Title	直感的制御に基づく生成AIによる3次元モデリングとその応用：建築設計から都市計画への展開
Author(s)	杜, 旭升
Citation	
Issue Date	2026-03
Type	Thesis or Dissertation
Text version	ETD
URL	https://hdl.handle.net/10119/20585
Rights	
Description	Supervisor: 謝 浩然, 先端科学技術研究科, 博士

氏　　　　　名	DU Xusheng
学 位 の 種 類	博士（情報科学）
学 位 記 番 号	博情第 572 号
学位授与年月日	令和 8 年 3 月 25 日
論 文 題 目	Generative AI-Driven 3D Modeling and Applications using Intuitive Control: From Architectural Design to Urban Planning
論文審査委員	謝　浩然　　北陸先端科学技術大学院大学　　准教授
	長谷川　忍　　　　同　　　　教授
	岡田　将吾　　　　同　　　　教授
	井之上　直也　　　同　　　　准教授
	HU Ruizhen　Shenzhen University　Distinguished Professor

論文の内容の要旨

Global cities and architectures are undergoing profound structural and functional transformations. New buildings must meet higher performance standards under tighter budgets and shorter design cycles, while existing building stocks require large-scale adaptive reuse to avoid wasteful demolition and reduce environmental impact. At the urban scale, historical spatial data and multi-decade development trends are increasingly used to project future urban form, making long-term forecasting indispensable for land allocation and policy formulation. However, the digital tools that should support these tasks have not been sufficiently explored or developed to address these emerging demands. High-fidelity 3D building modeling remains labor-intensive; Level of Detail (LoD) architectural representations are fragmented and inconsistent; facade renovation workflows depend heavily on expert judgment; and city-level prediction tools struggle to integrate density, height, transportation networks, and historical evolution.

Despite the rapid progress of generative Artificial Intelligence (AI), particularly diffusion models and multimodal vision–language models (VLM), these technologies remain fundamentally misaligned with the requirements of the built environment. Most existing models are trained on natural images and therefore optimize primarily for visual plausibility rather than structurally meaningful representation. As a result, they struggle to capture part–whole relationships, spatial organization, and cross-view consistency, which are essential for architectural and urban modeling. Current generative models also lack mechanisms to maintain semantic and geometric continuity across LoD. Sketches, massing models, and detailed facades are treated as unrelated conditions rather than coordinated expressions of the same underlying structure, even though early-stage sketches play a critical role in representing design intent, boundary logic, and spatial hierarchy. Likewise, these models cannot connect sketch intent, component-level geometric logic, and long-term urban evolution into a unified cross-level reasoning process. Yet effective spatial design requires cross-level continuity between buildings and the cities they constitute, rather than handling them as isolated scales. In practice, existing generative AI methods can generate images that look like buildings, but they cannot yet interpret how buildings are composed, how their components relate, or why urban patterns develop over time. These limitations highlight the need for generative approaches grounded not only in visual appearance but also in structural relationships, hierarchical representation, and temporal dynamics, which are core attributes shaping how buildings are designed and how cities evolve. More fundamentally, these challenges stem from the absence of a unified cross-level modeling paradigm that can consistently connect object composition, building-scale organization, and city-scale evolution within a generative approach.

To address these challenges, this dissertation argues that generative modeling for the built environment must be formulated as a cross-level problem and accordingly proposes a unified approach that systematically integrates object-, building-, and city-level spatial representations within a computational paradigm.

- **Object-Level: Model Generation.** As the foundation of the proposed cross-level approach, this part introduces DualShape, a hybrid retrieval–generation framework that combines sketch-guided component retrieval with implicit SDF-based synthesis. By resolving ambiguity and incompleteness in freehand sketches, DualShape enables efficient, topology-consistent 3D component modeling, offering a new pathway for rapid geometric prototyping in early design.
- **Building-Level: Architectural Design and Renovation.** Building upon the object-level understanding of three-dimensional spatial structure and part–whole relationships, this part extends the approach to buildings as a specific class of objects and proposes methods for architectural design and renovation. It first constructs a multi-view, geometrically aligned Level-of-Detail (LoD) sketch dataset, addressing the long-standing lack of consistent LoD data in architectural research. In addition, a multi-view consistent 3D generation framework is developed, capable of reconstructing building models that remain geometrically aligned across different viewpoints. Furthermore, a facade renovation framework that integrates VLM-based structural reasoning, sketch-conditioned diffusion, and ControlNet refinement provides efficient and structure-aware design support for the adaptive reuse of aging industrial buildings.
- **City-Level: Urban Evolution.** In addition to object- and building-level modeling, the city scale represents a critical level of analysis for the built environment, where long-term spatial dynamics and collective urban patterns emerge. This part proposes MMCN, a Memory-aware Multi-Conditional generation Network that integrates building density, height, transportation networks, and historical urban patterns through multi-ControlNet, semantic fusion, and a spatial memory mechanism. Experiments on Shenzhen’s dataset from 2005–2024 demonstrate high predictive performance (SSIM 0.885, Boundary IoU 0.642) and robust generalization to Shanghai and Tianjin, enabling coherent cross-year and cross-city urban evolution forecasting.

Collectively, these three levels constitute a coherent cross-level generative foundation, within which geometric primitives, architectural semantics, and urban dynamics are modeled under a unified representational logic. This dissertation provides a new computational foundation for interpretable, controllable, and scalable generative design, supporting future applications in architectural practice, adaptive reuse, sustainable urban planning, and intelligent digital twins.

Keywords: Generative AI, Diffusion Models, Sketch-Based Modeling, 3D Generation, Level of Detail (LoD), Architectural Design, Facade Renovation, Urban Evolution, Cross-Level Generation.

論文審査の結果の要旨

近年，生成系人工知能技術（生成 AI）はマルチモーダル言語推論や画像生成において著しい進展を遂げているが，建築・都市分野への応用においては，構造的理解の欠如，幾何学的整合性の不足，および時間的連続性の欠落といった本質的な課題が残されていた．本論文は，生成 AI 技術を用いてこれらの建築デザイン及び都市計画の挑戦的な研究課題に向け，オブジェクト・建築・都市という 3 つの空間スケールにわたって統合的に扱う計算的方法論を提案し，その有効性を体系的に実証したものである．

具体的には，物体レベルにおいては，手描きスケッチを入力とした 3D 形状生成の手法を提案した．本手法は，暗黙的形状表現（生成モデル）と明示的な幾何学的分解（検索ベース）を組み合わせることにより，部品単位での意味的・構造的推論の技術を開発した．このパーツと全体の関係性に基づく表現手法は，建築や都市といったより大きな空間スケールにおいても拡張可能である．

また，建築レベルにおいては，物体レベルで得られた幾何学的理解を，建築物という設計対象へと拡張し，構造的整合性と表現の一貫性を高めるための 3 つの主要な貢献を行った．第一に，大規模なマルチ LoD（Level of Detail）スケッチデータセットを構築した．このデータセットは，詳細な建築形状から段階的に抽象化された階層的表現をモデルに学習させることを可能にし，建築形態の多段階的な抽象化能力の獲得に成功した．第二に，建築レンダリングを生成するための多視点生成フレームワークを提案した．多視点条件付き画像生成，深度情報に基づく幾何学的事前知識，および画像空間における整合性メカニズムを統合することで，複数の視点間におけるアライメントの維持を実現した．第三に，大規模視覚言語 VLM モデルに基づく建築改修ワークフローを開発した．本ワークフローは，意味的推論，パーツの生成，および画像の精緻化を組み合わせることで，初期設計段階の適応的再利用シナリオを支援する能力を示した．

最後に，都市レベルにおいては，建築レベルにおける制約付き生成を都市スケールの空間進化予測へと拡張した．本研究では， MMCN モデルを提案し，歴史的レイアウト，密度分布，高さパターン，交通ネットワークなどの異種空間要因を，多条件拡散過程の中で統一的に扱う枠組みを構築した．深圳市の多時期データセットに加えて，上海市および天津市のデータセットを用いた追加評価により，提案手法の汎化能力が実証された．

以上，本論文は，初期の建築デザインから長期的な都市計画に至るまで，生成 AI がより深く統合される技術フレームワークに関する包括的手法を提案した．建築情報学において生成的アプローチの発展に寄与するものであり，学術的に貢献するところが大きい．よって博士（情報科学）の学位論文として十分価値あるものと認めた．