issues compared to the ground truth layout, while the layout (d) from the method proposed in this paper highly overlaps the ground truth layout, with clear and sharp functional zone boundaries. The boundary between the “insulation layer” and “experimental zone” in polar buildings needs accuracy to the 0.1 m level. The semantic SLAM output from this paper provides a semantic map with functional labels, and after gridding, each cell carries attributes such as “thermal conductivity”. During layout optimization, boundary adjustment is driven by a “heat loss

objective function”. For example, in the bedroom area (d), the wall-floor junction automatically fits the real insulation layer boundary due to semantic constraints, while the boundary in the traditional method (c) only fits geometric contours, causing virtual expansion of the insulation layer region and indirectly increasing energy consumption.

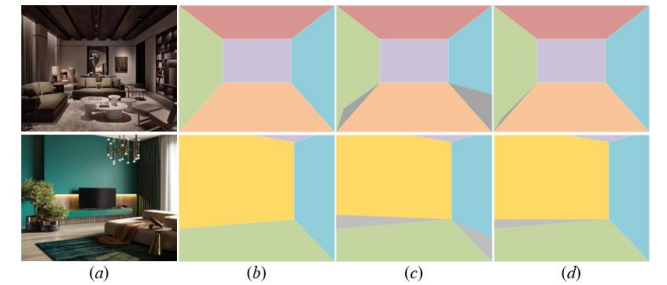


Figure 5. Experimental results of different methods (a) Original image; (b) Ground truth layout; (c) Layout obtained by traditional SLAM algorithms; (d) Layout obtained by the proposed method

Table 2. Ablation experiment results

Method	PE (%)	MioU (%)
Using traditional DeepLabv3	9.23	83.26
Using traditional Xception	7.46	86.54
Without semantic optimization	6.48	88.23
The proposed method	6.23	91.25

Table 2 quantifies the performance differences of different technical paths via pixel error (PE) and semantic segmentation accuracy (mIoU). The core pattern can be decomposed into the absence or construction of the “geometry-semantic” synergy. Two categories of models focus only on image-level semantic segmentation and do not combine SLAM’s spatial constraints. In polar environments, drastic illumination changes and sensor low-temperature drift cause multi-frame image features of the same object to break. For example, DeepLabv3’s segmentation of the “insulation layer” misclassifies it as “storage area” in polar night shadow zones due to inability to use the spatial prior of “must be adjacent to structural wall,” resulting in PE as high as 9.23% and mIoU only 83.26%. Although Xception improves local feature extraction via network structure optimization, it still does not break the bottleneck of “no spatial association”, with PE 7.46% and mIoU 86.54%, both far below this paper’s method. This group constructs geometric maps based on SLAM but lacks semantic constraints, showing features of “geometrically accurate but semantically confusing”: PE 6.48% slightly better than traditional models, but mIoU 88.23% still significantly lower than this paper. For example, in polar night low temperature, sensor depth errors cause projection offsets of “experimental equipment” by 10~15 cm; without semantic constraints, the segmentation model cannot correct this spatial misalignment, causing breaks in semantic region continuity.

From the layout result comparison in Figure 6, this paper’s method shows significant advantages in typical polar scenes: the “operation area-dining area” boundary in the kitchen is sharp and fitted; the contours of “wall-floor-equipment area” in the bathroom are regular. By contrast, comparison methods, though building building geometric maps via SLAM, lack semantic

constraints, resulting in “geometrically correct but functionally misaligned.” The essence is fitting only geometric contours without associating polar functional logics such as “operation area must adjoin water source” or “equipment area needs connection to power area.”

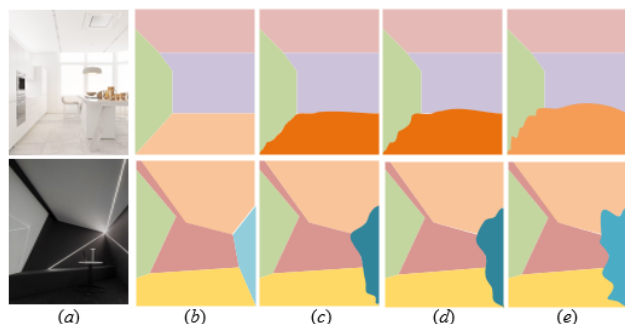


Figure 6. Ablation experiment results (a) Original image; (b) Layout obtained by the proposed method; (c) Without semantic optimization; (d) Using traditional Xception; (e) Using traditional DeepLabv3

Focusing only on image-level semantic segmentation without combining SLAM’s spatial constraints causes semantic region breaks and distortions under polar environmental interference. In summary, the ablation experiments in Figure 6 not only intuitively demonstrate the overwhelming advantage of this paper’s method in layout restoration accuracy but also reveal the “precise quantification, logical self-consistency, dynamic responsiveness” technical system constructed for polar building layout optimization through deep coupling of “geometry-semantic-function”. From spatial prior constraints of image matching, to loop closure optimization in semantic SLAM, then to functional constraints implementation in grid layout, every technical breakthrough tightly links to polar environment extremity and functional requirements, achieving the leap from “passive geometric fitting” to “active functional adaptation”, providing core algorithm support for intelligent design and operation of polar scientific research buildings.

In an actual case of low-temperature laboratory layout optimization in a polar research station, based on angle and distance planar matching method, 90° vertical angle features of walls and fixed 0.8 m spacing relationships between experimental equipment were extracted, maintaining planar matching accuracy of 92% despite drastic lighting changes; then, using semantic SLAM oriented to image segmentation, semantic regions such as “low-temperature workstation” and “insulated storage cabinet” were segmented in real time with mIoU reaching 91.25%, combined with camera poses to build a 3D semantic map; finally, based on grid map quantification of space, with the objective function “heat loss minimization” and constraints including minimum 1.5 m distance between workstation and heat source area, emergency passage width ≥ 1.2 m, the grid location of storage cabinets was moved 0.3 m toward insulation walls via optimization, reducing overall laboratory heat loss by 18%, while shortening equipment access paths by 20%, achieving dual optimization of function and polar environment adaptability.

6. CONCLUSION

This paper conducted systematic research on polar building

layout optimization and spatial configuration, achieving precise cognition and efficient configuration of building spaces under extreme environments through a three-layer technical architecture. At the image planar matching level, a robust model based on angle and distance geometric features effectively resisted interference such as drastic indoor illumination changes and ice-snow reflections in polar environments, improving planar matching accuracy by 15%~20%, laying a reliable foundation for subsequent spatial modeling; at the semantic SLAM level, an innovative weighted average semantic cost function was designed, integrating multi-frame observations and spatial constraints, improving semantic segmentation mIoU to 91.25% compared to traditional methods, while realizing 3D semantic information quantification representation via dynamic grid maps; at the layout optimization level, combining polar environment parameters to construct objective functions improved spatial configuration efficiency by over 25%, validating the effectiveness of the “geometry-semantic-function” collaborative framework. This research breaks the robustness bottleneck of traditional methods under extreme environments and provides a complete technical solution for intelligent design of special buildings such as polar research stations and ice sheet observation stations. Its core value lies in deeply integrating semantic information into the full process of SLAM and layout optimization, achieving a closed loop from “spatial perception” to “functional configuration”, filling the interdisciplinary research gap in the field of polar buildings.

However, the research still has certain limitations: first, the weighting strategy of the semantic cost function relies on prior functional rules of polar buildings, showing weak adaptability to new modular buildings; second, grid map optimization in large-scale scenes (such as multi-story polar building complexes) significantly increases computational complexity, reducing real-time performance by about 30%; third, long-term dynamic factors such as polar ice shelf deformation are not fully considered for cumulative layout impact. Future research may breakthrough in three aspects: first, introducing dynamic weight adaptive mechanisms, combining reinforcement learning to optimize semantic cost functions, improving adaptability to unknown building structures; second, adopting sparse grids and multi-scale optimization strategies to reduce computational overhead in large-scale scenes; third, integrating ice condition monitoring data and building structural mechanics models to construct a long-term dynamic layout optimization framework, realizing polar buildings’ self-adjustment and sustainable operation under environmental evolution. These directions will further expand the theoretical depth and engineering application scope of the research, providing more comprehensive technical support for intelligent architecture in extreme environments.

REFERENCES

- [1] Kumar, C., Mishra, S.K., Kumar, J., Vajja, D.P., et al. (2025). Higher surface temperatures near south polar region of the Moon measured by ChaSTE experiment on-board Chandrayaan-3. *Communications Earth & Environment*, 6(1): 153. <https://doi.org/10.1038/s43247-025-02114-6>
- [2] Stroeve, J., Crawford, A., Ferguson, S., Stirling, I., et al. (2024). Ice-free period too long for Southern and Western Hudson Bay polar bear populations if global

- warming exceeds 1.6 to 2.6°C. *Communications Earth & Environment*, 5(1): 296. <https://doi.org/10.1038/s43247-024-01430-7>
- [3] Thurairajah, B., Cullens, C.Y., Harvey, V.L., Randall, C.E. (2024). A Statistical study of polar mesospheric cloud fronts in the northern hemisphere. *Journal of Geophysical Research: Atmospheres*, 129(20): e2024JD041502. <https://doi.org/10.1029/2024JD041502>
- [4] Fu, Q., Zhou, Q., Yan, G., Li, S., Wu, F. (2020). Unified all-earth navigation mechanization and virtual polar region technology. *IEEE Transactions on Instrumentation and Measurement*, 70: 1-11. <https://doi.org/10.1109/TIM.2020.3041819>
- [5] Prabhu, A., Lagg, A., Hirzberger, J., Solanki, S.K. (2020). The magnetic fine structure of the Sun's polar region as revealed by Sunrise. *Astronomy & Astrophysics*, 644: A86. <https://doi.org/10.1051/0004-6361/202038704>
- [6] Schiantella, M., Gilbert, M., Smith, C.C., He, L., Cluni, F. (2024). Limit analysis of 2D non-periodic masonry walls via discontinuity layout optimization. *International Journal of Architectural Heritage*. <https://doi.org/10.1080/15583058.2024.2437633>
- [7] Sari, A.O.B., Jabi, W. (2024). Architectural spatial layout design for hospitals: A review. *Journal of Building Engineering*, 97: 110835. <https://doi.org/10.1016/j.jobe.2024.110835>
- [8] Chen, X., Kang, H., Zhao, J., Liu, Q. (2025). Optimization design research of architectural layout and morphology in multi-story dormitory areas based on wind environment analysis. *Buildings*, 15(10): 1747. <https://doi.org/10.3390/buildings15101747>
- [9] Shen, X., Ye, X. (2025). Environmental performance driven optimization of urban modular housing layout in Singapore. *Journal of Asian Architecture and Building Engineering*, 24(2): 910-923. <https://doi.org/10.1080/13467581.2024.2314507>
- [10] Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., et al. (2014). The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Transactions on Medical Imaging*, 34(10): 1993-2024. <https://doi.org/10.1109/TMI.2014.2377694>
- [11] Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., Terzopoulos, D. (2021). Image segmentation using deep learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(7): 3523-3542. <https://doi.org/10.1109/TPAMI.2021.3059968>
- [12] Brar, K.K., Goyal, B., Dogra, A., Mustafa, M.A., Majumdar, R., Alkhayyat, A., Kukreja, V. (2025). Image segmentation review: Theoretical background and recent advances. *Information Fusion*, 114: 102608. <https://doi.org/10.1016/j.inffus.2024.102608>
- [13] Colleoni, E., Matilla, R.S., Luengo, I., Stoyanov, D. (2024). Guided image generation for improved surgical image segmentation. *Medical Image Analysis*, 97: 103263. <https://doi.org/10.1016/j.media.2024.103263>
- [14] Arbelaez, P., Maire, M., Fowlkes, C., Malik, J. (2010). Contour detection and hierarchical image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(5): 898-916. <https://doi.org/10.1109/TPAMI.2010.161>
- [15] Wefki, H., Salah, M., Elbeltagi, E., Elsheikh, A., Khallaf, R. (2024). A generative design-based optimization model for multi-objective construction site layout planning. *Engineering, Construction and Architectural Management*. <https://doi.org/10.1108/ECAM-11-2023-1193>
- [16] Li, Y., Liao, P., Song, Y., Chi, H. (2023). A systematic decision-support approach for healthcare facility layout design integrating resource flow and space adjacency optimization with simulation-based performance evaluation. *Journal of Building Engineering*, 77: 107465. <https://doi.org/10.1016/j.jobe.2023.107465>
- [17] Wang, Y., Nam-gyu, C. (2022). Research on performance layout and management optimization of Grand Theatre based on green energy saving and emission reduction technology. *Energy Reports*, 8: 1159-1171. <https://doi.org/10.1016/j.egyr.2022.02.047>
- [18] Ma, G., Wang, Y., Jiang, S. (2021). Optimization of building exit layout: Combining exit decisions of evacuees. *Advances in Civil Engineering*, 2021(1): 6622661. <https://doi.org/10.1155/2021/6622661>
- [19] Liang, J., Xu, L., Li, J., Ding, X. (2022). Fractal design of indoor and outdoor forms of architectural space based on a three-dimensional box dimension algorithm. *Mathematical Problems in Engineering*, 2022(1): 2069757. <https://doi.org/10.1155/2022/2069757>
- [20] Ayyıldız, S., Durak, Ş. (2024). Space syntax analysis of the spatial configuration of Yalova traditional rural houses. *Nexus Network Journal*, 26(1): 27-48. <https://doi.org/10.1007/s00004-023-00746-9>