Intelligent Urban Design: Self-organized Block Form Generation Using Reinforcement Learning - An Empirical Study from Nanjing, China

Yuyue HUANG
Affiliation: School Of Architecture, Southeast University, Nanjing, China
Email: huangyy0628@163.com

Abstract
With the rapid rise of AI technologies, intelligent design has enhanced urban planning through algorithm-driven diversity, flexibility, and efficiency. This study proposes a reinforcement learning-based method for the self-organized generation of block forms, enabling adaptive spatial layout optimisation. Using Nanjing, China, as a case study, geometric calculations extract topological relationships to build a prototype database. A reinforcement learning model is then applied, incorporating design constraints like development intensity, building density, height, and green space ratio. Generative design experiments on three Nanjing blocks demonstrate how feedback allows dynamic adjustments, achieving economic, ecological, and social benefits. Unlike traditional rule-based approaches, this model learns through autonomous interaction with the environment and can better adapt to environmental characteristics, offering a novel framework and practical support for intelligent urban design.

Keywords: Block form; self-organized generation; reinforcement learning; intelligent optimisation; urban design

1. Introduction
With the acceleration of global urbanisation, urban design systems are facing the continuous impact of multiple dynamic constraints, such as population mobility and demand for mixed land use [1]. Traditional static planning methods (e.g., zoning planning or morphological guidelines) gradually reveal fundamental flaws in dealing with these real-time changes: First, the construction of the rule base is highly dependent on the experience of experts, and it is difficult to cover the special requirements of different cultural backgrounds and geographic conditions; Second, when faced with the multi-objective optimisation problem, the fixed weight allocation mechanism will easily lead to the solution falling into the local optimum, and often produces a single morphologically homogeneous solutions. This dilemma has given rise to the urgent demand for adaptive generative design methods, to dynamically adjust the spatial form through a real-time feedback mechanism instead of fixed rules [2,3].

Current intelligent design methods are mainly presented as two types of technological paths: multi-objective optimisation based on evolutionary algorithms and generative adversarial networks based on deep learning [4,5]. The former can generate Pareto frontier solution sets, but in practice, it is found that a single iteration is too time-consuming and the morphological diversity is limited; the latter can quickly generate visually reasonable solutions, but there are response delays and the decision-making process lacks transparency [6]. In essence, the existing methods have not yet broken through two key technical bottlenecks: (1) How to maintain the coherence of spatial topological relationships (e.g., the connectivity of the street network); (2) How to achieve real-time integration of dynamic constraints (e.g., dynamic adjustments of construction indicators). These shortcomings seriously restrict the practical application value of self-organised generation of neighbourhood morphology [7].

Reinforcement Learning (RL) provides new possibilities to solve the above problems. Its 'trial-and-error-feedback-optimisation' mechanism is intrinsically compatible with the process of self-organization of urban space: by simulating environmental interactions (e.g., crowd activity scenarios), intelligences can autonomously explore morphological evolution paths that satisfy multiple constraints [8]. The breakthrough research of Zheng et al. (2023) [9] indicates that RL can significantly enhance the working efficiency of human planners in architectural layout optimisation and efficiently generate design schemes of different styles. This lays a theoretical foundation for constructing an adaptive design system at the plot scale. Based on this, this study proposes a reinforcement learning framework that integrates spatial topological relations and dynamic constraints. Its technical implementation includes three core links: Firstly, the spatial topological characteristics of typical blocks in Nanjing are extracted through geometric operations to construct a design prototype database; Secondly, the constraint conditions such as development intensity and building density are quantified as reward functions; Ultimately, generative design was implemented in three districts in Nanjing, and the dynamic optimisation of form was achieved through the real-time feedback of the revenue function.

Compared with the traditional rule-driven methods, the innovation of this study is reflected in: (1) Mapping the theory of spatial self-organization systematically into the state-action space of reinforcement learning for the first time, which makes the morphogenesis process both environmentally adaptive and topologically stable; (2) Developing an interpretable reward mechanism, which achieves the comprehensive optimisation of economic, ecological and social benefits in the Nanjing case; (3) Constructing a design assistance link to support planner interaction, and the response time of the system was controlled within 30 minutes. The empirical results show that the method can not only meet the current design constraints, but also respond to future demand changes through continuous learning, providing a new technical paradigm for intelligent urban design.

2. Methods

2.1 Technical framework

The Dynamic Coupled Reinforcement Learning (DCRL) framework proposed in this study realizes the intelligent generation and optimisation of building layout schemes through the collaborative operation of four core modules (Fig. 1).

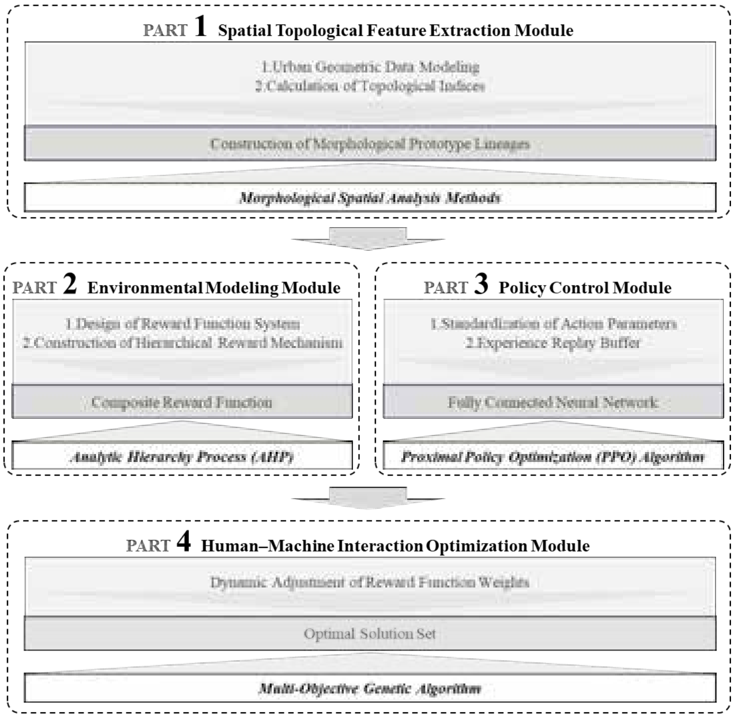

Fig.1. Technical framework

Firstly, the spatial topological feature extraction module conducts three-dimensional modeling of the geometric data of urban blocks, integrates spatial elements such as plot boundaries, street and lane networks, and the contours of existing buildings, and extracts topological indicators such as the curvature of streets and lanes, the fractal dimension of plots, and the proportion of building base area through computational geometric algorithms. Then, the Morphological Spatial Analysis (MSA) method is used for clustering to form a benchmark library containing multiple basic morphological prototypes [10], covering typical patterns such as “grid-type blocks - low-density development” and “radial road networks - mixed functions”, providing multi-dimensional morphological evolution benchmarks for reinforcement learning.

Secondly, the environmental modeling module converts the control elements of urban design into a calculable reward function system, builds a hierarchical reward mechanism for legal indicators such as floor area ratio and green space ratio, penalizes the indicators deviating from the target interval through the exponential attenuation function, and finally forms a comprehensive reward function through hierarchical weighting, where the weight coefficients are determined by the initial values of the AHP analytic hierarchy process.

Thirdly, the strategy control module adopts the improved Proximal Policy Optimisation (PPO) algorithm [11] to construct the state input layer containing multi-dimensional spatial features and multi-dimensional constraint indicators. Feature fusion is achieved through the fully connected neural network, and the output layer generates standardized action parameters using the Tanh function. Design a double-loop training mechanism: The inner loop synchronously updates the policy network and the value network, and reduces the variance through dominance estimation; For external environmental storage, the strategy transfer is carried out with the optimal 5% scheme of each generation to avoid local optima. In the specific implementation, a 10,000-capacity experience replay buffer is set up, the mini-batch update model is adopted, and the learning rate decays according to the cosine annealing strategy.

Finally, the human-computer interaction optimisation module builds a two-way interaction system. Designers can dynamically adjust the weight of the reward function (such as increasing the green space ratio weight to 0.4). The scheme optimisation module uses the multi-objective genetic algorithm for Pareto frontier optimisation. The iteration termination mechanism is set so that the comprehensive score change rate of the optimal scheme for 10 consecutive generations is less than 1.5% or triggered by manual intervention. Finally, output the solution set of the preferred scheme.

2.2 Case selection and feature extraction

The study selected three typical blocks in Nanjing as the experimental objects: The Gulou Historical Block (1.2 km²) retains the "fishbone-shaped" street and alley layout of the Republic of China era, contains three cultural relics protection units, and has a floor area ratio of 1.2-1.8, representing the protective development of the historical urban area; The CBD of Hexi Area (1.5 km²) features a modern grid-type road network with a plot ratio of 2.5 to 3.5, reflecting high-intensity business development. The Zhujianglu Area (0.8 km²) integrates multiple functions through a radial road network, with a plot ratio of 1.8 to 2.8, reflecting the mixed functional scenarios in urban renewal (Fig. 2).

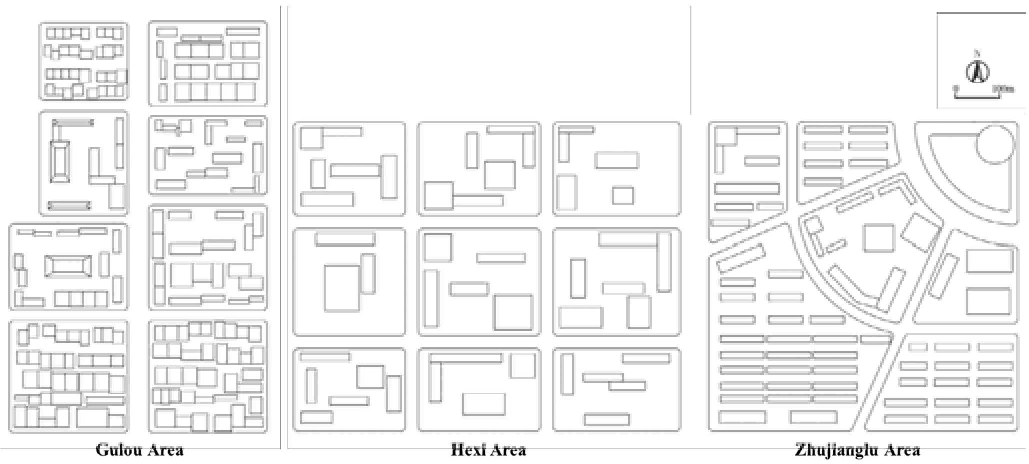

Fig.2. Three experimental areas

The feature extraction process is divided into two stages: Firstly, the ArcGIS platform is used to conduct geometric analysis of the basic spatial data of the study area, quantitatively extracting topological parameters such as street and alley connectivity, plot division patterns, and plot construction intensity. Secondly, K-means clustering analysis was conducted on the topological characteristics of buildings and plots in the main urban area of Nanjing, and they were classified into different prototype categories to construct a design prototype database. This database not only provides the initial state set for reinforcement learning, but also can generate composite topological structures through parameter combinations, effectively enhancing the diversity basis of form generation.

2.3 Reinforcement learning model construction

The state space design integrates three layers of feature vectors: static features, dynamic constraints, and spatial performance. The static layer includes the One-hot coding of the plot prototype and location features (standardized coordinate values, unique hot coding of road grades, and land use mixing degree index). The dynamic layer covers deviations of legal indicators, simulated values of environmental performance, and social indicators (such as walkability; The performance evaluation layer determines the weights of secondary indicators through the entropy weight method and linearly weights to generate the comprehensive score as the measurement of state value.

The action space defines four types of explainable form adjustment operations: ① Building volume reorganization (base area ±20%, height ±50%, step lengths 5% and 10% respectively), which needs to meet the requirements of fire fighting surfaces and aviation height limits; ② Interface offset (street-back line ±4m, step length 1m, continuous interface ±15%), it is necessary to ensure the width of the sidewalk and the continuity of the interface. ③ Public space replacement (minimum 100 square meters unit, conversion between green space and hard ground) must comply with the lower limit of the green space ratio and the proportion limit of public space in the plot. ④ Development intensity adjustment (FAR±0.2 for adjacent plots, not exceeding 20% of the original index), it is necessary to ensure that the indicators of each plot are compliant and the total construction volume change is ≤5%. Before each action is executed, a multi-constraint verification module that includes 3D spatial conflict detection, planning regulation inspection, and engineering feasibility assessment is adopted to ensure the implemtability of the plan.

3. Results

3.1 Spatial Topological Relationship Graph

The 180 plots of the three typical blocks in Nanjing were geometrically analyzed through the ArcGIS platform, and four types of spatial parameters including Road network, Land function, Development intensity and Building density were extracted. Through K-means clustering, a total of 54 spatial combination prototypes were formed (Fig. 3).

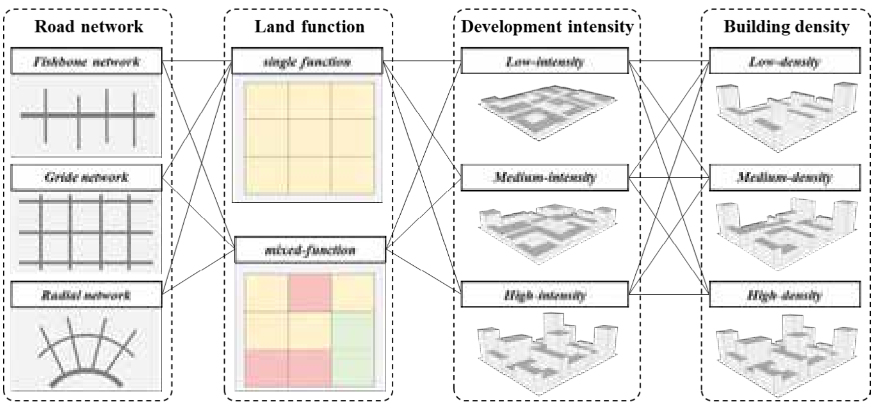

Fig.3. Space prototype combination mode

Among them, the Gulou Area is mainly based on the prototype of "fishbone-shaped streets and alleys + small-scale plots (average 3,000 square meters) + high building density (35%-45%)", and the integration degree of the streets and alleys (spatial syntax α=0.5) reaches 0.82, reflecting the high accessibility of the historical space. The Hexi Area is mainly based on the prototype of "grid-type road network + medium and high-intensity development (FAR=2.5±0.3) + low building density (20%-30%)", and the plot compactness index (0.75±0.05) reflects the regularized form of the modern business district. The Zhujianglu Area presents the characteristics of "radial streets and alleys + mixed functional layout + medium development intensity (FAR=2.2±0.3)", and the variation coefficient of the plot shape coefficient reaches 0.38, reflecting the morphological diversity brought by functional mixing. The prototype database contains various types of geometric parameter thresholds (such as the building aspect ratio 1:1.2-1:1.5) and planning constraints (such as the building height of the plot ≤150m), providing a multi-input space for the initial state vector for reinforcement learning.

3.2 Generation of the design plan for the Gulou Area

In response to the protection and renewal requirements of the Drum Tower Historical Block, the agent generated 32 effective solutions through 150 generations of iterations, and the Pareto optimal solution scheme was obtained through NSGA-II screening (Fig. 4). The optimal plan, while maintaining the integrity of the historical street and alley framework (92%), increases the green space ratio from 18% to 25%, and at the same time reduces the number of households with substandard sunlight exposure in cultural relics protection units from 12 in the initial plan to 0. The specific optimisation strategies include: ① Through the "building volume reorganization" action, reduce the base area of the three plots by 10%-15% to release the space for the construction of pocket parks at street corners; ② Utilize the "interface offset" operation to adjust the receded lines of the buildings along the street (with an average retracement of 2.5 meters), and widen the pedestrian passage to 4.5 meters without disrupting the continuity of the historical interface (maintaining a line alignment rate of 85%). ③ By adjusting the development intensity across plots (reducing the FAR from 1.8 to 1.6), the saved construction volume is transferred to non-protected plots to balance the protection requirements and development benefits. Compared with the traditional rule-driven method, the sunshine compliance rate of the DCRL scheme is increased by 23%, and it avoids the conflict between the green space ratio and building density caused by manually setting weights (the traditional scheme sacrifices an average of 5% of the green space ratio to meet the density requirements).

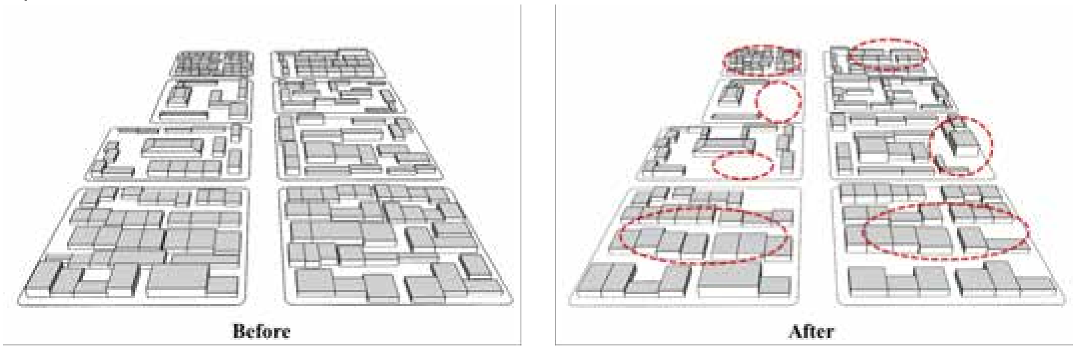

Fig.4. The design optimization plan of the Gulou Area

3.3 Generation of the design plan for the Hexi Area

In the high-intensity development scenario of the Hexi Area, the agent generated 56 compliant schemes after 200 generations of training. The optimal scheme achieved a comprehensive balance of a plot ratio of 3.0 and a building density of 25% (Fig. 5). The key optimisation paths include: ① Converting idle and inefficient land into green space in the center of the plot through "public space replacement", increasing the green space ratio of the block from 18% to 22%; ② Increase the floor area ratio to meet the demands of high-intensity development by raising the FAR of commercial plots from 1.5 to 4.0-7.0; ③ Add landmark buildings. Add landmark high-rise buildings to the plots close to the main roads to enhance the visual experience of the economic core area and economic corridors.

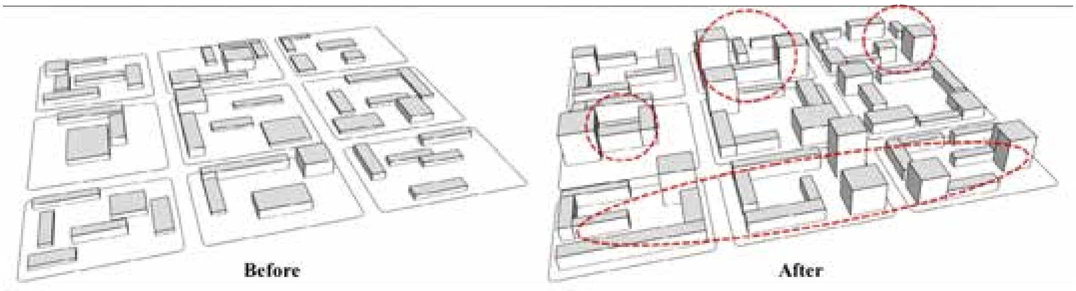

Fig.5. The design optimization plan of the Hexi Area

3.4 Generation of the design plan for the Zhujianglu Area
In response to the complex constraints of the Zhujianglu Area, the agent generated 45 effective schemes. The optimal scheme achieved a functional ratio of commercial building area accounting for 35%, residential building area accounting for 50%, and office building area accounting for 15% (Fig. 6). At the same time, the coverage rate of bus stops within a 500-meter walk reaches 100% (Figure 5). The core optimisation methods include: ① Through the "interface offset" operation, continuous arcade interfaces (200m in length) are set up in commercial plots to enhance the commercial atmosphere of the street, while meeting the requirement that the width of the sidewalk is ≥3m; ② Utilize "building volume reorganization" to adjust the building height of the residential plot (increase by 15%), expand the residential area without exceeding the height limit (50m), and maintain the uniformity of the street interface simultaneously through the control of the line connection rate (75%). ③ Through cross-plot development intensity adjustment (FAR±0.2), the excess 0.3 FAR of residential plots was transferred to commercial plots to achieve functional optimisation without changing the total construction volume. Compared with the traditional scheme, the scheme generated by DCRL has significant improvements in social indicators such as the functional mixing degree (the mixing degree index calculated by the entropy method is 0.89) and the efficiency of the walking network (the average shortest path is reduced by 120m).

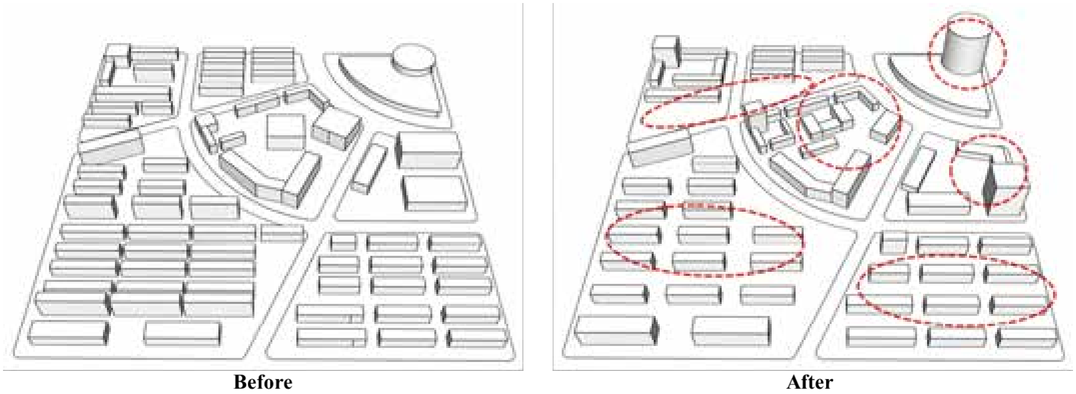

Fig.6. The design optimization plan of the Zhujianglu Area

Experimental data show that the optimal schemes of the three types of blocks all meet all the legal constraints (floor area ratio deviation ≤±0.1, green space ratio ≥ the legal lower limit), and on average increase by 18%-25% in environmental and social indicators compared with the initial schemes. The average training time was shortened by 15% compared with expectations, verifying the improvement effect of the elite solution set transfer strategy on computational efficiency. Visual analysis shows that the retention degrees of the morphological prototypes of each block generation scheme all exceed 70%, proving that DCRL has achieved innovative optimisation while maintaining the stability of spatial topology.

4. Discussion

4.1 Method Innovation

The core difference between the DCRL proposed in this study and the traditional generative design method lies in the construction of an intelligent decision-making mechanism of "dynamic feedback - continuous optimisation", breaking through the limitation of the traditional method that relies on "static rule base + heuristic search". Traditional generative design mostly generates schemes based on preset form rules (such as building regression lines, floor area ratio zoning, etc.), and its optimisation process is essentially parameter traversal under fixed constraints, making it difficult to respond in real time to complex environmental changes and multi-objective conflicts (such as the dynamic balance between economic benefits and ecological benefits). The DCRL converts urban design constraints into quantifiable dynamic reward signals through the environmental modeling module. Combined with the PPO algorithm of the policy optimisation module, it enables the agent to independently learn the optimal layout strategy in the interaction with the virtual environment, achieving a paradigm shift from "rule-driven" to "data-rule dual-driven". For example, when dealing with the contradiction between floor area ratio and green space ratio, traditional methods require manual setting of priority weights. However, DCRL, through the adaptive adjustment of the reward function, can dynamically search for the optimal solution that takes both into account in the Pareto front solution space, avoiding the local optimal problem caused by static rules.

The design of the reward mechanism plays a key coordinating role in multi-objective optimisation. The comprehensive reward function constructed in the research integrates the legal control indicators, environmental performance and social effects. Through hierarchical weighting (determining the initial weight by AHP) and dynamic adjustment (the human-computer interaction module supports real-time weight modification), the systematic quantification of multiple design goals is achieved. Compared with the linear weighting method commonly used in traditional multi-objective optimisation (with fixed weights and difficulty in reflecting the nonlinear relationship between objectives), the reward mechanism of DCRL has stronger dynamic adaptability: On the one hand, the implicit design requirements are transformed into calculable optimisation objectives through continuous reward signals (such as real-time simulation values of sunlight accessibility and pedestrian network convenience); On the other hand, combined with the generation and visualization of Pareto frontier solutions (the NSGA-II algorithm supports multi-scheme trade-offs), it provides designers with clear solutions to target conflicts, enabling the balance process of "economic benefits - ecological benefits - social benefits" to shift from empirical judgment to data-driven. This method of encoding urban design knowledge into computable reward signals provides a universal framework for the intelligent generation of complex built environments.

4.2 Practical Significance

The DCRL framework has shown significant application potential in the scenario of urban stock renewal. Stock renewal faces challenges such as complex spatial ownership, numerous constraints of existing facilities, and diverse demands of stakeholders. Traditional design methods are difficult to efficiently handle the dynamic coupling problem of multi-dimensional constraints. Take the Nanjing Gulou Historical Block as an example. Its spatial prototype of "fishbone-shaped streets and alleys + high-density plots" has strict protection requirements for building volume and interface continuity, and at the same time needs to meet the demands of modern traffic organization and public service improvement. The prototype database constructed by the spatial topological feature extraction module of DCRL can accurately identify existing morphological genes (such as small-scale plot boundaries and historical interface alignment rates), and combine the reward mechanism to quantitatively balance the protection elements (sunlight spacing of cultural relics protection units) and the renewal demands (new green space ratio). Generate a layout plan that not only meets the requirements of historical style protection but also enhances spatial performance. This workflow of "prototype recognition - constraint translation - intelligent optimisation" is particularly suitable for the refined design of complex existing Spaces such as old urban areas and industrial heritage areas, providing technical support for urban renewal to shift from "incremental expansion" to "existing quality improvement".

In terms of interface compatibility with the urban planning management system, DCRL has achieved seamless integration of intelligent design achievements with the existing planning control system through standardized data interaction mechanisms and constraint translation methods. Firstly, the spatial topological feature extraction module adopts the CityGML standard to construct a three-dimensional spatial information model, and forms data intercommunication with the ArcGIS platform commonly used in the planning management system to ensure the consistency of basic parameters such as plot division and construction intensity. Secondly, the environmental modeling module directly converts legal indicators such as floor area ratio and building density into the core variables of the reward function, enabling the schemes generated by the agent to naturally comply with the planning control requirements (such as the development intensity deviation controlled within ± 0.2FAR), reducing the frequent compliance verification costs in the generation of traditional schemes. Finally, the human-computer interaction module enables designers to adjust the constraint weights in real time (such as adjusting the priority of the green space ratio according to the control plan), enabling the optimisation process to respond synchronously to changes in planning management policies and forming a closed-loop workflow of “intelligent generation - manual guidance - policy adaptation”. This compatibility lays the foundation for the deep integration of the planning approval system and intelligent design tools, promoting the transformation of urban design from “experience-based decision-making” to “data-driven intelligent decision-making”.

4.3 Limitations and Future Prospects

Although the DCRL framework shows advantages in method innovation and practical application, its performance is still limited by the quality of training data and computational efficiency. Firstly, the construction of the spatial prototype library relies on the typical morphological features obtained by K-means clustering. When facing small sample scenarios (such as special functional blocks and extreme terrain conditions), the limited training data may lead to insufficient prototype coverage, thereby affecting the diversity and adaptability of the generation schemes. For example, if a certain block has a unique form of “valley landform + irregular road network”, and there is a lack of similar samples in the prototype library, the agent may not be able to effectively identify the key constraints, resulting in spatial performance deficiencies in the scheme. Future research can introduce transfer learning techniques. By using the pre-trained basic prototype model combined with a small amount of data in the target area for fine-tuning, the generalization ability of the model in small sample scenarios can be enhanced. Meanwhile, explore self-supervised learning methods to automatically discover potential morphological patterns and dynamically expand the coverage of the prototype library.

Secondly, the balance between the real-time computing efficiency and the design cycle of the reinforcement learning model urgently needs to be optimized. The PPO algorithm adopted in the current strategy optimisation module takes a long time in a single iteration when dealing with the high-dimensional state space and the continuous action space (a single case requires approximately 4 to 6 hours to complete 200 generations of training), and it is difficult to meet the requirements of rapid scheme comparison and selection in actual projects. Although the research has improved the training efficiency through the experience replay buffer and the elite solution set transfer mechanism, in the design of large-scale urban districts, the consumption of computing resources still increases exponentially. In the future, breakthroughs can be made in two aspects: algorithm optimisation and hardware acceleration. On the one hand, develop lightweight neural network architectures (such as the combination of deep reinforcement learning and graph neural networks) to reduce the computational complexity of state processing. On the other hand, the distributed computing platform is utilized to process the optimisation tasks of multiple plots in parallel. Combined with GPU acceleration technology, the time of a single iteration is shortened, promoting the framework to be implemented from the experimental environment to engineering application scenarios. Furthermore, exploring the two-tier architecture of “offline training - online rapid generation”, pre-training the universal layout strategy model, and conducting rapid fine-tuning only for site-specific constraints in specific projects can further balance computational efficiency and design flexibility.

References

Lei, W., Jiao, L., Xu, Z., & Zhu, X. (2024). Evolution of urban land and population system coupling micro–dynamics and macro-stability: Trends and paths. Land Use Policy, 141, 107152.

Valença, G., Moura, F., & de Sá, A. M. (2021). Main challenges and opportunities to dynamic road space allocation: From static to dynamic urban designs. Journal of urban mobility, 1, 100008.

Lee, J. (2022). A multi-scale perspective on production of space: A critical review of urban design. Cities, 121, 103497.

He, W., & Chen, M. (2024). Advancing urban life: a systematic review of emerging technologies and artificial intelligence in urban design and planning. Buildings, 14(3), 835.

Wang, H. (2024). Application of new features based on artificial intelligent robot technology in medium-scale urban design pedigree and intelligent management and control. Intelligent Systems with Applications, 22, 200379.

Gan, W., Zhao, Z., Wang, Y., Zou, Y., Zhou, S., & Wu, Z. (2024). UDGAN: A new urban design inspiration approach driven by using generative adversarial networks. Journal of Computational Design and Engineering, 11(1), 305-324.

Kakooee, R., & Dillenburger, B. (2024). Reimagining space layout design through deep reinforcement learning. Journal of Computational Design and Engineering, 11(3), 43-55.

Adibhesami, M. A., Karimi, H., & Sepehri, B. (2024). Optimizing urban design for pandemics using reinforcement learning and multi-objective optimization. In City Information Modelling (pp. 77-94). Singapore: Springer Nature Singapore.

Zheng, Y., Lin, Y., Zhao, L., Wu, T., Jin, D., & Li, Y. (2023). Spatial planning of urban communities via deep reinforcement learning. Nature Computational Science, 3(9), 748-762.

Scorza, F., & Fortunato, G. (2021). Cyclable cities: building feasible scenario through urban space morphology assessment. Journal of urban planning and development, 147(4), 05021039.

Wang, Y., Guo, J., Jiang, Y., & Sun, C. (2024). Multi-objective optimisation of buildings in urban scale for early stage planning and parametric design. Sustainable Cities and Society, 113, 105714.