Abstract

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. This part 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.