Generative Design of Cold-Region Waterfront Ice and Snow Cultural Scenes Based on LiDAR Point Clouds and Generative Adversarial Network (GAN)

Yue Ge
Qiqihar University, Qiqihar 161006, Heilongjiang, China
Corresponding author: Yue Ge (e-mail: GEYUE15846205817@163.com)

ABSTRACT Accurate reconstruction and generation of complex cold-region waterfront environments remain challenging due to the coupled influence of geometric structures, environmental conditions, and physical constraints. This study proposes a multimodal conditional latent-space disentanglement and physically constrained adversarial generation framework (MC-PAG) for LiDAR-based 3D scene modeling and synthesis. The framework integrates LiDAR point-cloud measurements, semantic descriptors, and environmental parameters into a structured latent representation through multimodal feature encoding and latent-space disentanglement. A cascaded generative adversarial architecture is subsequently employed to progressively reconstruct dense spatial point-cloud distributions, while differentiable physical-consistency modules are embedded into the generation process to enforce environmental plausibility and geometric continuity. Furthermore, a multi-scale discrimination strategy is introduced to jointly evaluate spatial fidelity, semantic consistency, and physical-rule compliance. Experimental results on a dedicated cold-region waterfront dataset demonstrate that the proposed framework achieves a semantic segmentation IoU of 0.845 for ice–snow structures, maintains an average physical-rule violation rate of only 3.37%, and generates spatial layouts with an average path curvature of 0.12. The generated scenes further exhibit high geometric fidelity, environmental consistency, and controllable semantic representation. By establishing a unified pipeline for LiDAR sensing information processing, spatial environment reconstruction, and physically constrained scene generation, the proposed framework provides an effective methodology for digital environment modeling and propagation-aware spatial representation in complex outdoor environments.

INDEX TERMS LiDAR sensing, point cloud reconstruction, multimodal feature fusion, propagation environment modeling

I. INTRODUCTION

WITH the increasing impact of climate change on seasonal landscapes, the winter utilization of cold-region waterfront spaces and the inheritance of ice and snow culture have become cutting-edge topics across landscape architecture, urban-rural planning, and material science. This interdisciplinary focus ensures that the preservation of cultural heritage and the deployment of industrial installations contribute to the resilience and aesthetic value of high-latitude waterfront environments. Traditional scene design methods heavily rely on manual experience and static modeling, making it difficult to efficiently respond to dynamic changes in ice and snow environments, complex terrain hydrological coupling relationships, and diverse cultural expression needs [1,2]. In this context, the use of artificial intelligence generation technology, especially 3D vision based generation models, for data-driven automated or semi automated creation of cold-region waterfront ice and snow cultural scenes, demonstrates enormous potential for driving design paradigm change [3,4]. This research not only aims to provide technical tools for climate-adaptive spatial planning but also opens new avenues for digitally preserving, reproducing, and innovatively developing regional ice and snow culture, thereby demonstrating clear scientific and practical value. Current research faces multi-dimensional challenges. At the data level, publicly available 3D datasets severely lack detailed semantic annotations of the complex interfaces of "ice-snow-water-shore" and cultural structures such as ice sculptures and snow castles, resulting in a lack of high-quality, multimodal specialized data foundation for model learning [5,6]. On a technical level, general 3D generative models have significant limitations in handling

such scenes: voxel or mesh based methods are difficult to depict delicate geometric features and complex topologies of ice and snow; Although point cloud based methods can maintain geometric flexibility, their generation process often ignores the hard constraints of basic physical rules such as hydrology and thermodynamics, which can easily result in structures that violate geological knowledge (such as suspended ice bodies) [7,8]. At the application level, existing methods mostly focus on simulating visual authenticity or single physical attributes, and fail to deeply integrate advanced design intentions such as cultural semantics and functional layout as controllable conditions into the generation process, resulting in a disconnect between the generated results and actual planning and design requirements [9,10]. In response to data challenges, some studies have used a combination of manual modeling and physical simulation to construct small synthetic datasets, or utilized drone photogrammetry for scene reconstruction. However, these methods have bottlenecks in terms of scale, efficiency, and detail richness [11,12]. Regarding generation technology, existing work mainly advances along two paths: one is to improve the architecture of GAN to enhance the quality of point cloud generation, such as using tree structure generators or introducing attention mechanisms [13,14]; The second is to embed differential equation solvers into neural networks to enhance physical consistency, such as applying neural operators in fluid simulations [15,16]. However, the former usually does not explicitly model the dynamic boundary conditions of phase transition processes such as ice and snow accumulation and melting, as well as their interaction with the shore; Although the latter performs well in simulating a single physical field, it is difficult to couple multiple physical processes and simultaneously meet the global layout requirements of cultural semantics. These methods still lack a unified framework to synchronously optimize the geometric fidelity, multi-physics rationality, and cultural semantic controllability of generated scenes in an end-to-end manner. Therefore, this article focuses on solving the core problem of generating three-dimensional ice and snow scenes in cold regions with high physical credibility and cultural semantic controllability in the absence of specialized data. Therefore, this paper proposes a hybrid framework MC-PAG that decouples multimodal conditional latent spaces and generates physical constraint adversarial interactions. This method first designs a multimodal conditional encoder that maps sparse LiDAR point clouds, symbolic cultural semantic labels, and parameterized physical conditions to a structured latent space, and guides the different dimensions of this space to represent geometric, semantic, and physical properties through decoupling loss functions. Subsequently, a cascaded GAN was used to synthesize dense point clouds from decoupled latent variables, with each upsampling layer of its generator connected to a differentiable physical consistency verification module. This module imposes constraints on the intermediate generated results based on simplified material transport and phase transition equations, and feeds back quantization errors that violate physical rules as regularization terms to the adversarial training loss function, thereby driving the generator to spontaneously follow geological laws such as water ice coexistence and snow distribution while pursuing visual realism. The generation mechanism proposed in this article, which combines explicit decoupling and implicit constraints, achieves closed-loop control of physical rules and cultural semantics from conditional input to geometric output.

II. GENERATIVE DESIGN METHOD FOR COLD-REGION WATERFRONT ICE AND SNOW SCENES BASED ON MC-PAG

To solve the problem of multimodal condition fusion and physical law embedding in the generation of cold-region waterfront ice and snow cultural scenes, this paper proposes a hybrid framework based on multimodal condition hidden space decoupling and physical constraint adversarial generation. The overall architecture is shown in Figure 1. This framework consists of three core modules that form a complete closed loop from conditional encoding to adversarial optimization. The multimodal conditional coding and hidden space decoupling module is responsible for feature extraction and fusion of the original LiDAR point cloud, semantic labels, and physical parameters, and generates structured hidden space representations through decoupling learning to achieve separation and control of geometric, semantic, and physical attributes. The cascaded point cloud generator module embedded with physical constraints takes decoupled latent variables as conditions and gradually synthesizes point clouds through multi-level upsampling. Its key innovation lies in coupling a differentiable physical simulator to each level, which feeds back physical rules such as ice and snow accumulation and waterfront boundaries as real-time normalized losses to the generation process, ensuring the physical rationality of the output scene. The multi-scale discrimination and physical perception adversarial training module constructs a multi-channel discrimination network to evaluate and generate point clouds from multiple dimensions, including overall structure, local details, and physical rule consistency. The adversarial training mechanism drives the generator and discriminator to collaborate and optimize, ultimately achieving high fidelity and high controllability scene generation. The entire process embodies the systematic design concept of "condition encoding, physical constraint generation, and multi-dimensional adversarial optimization".

A. HIDDEN SPACE DECOUPLING CODING FOR MULTIMODAL CONDITIONAL INPUT

This section aims to build a structured conditional representation foundation that encodes heterogeneous information from LiDAR point clouds, semantic labels, and physical parameters into an internally decoupled joint hidden space. The decoupling property of this space is a key prerequisite for achieving high-dimensional controllable generation in

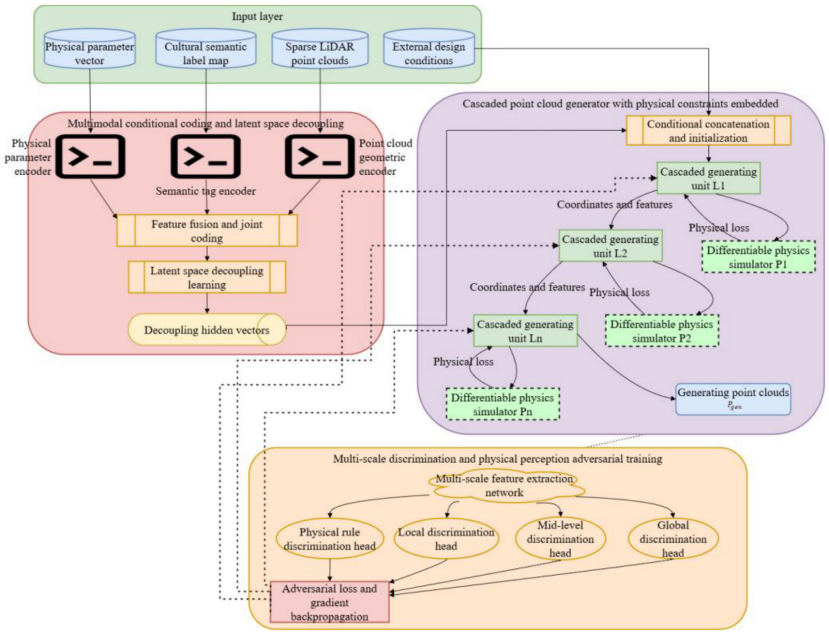

FIGURE 1. Intelligent generation framework for cold-region waterfront ice and snow scenes based on multimodal conditional latent space decoupling and physical constraint adversarial generation

the future, ensuring that the different attributes of the generator can be independently and accurately guided by the separated conditional signals. The input raw sparse LiDAR point cloud data is denoted as $P_{\text{raw}} \in \mathbb{R}^{N\times 3}$, where N is the number of points. The cultural semantic information is provided by a two-dimensional label map S aligned with the point cloud space, where each pixel corresponds to a predefined semantic category (such as ice surface, snow sculpture, water body). The physical conditions are represented by a set of scalar parameter vector $\Phi \in \mathbb{R}^K$, including temperature, wind speed, and other environmental variables. Extracting modality specific features through three parallel encoding networks. For geometric features, a point cloud encoder $E_g(\cdot)$ based on PointNet++ architecture is used to process P_{raw} [17,18]:

$$f_g = E_g(P_{\text{raw}}) \tag{1}$$

The encoder outputs a multi-scale geometric feature vector f_g that captures the scene from local details to global contours through hierarchical farthest point sampling and feature aggregation. For semantic features, use a lightweight convolutional neural network E_s to process the semantic label graph S:

$$f_s = E_s(S) \tag{2}$$

The encoder E_s extracts the semantic layout and composition relationship of the scene, and outputs the semantic feature vector f_s. The physical parameter vector Φ is encoded by a multi-layer perceptron E_p:

$$f_p = E_p(\Phi) \tag{3}$$

This network maps discrete physical parameters to continuous physical feature vectors f_p. Subsequently, the extracted multimodal feature vectors f_g, f_s, and f_p are fused. Adopting a strategy of splicing followed by nonlinear transformation, implemented through a multi-layer perceptron $\text{MLP}_{\text{fusion}}$:

$$f_{\text{joint}} = \text{MLP}_{\text{fusion}}([f_g; f_s; f_p]) \tag{4}$$

[;] represents the vector concatenation operation, and the resulting f_{joint} is a joint feature vector containing multi-source information interaction. The joint feature f_{joint} is fed into the encoder E_{joint} composed of fully connected layers and mapped into a joint latent vector $z \in \mathbb{R}^d$ that follows a Gaussian prior distribution [19,20]:

$$z = E_{\text{joint}}(f_{\text{joint}}) \tag{5}$$

To make the hidden space z interpretable and controllable, decoupling learning constraints based on minimizing Total Correlation (TC) are introduced. The total correlation TC(z) is used to measure the statistical dependence between the dimensions of latent variables, defined as the KL (Kullback Leibler) divergence of the joint distribution $q(z)$ multiplied by the marginal distribution $\prod_{i=1}^{d} q(z_i)$ [21,22]:

$$L_{\text{disent}} = TC(z) = D_{\text{KL}}\left(q(z) \middle\| \prod_{i=1}^{d} q(z_i)\right) \tag{6}$$

Minimizing this loss aims to force information to separate into different dimensions, achieve independence of hidden variables in each dimension, and spontaneously form decoupled subspaces corresponding to different attributes. To ensure that the hidden space encoding process does not lose key input information, while applying reconstruction loss as a constraint. Request to reconstruct the joint feature f_{joint} from the hidden vector z through a symmetric decoder network $D(\cdot)$:

$$L_{\text{recon}} = \|D(z) - f_{\text{joint}}\|_2^2 \tag{7}$$

By jointly optimizing the objective function that includes decoupling loss L_{disent} and reconstruction loss L_{recon}, this module ultimately learns a structured joint latent space. In this space, geometric, semantic, and physical information are decoupled into different subsets of dimensions, providing clear and separated control signals for the generator, making it possible to precisely control the corresponding attributes of the generated scene by independently modifying the hidden codes of specific dimensions.

B. CASCADED POINT CLOUD GENERATOR INTEGRATING PHYSICAL CONSTRAINTS

The cascaded point cloud generator G in this section takes the decoupled latent vector z and design conditions as inputs to generate a high-density point cloud P_{gen}. Its core is to synchronously embed differentiable physical simulations as process constraints at each stage of cascading upsampling, ensuring that the generated scene form is reasonable in both geometric details and physical laws. After linear projection of the input conditions, the initial point cloud seed $P_0 \in \mathbb{R}^{M_0 \times 3}$ and its feature F_0 are formed. The generation process is completed by L cascaded units. At the l-th stage, the output of the previous stage (P_{l-1}, F_{l-1}) is received and upsampled to the target point M_l through feature interpolation to obtain F_l^{up}. Subsequently, the coordinate prediction network Δ_l predicts the coordinate offset based the feature F_l on the features and the current position:

$$\Delta P_l = \Delta_l([P_{l-1}; F_l^{\text{up}}]) \tag{8}$$

Thus obtaining the preliminary point cloud $\widetilde{P}_l$ for this level. The key step is to embed the differentiable physics simulator $\mathcal{P}_l$ into this process. Apply constraints to P_l based on the current physical parameter Φ. For example, to simulate the stability of ice and snow accumulation, calculate the difference between the component of gravity acting on local point i along the normal direction n_i and the support force represented by the local density ρ_i:

$$L_{\text{phy}}^{(l)} = \frac{1}{M_l} \sum_{i=1}^{M_l} \text{ReLU}(g \cdot n_i - \kappa \cdot \rho_i) \tag{9}$$

κ is a parameter related to material strength, and ReLU (Rectified Linear Unit) ensures that only violations of physical equilibrium (i.e. gravity component greater than support force) are penalized [23,24]. This loss $L_{\text{phy}}^{(l)}$ is fed back in real-time to the training gradient of Δ_l, forcing the network to spontaneously follow physical rules when predicting coordinates. After adjusting the physical constraints, the final output point cloud P_l and feature F_l of this stage are obtained and used as inputs for the next stage. Finally, the generator outputs $P_{\text{gen}} = P_L$. Through this cascaded embedding mechanism, physical rules are integrated as an intrinsic prior during generation, rather than being applied as a separate post-processing step. This ensures that the output point cloud P_{gen} achieves both high geometric fidelity and physical plausibility.

C. MULTI-SCALE DISCRIMINATION AND PHYSICAL PERCEPTION ADVERSARIAL TRAINING MECHANISM

The discriminator network D in this section adopts a multi-scale architecture to jointly discriminate the point cloud P_{gen} output by generator G from three levels: overall structure, local details, and physical consistency. The core of this design is to drive generator optimization through adversarial training, while using a physical perception assisted discrimination head to explicitly incorporate the compliance of physical rules into the objective function of adversarial games, thereby forcing the generated results to meet predetermined physical laws while improving visual realism. The input of discriminator D is point cloud P. Firstly, three point cloud feature extraction networks with different receptive fields are used to process the input point cloud in parallel, extracting multi-scale features of global, intermediate, and local levels respectively. After independent pooling operations, these features are mapped to the corresponding scale's authenticity scalar scores s_{global}, s_{mid}, s_{local}. The multi-scale discrimination mechanism forces the generator to match the true data distribution at various levels from macro layout to micro geometry. On the basis of obtaining multi-scale authenticity features, the discriminator adds a physical rule consistency discrimination head. This module takes mid-level features as input and predicts the probability of the input point cloud violating rules on several key physical relationships through a classification network. The calculation of the physics violation score s_{phy} is based on the aggregation of the probability of violating these rules. Therefore, the final discrimination output of discriminator D for point cloud P is a composite vector $s = [s_{\text{global}}, s_{\text{mid}}, s_{\text{local}}, s_{\text{phy}}]$. Correspondingly, the adversarial training objective functions of generator G and discriminator D are extended to include multi-scale authenticity adversarial loss and physical consistency adversarial loss [25,26]. The goal of the generator is defined as minimizing the following losses:

$$\begin{aligned} L_G^{\text{adv}} = -\mathbb{E}_{z\sim p(z), c\sim p(c)} \Bigg[& \sum_{k\in\{\text{glob},\atop \text{mid},\atop \text{loc}\}} D_k(G(z,c)) \\ & - \lambda_{\text{phy}} \cdot D_{\text{phy}}(G(z,c)) \Bigg] \end{aligned} \tag{10}$$

D_k represents the authenticity discrimination score of the corresponding scale, and D_{phy} represents the physical violation score. λ_{phy} is the weight coefficient that balances the two terms. The goal of discriminator D is to accurately distinguish between real and generated point clouds, and to correctly evaluate their physical consistency. Its loss

function is:

$$
\begin{aligned}
L_D^{\text{adv}} = \mathbb{E}_{P_{\text{real}}} \Bigg[\sum_k \text{ReLU}(1 - D_k(P_{\text{real}})) \\
- \lambda_{\text{phy}} \cdot D_{\text{phy}}(P_{\text{real}}) \Bigg] \\
+ \mathbb{E}_{P_{\text{gen}}} \Bigg[\sum_k \text{ReLU}(1 + D_k(P_{\text{gen}})) \\
+ \lambda_{\text{phy}} \cdot D_{\text{phy}}(P_{\text{gen}}) \Bigg]
\end{aligned} \tag{11}
$$

By iteratively optimizing the above adversarial objectives, the final output point cloud P_{gen} of the generator achieves a balance between visual realism and physical rationality.

III. EXPERIMENTAL SETUP AND DATASET CONSTRUCTION

A. CONSTRUCTION AND PREPROCESSING OF DATASET FOR COLD-REGION WATERFRONT ICE AND SNOW CULTURAL SCENES

To address the lack of specialized data on cold-region waterfront ice and snow cultural scenes in public datasets, this study constructed a finely annotated multimodal dataset. The data was collected in typical cold urban waterfront areas, spanning multiple winter cycles to cover natural changes in ice and snow and artificial structures during festive periods. The collection adopts a collaborative scheme of ground and airborne LiDAR, and through the deployment of ground control points and multi-level registration processes, multiple period and multi-source point clouds are unified into the same coordinate system, forming a complete dense point cloud of the scene. The point cloud semantic segmentation annotation defines five core categories: natural water bodies, ice surfaces, snow cover layers, conventional vegetation and buildings, and artificial ice and snow cultural structures. The "artificial ice and snow cultural structures" have been subdivided into functional subcategories. The annotation is independently completed by multiple individuals and cross checked to ensure consistency. To adapt to the model input, the large-scale point cloud is cut into regular scene blocks, voxel downsampling and spatial normalization are performed, and corresponding two-dimensional semantic projections and physical parameter vectors are generated. The final dataset partitioning and statistical information are shown in Table 1.

TABLE 1. Statistical information of cold-region waterfront ice and snow scene dataset

Category	Number of scenes	Average points/scene	Number of semantic categories	Attached physical parameters
Training set	170	1.2M	6	Temperature, wind speed, humidity, time
Validation set	20	1.1M	6	Temperature, wind speed, humidity, time
Test set	30	1.3M	6	Temperature, wind speed, humidity, time

As shown in Table 1, the dataset is divided into training set, validation set, and testing set. The training set contains 170 scenarios, providing sufficient samples for model learning. The validation set and testing set are used for tuning and evaluation, respectively. All scenarios have a point cloud density of millions and six semantic labels, ensuring data richness and task specificity. The synchronized collection of physical parameters provides key conditional inputs for generating models and establishes a reliable data foundation for method validation.

B. COMPARISON METHODS AND EXPERIMENTAL PARAMETER SETTINGS

To comprehensively evaluate the performance of the proposed framework, this study selected four representative 3D scene generation methods as baselines for comparison. 1) The unconditional point cloud generation models are represented by r-GAN (recurrent Generative Adversarial Network) and TreeGAN (Tree-Generative Adversarial Network). R-GAN generates point sequences through a recurrent neural network structure, while TreeGAN captures the hierarchical structure of point clouds using a tree graph convolutional network. They do not accept any conditional input and only learn the overall distribution of the training data [27,28]. 2) The conditional scenario generation model selects SP-GAN (Scene Planning Generative Adversarial Network) [29]. This method first generates semantic layout diagrams, and then generates corresponding 3D point clouds based on the layout, which can respond to semantic conditions. 3) The two-stage baseline (SP-GAN+PBD) model is generated and simulated using SP-GAN as the conditional scene generator, with the same semantic layout as input to generate the initial 3D point cloud. Subsequently, an offline physics simulator based on the Position Based Dynamics (PBD) framework was introduced to apply gravity, collision, and basic material constraints to the initial point cloud. Through iterative solving, it reached static equilibrium and obtained a physically corrected scene. These comparison methods provide a benchmark for comparison from three dimensions: unconditional generation, conditional generation, and physical processing, as shown in Table 2.

TABLE 2. Overview of Core Characteristics of Comparative Methods

Method name	Type	Conditional control mode	Physical rule processing method
r-GAN	Unconditional generation	None	None
TreeGAN	Unconditional generation	None	None
SP-GAN	Conditional generation	Semantic layout	None
SP-GAN + PBD(Two-stage)	Conditional generation + post-processing	Point cloud hidden vectors	Offline posterior simulation and correction
MC-PAG (Proposed)	Conditional generation	Multimodal (geometric, semantic, physical)	End-to-end online constraints

All experiments were conducted in a unified computing environment, and the models were implemented based on the PyTorch framework [30]. For the MC-PAG framework, Adam is selected as the optimizer, with an initial learning rate of 0.0001 and a batch size of 8. The physical constraint weights in multi-scale discriminant loss are set to 0.5 through grid search. The final joint training objective for the encoder and generator is a weighted sum of the following components: Adversarial Loss: Weight = 1.0 Disentanglement Loss: Weight = 0.1 Reconstruction Loss: Weight = 10.0 Per-level Physical Constraint Loss: Weight = 2.0 These hyperparameters were determined via grid search on the validation set to balance the convergence and performance of each term. The level L of the cascaded generator is set to 4, and the upsampling multiples for each level are 4, 4, 2, and 2, respectively. Comparative methods are trained on the same training set until convergence, and the same validation set is used for early stopping and hyperparameter tuning to ensure fairness in comparison. When evaluating, 50 sets of multimodal condition inputs are randomly selected from the test set, and corresponding scenarios are generated for each trained model. All generated results are subjected to unified quantitative evaluation and visual analysis.

IV. RESULTS AND ANALYSIS

A. GEOMETRIC FIDELITY ANALYSIS OF GENERATED SCENES

To quantitatively evaluate the restoration quality of the generated scene in three-dimensional geometric form, this section conducted a systematic geometric fidelity analysis of the outputs of MC-PAG and various baseline methods. The experiment calculates four core indicators, namely chamfer distance (CD), volume intersection ratio, Hausdorff distance, and normal vector consistency, and compares them comprehensively from three dimensions: overall error, substructure error, and profile geometry, to verify the advantages of the method in capturing complex terrain and fine ice and snow structures. In the comparison of structural errors, this section further subdivides and analyzes key components, summarizing them into four representative dimensions: ice and snow cultural structures, coastline natural terrain, conventional buildings, and overall scenes. The analysis results are shown in Figure 2. The comprehensive analysis in Figure 2 reveals the overall superiority of the MC-PAG method in geometric fidelity. Regarding the overall error (Figure 2a), MC-PAG's chamfer distance is 1.37×10^{-3}, a reduction of 18.5% compared to the second-best SP-GAN+PBD, and its Hausdorff distance is 4.12×10^{-2}, a reduction of nearly 30%. This is attributed to MC-PAG's cascaded generation structure and physical constraint mechanism effectively suppressing outliers and overall deformation in the generated point cloud, making the generated surface more closely conform to the real geometry. Regarding structural similarity (Figure 2b), MC-PAG achieves a volume IoU of 0.842 and a normal vector consistency of 0.923, both the highest values. The embedding of physical constraints makes the normal field of the generated point cloud more consistent with the continuous variation of the real surface, thereby improving the rationality of local geometric details. Figure 2c further shows that MC-PAG has a CD (Chamfer Distance) error of 1.52×10^{-3} in the most challenging category of "ice and snow cultural structures", which is significantly lower than other methods. This is due to the explicit separation and precise control of cultural semantic features by the multimodal encoder, which enables the model to focus on the geometric reconstruction of complex artificial structures. The profile line comparison in Figure 2d intuitively shows that the elevation profile generated by MC-PAG matches the real terrain undulations best, while the profile lines of other methods such as r-GAN show obvious high-frequency jitter and systematic shifts, reflecting that unconditional generative models are difficult to stably reconstruct the terrain features of specific scenes without the guidance of multimodal conditions. In summary, MC-PAG has achieved high fidelity geometric reconstruction of cold-region waterfront scenes, especially complex ice and snow cultural elements, through its multimodal condition encoding and physical constraint generation mechanism.

B. CULTURAL SEMANTIC CONSISTENCY IN GENERATING SCENARIOS

To quantitatively evaluate the accuracy of the generated scene at the semantic level, the experiment performed fine semantic segmentation on the scene point cloud and calculated the intersection to union ratio on key categories. This analysis aims to verify the learning and control ability of multimodal conditional coding on cultural semantic features, and measure whether the layout and composition of the generated scene conform to the semantic logic of the real world. The quantitative comparison results of IoU for each category are shown in Figure 3. Figure 3 presents the IoU accuracy of r-GAN, SP-GAN, and MC-PAG across five core semantic categories. The five axes correspond to the categories of ice and snow structures, natural water bodies, ice surfaces, snow cover, and vegetation structures. MC-PAG exhibits the highest IoU value across all categories, with its polygonal outline closest to the outer edge. In the most challenging category, "ice and snow structures," MC-PAG achieves an IoU of 0.845, significantly higher than SP-GAN's 0.685 and r-GAN's 0.512. This advantage is attributed to the explicitly decoupled cultural semantic latent space within the MC-PAG framework. This design allows the model to separate and precisely control the generation of cultural elements from the input conditions, avoiding the confusion and omissions in complex cultural features inherent in general generative models. In the categories of "natural water bodies" and "ice surfaces," MC-PAG's IoU is 0.892 and 0.868, respectively. Its physical constraint module reduces the abnormal generation of water body and ice surface geometry by enforcing waterfront boundaries and phase transition laws, thereby improving the accuracy of semantic segmentation. Quantitative results demonstrate that

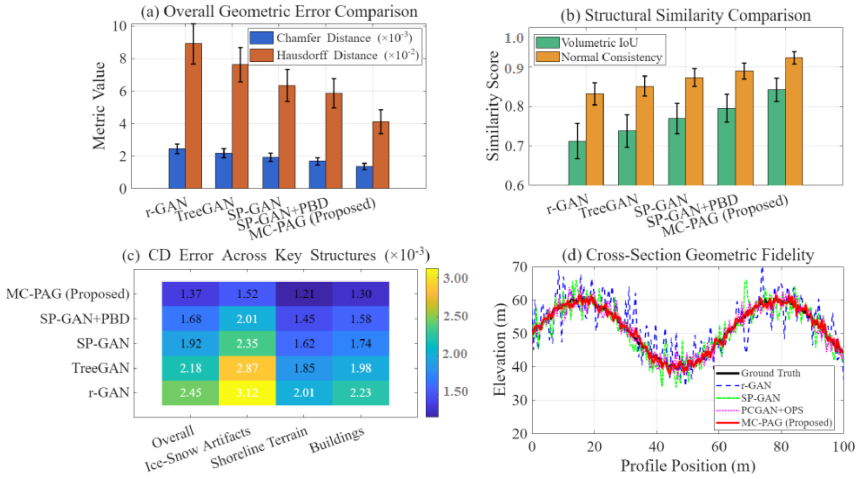

FIGURE 2. Multi-dimensional comparative analysis of geometric fidelity of generated scenes: (a) Overall geometric error comparison; (b) Structural similarity comparison; (c) Sub-structure CD error analysis; (d) Profile geometric fidelity comparison.

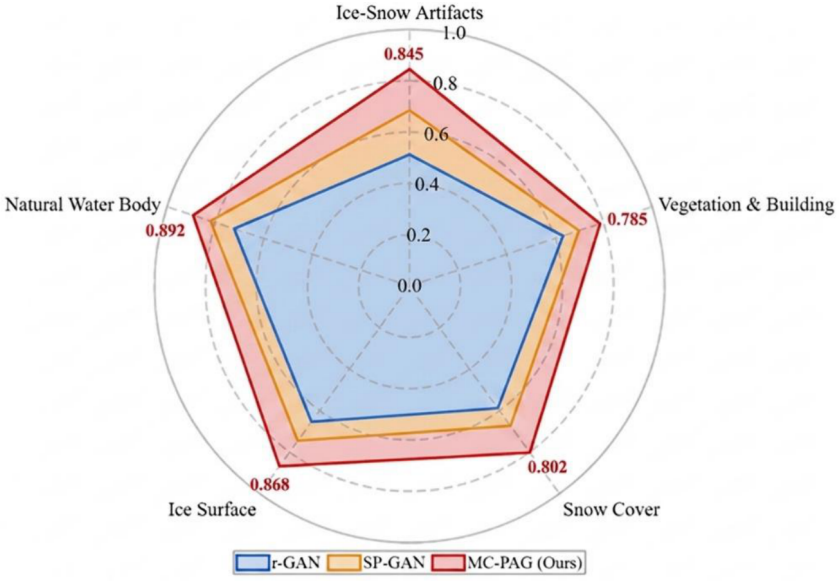

FIGURE 3. Semantic Segmentation Accuracy Comparison

MC-PAG, by integrating semantic conditions and physical constraints, achieves high-fidelity and high-consistency semantic generation for cold-region waterfront scenes, especially those containing cultural semantic elements.

C. PHYSICAL RULE COMPLIANCE IN GENERATING SCENARIOS

To verify the compliance of the generated scenario with the geological and physical laws of cold-region waterfront environments, this section evaluates the physical rationality of the scenario from three dimensions: static rule violation, quantification of key interfaces, and dynamic condition response. The experiment calculated the violation rate of core physical rules such as gravity stability and waterfront coherence, quantified indicators such as elevation gradient and snow slope at the waterfront boundary, and tested the response characteristics of the generated results to continuous changes in temperature conditions. The multidimensional evaluation results are shown in Figure 4. Figure 4 illustrates the advantages of MC-PAG in terms of physical rationality. The data in Figure 4 (a) shows that the average physical rule violation rate of MC-PAG is the lowest, at 3.37%, and only 2.8% violates the "gravity and stacking stability" rule. The embedded microphysics simulator guides the generation process in a forward constraint manner, directly avoiding a large number of geometric errors that violate fundamental mechanical laws. Figure 4 (b) shows that the polygon formed by MC-PAG on the three normalized indicators of contact point elevation gradient, ice surface flatness, and snow slope is closest to the outer circle and most balanced. The normalized value of the generated contact point elevation gradient is the highest, which is due to the physical constraint adversarial training that forces the elevation continuity at the intersection of water, ice, and shore, significantly reducing unreasonable steep slopes or suspensions. Figure 4 (c) shows that when the temperature

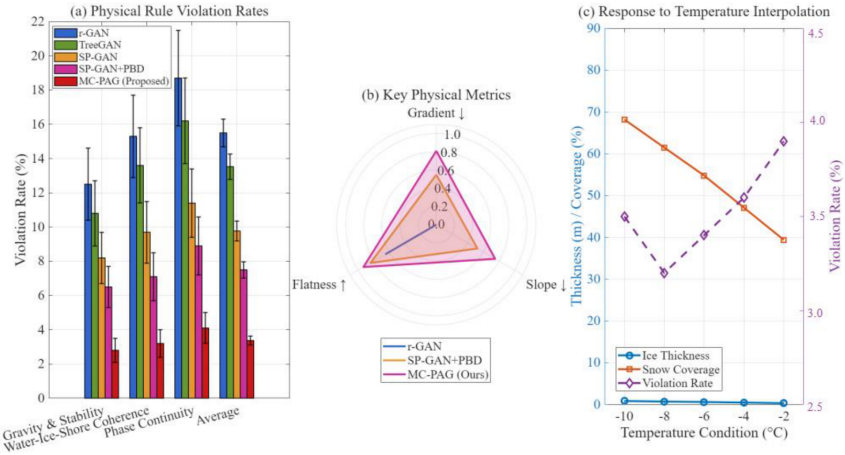

FIGURE 4. Multi-dimensional evaluation of physical rationality of generated scenes: (a) Comparison of violation rates of physical rules; (b) Key physical indicators; (c) Interpolation response analysis of temperature conditions

conditions are interpolated from -10 °C to -2 °C, the ice thickness generated by MC-PAG smoothly decreases from 0.85m to 0.33m, and the snow coverage rate synchronously decreases from 68.2% to 39.4%, while the physical violation rate remains stable between 3% and 4%. This interpolation simulates a gradual seasonal warming trend over weeks (e.g., from mid-winter to late winter/early spring), where sustained sub-freezing but rising temperatures can lead to sublimation, compaction, and reduced snow accumulation, thereby decreasing the apparent ice thickness (snow-ice mixture) and coverage. This proves that the decoupled physical latent space of the model can learn and extrapolate the continuous mapping relationship between ice and snow phase transition and temperature, ensuring the rationality of condition generation. Comprehensive analysis shows that MC-PAG generates scenes with high physical credibility through an end-to-end physical constraint embedding mechanism, and can make reasonable responses to dynamic environmental parameters.

D. DESIGN RATIONALITY VERIFICATION OF GENERATED SCENARIOS: FUNCTIONALITY AND SPATIAL EXPERIENCE

To verify the spatial functionality of generating scenes that support actual cultural activities, this section conducted functional accessibility and spatial configuration analysis from the perspectives of movement and perception in the scene. The experiment evaluated traffic efficiency by simulating path planning in the generated scene, and quantitatively analyzed the spatial organization clarity and visual guidance of the scene using spatial syntax theory. The comparative results are presented in Figure 5. Figure 5 comprehensively evaluates the practicality of generating scenarios as design solutions. Figure 5 (a) visualizes a schematic diagram of a waterfront ice and snow plaza generated by MC-PAG, where the red path shows the efficient connection from the main entrance through the core observation point to the service area, and the blue sector reveals the good visual control range between key nodes. This clear spatial structure originates from the precise analysis of design semantics such as "square" and "streamline" by multimodal conditional coding, enabling the generator to layout functional zones that conform to cognitive habits. The quantitative data in Figure 5 (b) confirms this advantage: the average path curvature of MC-PAG generated scenes is 0.12, much lower than the 0.22 of SP-GAN; Its visual integration is 0.82, which is also higher than the latter's 0.67. The lower curvature of the path is an objective metric reflecting that the physical constraint module eliminates unrealistic obstacles and extreme terrain, resulting in a navigable and coherent circulation baseline. This provides a rational foundation for subsequent detailed design; The higher visual integration is attributed to the model implicitly learning excellent layout patterns with strong spatial permeability through adversarial training. This analysis is grounded in Space Syntax theory, which links spatial configuration metrics like visual integration to empirically observed patterns of human movement and co-presence, thereby providing a theoretical basis for evaluating spatial experience. Meanwhile, the success rate of path planning in the MC-PAG scenario is as high as 98.5%, indicating that its spatial connectivity is more complete. Quantitative and qualitative analysis jointly demonstrate that the scenarios generated by MC-PAG are superior to other baseline methods in supporting efficient traffic and providing clear spatial perception, and have the potential to directly serve subsequent deepening designs.

V. CONCLUSION

This article proposes a hybrid framework based on multimodal conditional latent space decoupling and physical constraint adversarial generation to solve the intelligent design problems of cold-region waterfront ice and snow cultural scenes. This method explicitly decouples the geometric, semantic, and physical attributes of the scene through an encoder, and embeds differentiable physical simulation constraints in the synthesis process using a cascade generator, supplemented by a multi-scale physical perception discriminator for adversarial optimization, achieving end-to-end gen-

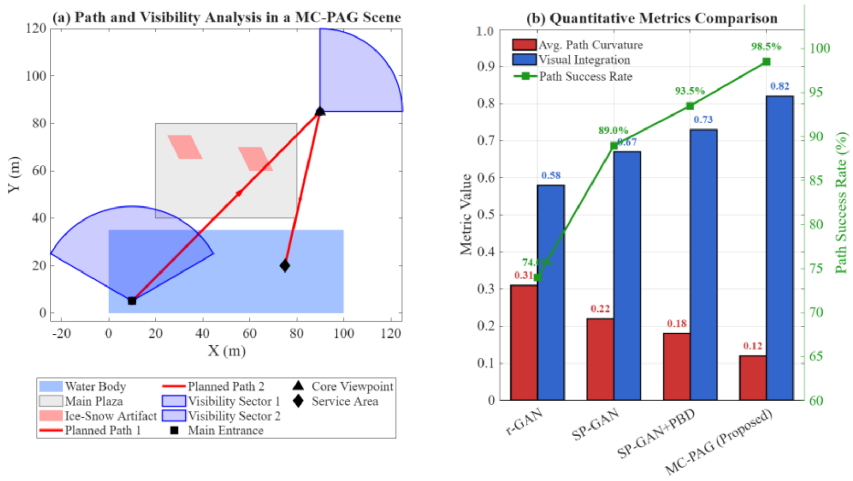

FIGURE 5. Comparative analysis of functional accessibility and spatial configuration: (a) Typical scene path and line of sight analysis; (b) Multi-method quantitative index comparison

eration of high fidelity and high rationality 3D point cloud scenes from multi-source conditions. The experiment shows that the framework performs well in key indicators: the IoU of the ice and snow culture construction object semantic segmentation generated by the scene reaches 0.845, the average violation rate of physical rules is as low as 3.37%, and the average curvature of the functional accessibility path is as low as 0.12. These data validate the significant advantages of this method in generating cultural semantic accuracy, physical rationality, and spatial functionality of the scene. This study provides a controllable and verifiable generative design paradigm for the seasonal planning and digital cultural innovation of cold-region waterfront spaces, integrating material intelligence and fiber-based structural systems to achieve a leap from "scene generation" to "design support." By embedding the physical constraints of technical into the generative process, the framework ensures that digital cultural reconstructions are functionally grounded in the material realities of architecture for extreme climates.

AUTHOR CONTRIBUTIONS

Yue Ge designed, collected and analyzed the data, and drafted the manuscript. Yue Ge conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Yue Ge participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDING

This research was supported by, 1. Basic Research Funds for Heilongjiang Provincial Universities 2024 Project "Research on the Implantation of Ice and Snow Culture in Urban Waterfront Landscape Belt Design in Heilongjiang Province" (Project No. 145409104). 2. The 2025 Annual Project of Heilongjiang Province Art and Science Planning Project "Research on Digital Filing and Active Inheritance of Traditional Costumes of Ewenki, Oroqen, and Hezhe Ethnic Groups in Cross-border China Russia Relations" (Project No. 2025B015). 3. The 2025 Project of Heilongjiang Provincial Higher Education Association (Project No. 24GJZXC096).

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] A. A. Fime, S. Mahmud, A. Das, M. S. Islam, and J.-H. Kim, "Automatic Scene Generation: State-of-the-Art Techniques, Models, Datasets, Challenges, and Future Prospects," IEEE Access, vol. 13, no. 1, pp. 95753-95796, 2025, doi: 10.1109/ACCESS.2025.3574298.

[2] Z. X. Chew, J. Y. Wong, Y. H. Tang, C. C. Yip, and T. Maul, "Generative design in the built environment," Automation in Construction, vol. 166, no. 1, Art. no. 105638, 2024, doi: 10.1016/j.autcon.2024.105638.

[3] X. Zhuang, Y. Ju, A. Yang, and L. Caldas, "Synthesis and generation for 3D architecture volume with generative modeling," International Journal of Architectural Computing, vol. 21, no. 2, pp. 297-314, 2023, doi: 10.1177/14780771231168233.

[4] D. Yang, J. Wang, and X. Yang, "3D Point Cloud Shape Generation with Collaborative Learning of Generative Adversarial Network and Auto-Encoder," Remote Sensing, vol. 16, no. 10, pp. 1772-1777, 2024, doi: 10.3390/rs16101772.

[5] X. Han, C. Liu, Y. Zhou, K. Tan, Z. Dong, and B. Yang, "WHU-Urban3D: An urban scene LiDAR point cloud dataset for semantic instance segmentation," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 209, no. 1, pp. 500-513, 2024, doi: 10.1016/j.isprsjprs.2024.02.007.

[6] H. Yoo, Y. Kim, J. H. Ryu, S. Lee, and J. H. Lee, "Semi-Automated Building Dataset Creation for 3D Semantic Segmentation of Point Clouds," Electronics, vol. 14, no. 1, pp. 108-114, 2024, doi: 10.3390/electronics14010108.

[7] K. A. Tychola, E. Vrochidou, and G. A. Papakostas, "Deep learning based computer vision under the prism of 3D point clouds: A systematic review," The Visual Computer, vol. 40, no. 11, pp. 8287-8329, 2024, doi: 10.1007/s00371-023-03237-7.

[8] Q. C. Xu, T. J. Mu, and Y. L. Yang, "A survey of deep learning-based 3D shape generation," Computational Visual Media, vol. 9, no. 3, pp. 407-442, 2023, doi: 10.1007/s41095-022-0321-5.

[9] Z. Qiu, J. Liu, Y. Xia, H. Qi, and P. Liu, "Text Semantics to Controllable Design: A Residential Layout Generation Method Based on Stable Diffusion Model," Developments in the Built Environment, vol. 23, no. 1, Art. no. 100691, 2025, doi: 10.1016/j.dibe.2025.100691.

[10] M. Shi, J. O. Seo, S. H. Cha, B. Xiao, and H. L. Chi, "Generative AI-powered architectural exterior conceptual design based on the design intent," Journal of Computational Design and Engineering, vol. 11, no. 5, pp. 125-142, 2024, doi: 10.1093/jcde/qwae077.
[11] R. Maskeliūnas, S. Maqsood, M. Vaškevičius, and J. Gelšvartas, "Fusing LiDAR and photogrammetry for accurate 3D data: A hybrid approach," Remote Sensing, vol. 17, no. 3, pp. 1-27, 2025, doi: 10.3390/rs17030443.
[12] J. Xu, S. Zhang, H. Jing, C. Hancock, P. Qiao, N. Shen, et al., "Improving real-scene 3D model quality of unmanned aerial vehicle oblique-photogrammetry with a ground camera," Remote Sensing, vol. 16, no. 21, pp. 3933, 2024, doi: 10.3390/rs16213933.
[13] Y. Shen, H. Xu, Y. Bao, Z. Xu, Y. Gu, J. Lu, et al., "Style-Based Tree GAN for Point Cloud Generator," IEEE Access, vol. 12, no. 1, pp. 27017-27028, 2024, doi: 10.1109/ACCESS.2024.3365519.
[14] W. Li, Y. Chen, Q. Fan, M. Yang, B. Guo, and Z. Yu, "I-PAttnGAN: An Image-Assisted Point Cloud Generation Method Based on Attention Generative Adversarial Network," Remote Sensing, vol. 17, no. 1, pp. 153-157, 2025, doi: 10.3390/rs17010153.
[15] W. Zhong and H. Meidani, "Physics-informed geometry-aware neural operator," Computer Methods in Applied Mechanics and Engineering, vol. 434, no. 1, Art. no. 117540, 2025, doi: 10.1016/j.cma.2024.117540.
[16] X. Hu, Q. Ma, P. Zhao, and X. Wang, "Physics-aware neural operator for high-fidelity fluid dynamics modeling with geometric and spectral priors," Physics of Fluids, vol. 37, no. 11, Art. no. 115111, 2025, doi: 10.1063/5.0299765.
[17] Y. H. Shin, K. W. Son, and D. C. Lee, "Semantic segmentation and building extraction from airborne LiDAR data with multiple return using PointNet++," Applied Sciences, vol. 12, no. 4, pp. 1975-1982, 2022, doi: 10.3390/app12041975.
[18] X. Nong, W. Bai, and G. Liu, "Airborne LiDAR point cloud classification using PointNet++ network with full neighborhood features," Plos One, vol. 18, no. 2, Art. no. e0280346, 2023, doi: 10.1371/journal.pone.0280346.
[19] S. Molnar and L. Tamas, "Variational autoencoders for 3D data processing," Artificial Intelligence Review, vol. 57, no. 2, pp. 42-48, 2024, doi: 10.1007/s10462-023-10687-x.
[20] J. Dai, Q. Guo, G. Wang, X. Liu, and Z. Zheng, "An optimized method for variational autoencoders based on Gaussian cloud model," Information Sciences, vol. 645, no. 1, Art. no. 119358, 2023, doi: 10.1016/j.ins.2023.119358.
[21] N. Balakrishnan, F. Buono, C. Calì, and M. Longobardi, "Dispersion indices based on Kerridge inaccuracy measure and Kullback-Leibler divergence," Communications in Statistics-Theory and Methods, vol. 53, no. 15, pp. 5574-5592, 2024, doi: 10.1080/03610926.2023.2222926.
[22] Z. Lu, B. Cao, and Q. Hu, "LiDAR-Camera Continuous Fusion in Voxelized Grid for Semantic Scene Completion," IEEE Transactions on Circuits and Systems for Video Technology, vol. 34, no. 12, pp. 12330-12344, 2024, doi: 10.1109/TCSVT.2024.3435045.
[23] S. Ye, X. Di, M. Liao, and X. Li, "Enhancing realism in LiDAR scene generation with CSPA-DFN and linear crossattention via Diffusion Transformer model," Neural Networks, vol. 189, no. 1, Art. no. 107503, 2025, doi: 10.1016/j.neunet.2025.107503.
[24] Q. Xu, X. Guan, J. Cao, Y. Ma, and H. Wu, "MPR-GAN: A novel neural rendering framework for MLS point cloud with deep generative learning," IEEE Transactions on Geoscience and Remote Sensing, vol. 60, no. 1, pp. 1-16, 2022, doi: 10.1109/TGRS.2022.3212389.
[25] Y. Jia, W. Yu, and L. Zhao, "Generative adversarial networks with texture recovery and physical constraints for remote sensing image dehazing," Scientific Reports, vol. 14, no. 1, Art. no. 31426, 2024, doi: 10.1038/s41598-024-83088-x.
[26] H. Lee, S. G. Lee, J. Park, and J. Shim, "Improving complex scene generation by enhancing multi-scale representations of gan discriminators," IEEE Access, vol. 11, no. 1, pp. 43067-43079, 2023, doi: 10.1109/ACCESS.2023.3270561.
[27] X. Zeng, A. Vahdat, F. Williams, Z. Gojcic, O. Litany, S. Fidler, et al., "Lion: Latent point diffusion models for 3d shape generation," in Koyejo S, Mohamed S, Agarwal A, Belgrave D, Cho K, Oh A, editors. Proceedings of the 36th International Conference on Neural Information Processing Systems, 28 November - 9 December 2022, New Orleans, LA, USA. Red Hook, NY, USA: Curran Associates Inc., 2022, pp. 10021-10039, doi: 10.48550/arXiv.2210.06978.
[28] Y. Li and G. Baciu, "SG-GAN: Adversarial self-attention GCN for point cloud topological parts generation," IEEE Transactions on Visualization and Computer Graphics, vol. 28, no. 10, pp. 3499-3512, 2021, doi: 10.1109/TVCG.2021.3069195.
[29] R. Li, X. Li, K. H. Hui, and C. W. Fu, "SP-GAN: Sphere-guided 3D shape generation and manipulation," ACM Transactions on Graphics (TOG), vol. 40, no. 4, pp. 1-12, 2021, doi: 10.1145/3450626.3459766.
[30] M. M. Yapıcı and N. Topaloğlu, "Performance comparison of deep learning frameworks," Computers and Informatics, vol. 1, no. 1, pp. 1-11, 2021, [Online]. Available: https://izlik.org/JA35GC87FU.