A Deep Reinforcement Learning Model-Based Optimization Method for Graphic Design

Qi Guo*, Zhen Wang
School of Art and Design, Henan Institute of Economics and Trade; Zhengzhou, Henan 450000 China
E-mail: guoqi390446118@126.com
*Corresponding author

Keywords: Topology Optimization, Graphics design, buildings, Deep reinforcement learning, and 2D-3D structure.

Received: October 16, 2023

The significance of Deep Reinforcement learning is sensibly represented in the method of optimizing the graphic design and space framework of buildings in context with the worldwide big data environment, wherein people have increasingly stringent requirements for building layout and design and conventional layout is increasingly inadequate. This research put out a novel approach to topology optimization using deep learning in geometry. Deep neural networks characterize the density distribution in the design domain. By employing a geometry-based deep learning approach to represent the density distribution function, we can successfully avoid the checkerboard phenomena and ensure a smooth border. With a deep learning reinforcement approach, the design variables may be drastically decreased. In adjusting the designs of neural networks, we may fine-tune not only the minimal length but also the structural complexity. The proposed model has provided an accuracy of 95% and a computation time of 61s. The effectiveness of the suggested technique is shown by several 2-dimensional and 3-dimensional numerical results ranging from minimal conformance to stress-constrained issues.

Povzetek: Predlagana je nova metoda vzpodbujevalnega učenja za topološko optimizacijo v grafičnih storitvah z uporabo globokih nevronskih mrež.

1 Introduction

In both academia and business, research on machine learning (ML) and artificial intelligence (AI) has grown significantly in the past ten years. As computer technology improved and the need to evaluate increasing amounts of data evolved, these methods, which were previously undervalued, found updated recognition. Reinforcement Learning (RL) aims for maximizing a numerical reward signal by retraining the system to relate actions to instances. The student must attempt each activity to determine which is most rewarding rather than being instructed which to choose. The issue of how agents should learn a strategy that acts in a way to maximize the cumulative reward through interaction with the environment is addressed by reinforcement learning (Tapeh & Naser, 2022). Figure 1 represents Deep Reinforcement Learning Implementation using the Interior Design Model. The article outlines the solution of multi-objective reinforcement learning (MORL) tasks with unknown weights and many conflicting objectives (Yamaguchi, Nagahama, Ichikawa, & Takadama, 2019). The research demonstration continues to grow because it enables robots to quickly acquire innovative abilities. In inverse reinforcement learning (IRL), demonstrations can benefit in a number of methods by having the robot make an effort to determine the objectives or reward from the human demonstrator (Das, Bechtle, Davchev, Jayaraman, Rai, & Meier, 2021).

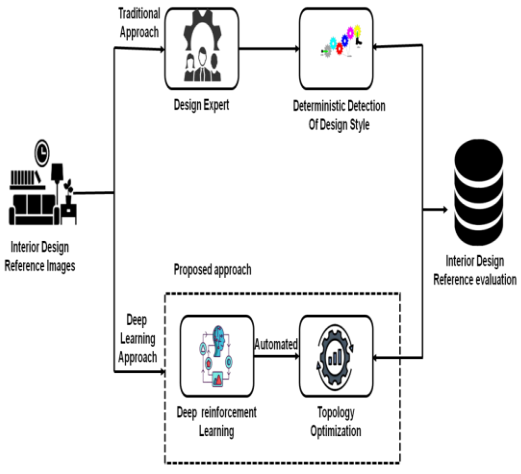

Figure 1: Deep reinforcement learning implementation using the interior design model.

The creations of completely autonomous agents that interact with their surroundings for learn the best behaviours and perfect them over time through trial and error. Making AI

systems that are responsive and can successfully learn has long been a problem, from software-only agents that can interact with spoken language and multimedia to robots that can perceive and respond to their environment (Zhou, Lee, Diao, Shi, Balyen, &Peto, et al, 2019). RL is a mathematical framework with guiding principles for experience-driven autonomous learning. While earlier iterations of RL had some success, they were fundamentally confined to rather low-dimensional issues and lacked scalability (Cioffi, Travaglioni, Piscitelli, Petrillo,& De Felice, et al, 2020).
AI will have a profound influence on human existence in the future due to the worldwide nature of the world, and it will be a key factor in designers' decision-making processes. Artificial intelligence is fundamentally a tool, and it should exercise its four main responsibilities of anticipation, contemplation, negotiation, and reaction throughout the process of design innovation (Bichu, Hansa, Bichu, Premjani, Flores-Mir, &Vaid, et al, 2021). Each designer has a preference, and ResNet artificial intelligence is suggested as a way to increase decision accuracy while also increasing the effectiveness of design selections based on individual designer preferences. To successfully prevent the negative consequences of designers' decision-making preferences, pattern recognition, and decision-making difficulties are combined (Wang, Tang, Huang, Chen, Zhang, & Huang, (2020)). The term "spatial layout design" describes the process of partitioning a given space into several tiny spaces or of logically placing certain things in the area within the framework of some objective and arbitrary design standards and layout conventions (Bouhamed, Ghazzai, Besbes, &Massoud, (2020)). The necessity for efficient design nowadays cannot be addressed by conventional approaches, which is why researchers are looking into spatial layout design. Designers often use interactive modeling tools or build traditional layouts by hand (Deng, & Chen, (2021)). The article suggested a novel Deep Reinforcement Learning-based topology optimization technique. The density dispersion in the design region is characterized by deep neural networks. Using geometric deep learning to define the density distribution function can ensure the smoothness of the border and successfully combat the checkerboard phenomena, in contrast to standard density-based methods (Brown, Garland, Fadel, & Li, et al, (2022)).

2 Related works

Table 1: Survey of related works

Author	Proposed	Result	Limitations
(Zhou, Lee, Diao, Shi, Balyen, &Peto, et al, (2019))	In the field of ophthalmology, AI, ML, and DL has been applied to verify medical diagnoses, interpret images, map the cornea, and compute intraocular lenses.	The existing DL, ML, and AI techniques and application on glaucoma treatment, AMD, DR, and other eye disorders early identification.	One of the main issues in many nations is the shortage of retina specialists and qualified human graders. Analysis of such images can be expensive, time-consuming, and prone to human error in population growth.
(Cioffi, Travaglioni, Piscitelli, Petrillo, & De Felice, et al, (2020))	The research has been designed to conduct a thorough analysis of scientific research about the industrial applications of AI and ML.	The significant outcome is the higher quantity of American-published works and the growing interest following the release of Industry 4.0.	It is essential to emphasize that this report was generated simply from two databases, namely WoS and Scopus and that only publicly accessible materials were included.

(Bichu, Hansa, Bichu, Premjani, Flores-Mir, &Vaid, et al, (2021))	The PRISMA-ScR standards were followed in the scoping assessment of the research.	The fields of diagnosis and treatment planning, development assessment, and treatment outcome evaluation were examined.	Some AI applications could have failed to appear in PubMed because of inclusion rules, search terms used, publication language rather than English.
(Wang, Tang, Huang, Chen, Zhang, & Huang, (2020))	The study developed the DNN framework and RL state area, action space, and multiple incentives.	The northeast power grid and the 36-node the China Electric Power Research establishment (CEPR) system are utilized to verify the efficacy of the technique.	The adjustment effect can be enhanced by raising the range of adjustments step per sample, with completing that could extend the learning and adjustments period.
(Bouhamed, Ghazzai, Besbes, &Massoud, (2020))	To enable the UAV to navigate over obstacles and the continuous area developed the Deep Deterministic Policy Gradient (DDPG).	The UAV is provided utilizing the DDPG in constant movement space to navigate over obstacles to achieve its designated destination.	The limited dimensions of mobility and action space for UAVs, which could lower their effectiveness in dealing with everyday environments.
(Deng, & Chen, (2021))	A policy-based RL model was developed in the investigation to depict the behaviour of controlling the thermostat and material level. To simulate the individuals' behaviour, a MDP used.	The behaviour of building occupants could be predicted reasonably well using the RL framework and transfer learning.	A limitation of the research was the RL occupant behaviour model's prediction difference, which could be partially justified.
(Brown, Garland, Fadel, & Li, et al, (2022))	An RL agent can sequentially determine in the specified environment whether to create a topology by eliminating components to most effectively accomplish compliance minimization requirements.	These results indicate that deep RL agents can acquire generalized design techniques to satisfy multi-objective design requirements.	Testing was done on the agent using a number of standard load instances, certain of which it did not see during training.

Contribution of the study Thus, this research contributes by demonstrating an implementation of the topology optimization to increase its effectiveness by Deep Reinforcement Learning and the field's relevance to making decisions through trial. The following are some of the particular accomplishments of this paper:

- The approach of interior design based on certain learning method is evaluated.
- To encourage the mathematical method of topology which is an optimized material layout within a given design space and assess the effectiveness of the process, an efficient Deep Reinforcement Learning component is suggested.

3 Application of deep learning in graphic design

The article presented an approach to electrical drive controller design that uses Deep Reinforcement Learning techniques. To effectively forecast the behavior of building occupants with high scalability and without the requirement for data gathering, the RL model was integrated with transfer learning (Ding, &Cerpa, (2020)). The article investigated how Designing 2D discredited topologies is automated by applying the optimum sequences of actions for RL agents to do to accomplish a goal learned from prior experiences. An RL agent may build a topology in the given environment by sequentially deciding which parts should be removed to best achieve compliance reduction goals (Zhang, Chen, Bernstein, Chintala, Graf, Jin, &Biagioni, et al, (2022)). The overview objective of the article conducted a series of using the lessons we've learned. By using a group of environment-conditioned neural networks, the piece was able to learn the dynamics of the building. Next, a brand-new control technique called Model Predictive Path Integral is used. In a five-zone office complex, we assess Energy Plus models. According to the report can save 8.23% more energy than the most advanced system while keeping a comparable level of thermal comfort (Zhang, Chintala, Bernstein, Graf, &Jin, et al, (2020).). The study desired a tailored scanning strategy that was learned using reinforcement learning (RL) to determine the angles and dosage for each selected angle for each patient. Modern deep RL techniques are used in the study to define the CT scanning procedure and then solve it. In addition to producing improved reconstruction outcomes, the learned tailored scanning technique also exhibits great generalizability when used in conjunction with other reconstruction algorithms (Shen, Wang, Yang, & Dong, et al, (2020)). The research downplayed the significance of sampling while determining the Q-return function, ensuring that the built-in techniques are more likely to acquire high-value lessons while being more resilient (Li, Zhu, Zhou, Feng, & Feng, et al, (2022)). Research enhanced the Building Information Model system and Python development tools, enabling cross-platform collaboration deep learning on computers and further design effort, The architectural design methodology of the BIM system and the interior design research carried out using the BIM building data platform were assessed in the article is shown using real-world examples (Luong, & Pham, 2021).The study paper's goal analyze the demand for interior space design has risen quickly along with the rate at which people are purchasing homes. In the domain of autonomous interior space design, computer science, and technology have infinite potential. The corresponding study suggested an automated way of designing spatial areas using convolutional neural networks (CNN) (Wu& Feng, 2022). The article investigated the CNN technique as a quick and effective approach. Iteratively finishing the automated arrangement of the internal spaces begins with the predicted living room. The paper examined several empirical interior design case studies, showing that this approach had similar results to professional designers' interior design floor plans (Predić, Manić, Saračević, Karabašević, &Stanujkić, 2022). Research classified the four different Machine Learning (ML) models created for the semi-arid region of Iraq's river flow forecasting. Investigated was the efficacy of data division's impact on the development of ML models. Three data division modeling scenarios—70%–30%, 80%–20%, and 90%–10%—were examined. To evaluate how well the models are performing, several statistical indicators are computed (Tao, Al-Sulttani, Salih Ameen, Ali, Al-Ansari, Salih, & Mostafa, 2020). Using 90%–10% data division, the article demonstrated the benefits of the hybrid support vector correlation model with a genetic algorithm over current machine learning forecasting models for monthly river flow predictions. Also, it was discovered to increase the accuracy of high-flow event predictions (Zhong, Zhang, Zhang, & Zhang, 2022). The study case developed the Support vector regression (SVR) model's internal parameters may be tuned by the optimizer, which results in a robust learning process. Compared to earlier developed hybrid models, the article has improved its ability to predict stochastic river flow behavior (Xu, Zhang, Liu, Nie, Su, Nie, & Zhang, 2019.) The Research compared the design of Adaptive Cruise Control (ACC) using Model Predictive Control (MPC) and Deep Reinforcement Learning (DRL) in car-following instances (Lin, McPhee, & Azad, 2019). The research explored the DRL approach as comparable to MPC with a large enough prediction horizon when modeling errors disappear and the training information range is occupied by the testing inputs (Zhu, Wang, Pu, Hu, Wang, &Ke, 2019). The study evaluated that DRL control performance declines when testing inputs are outside of the training data range, which is a sign that machine learning generalization is insufficient (Chen, Tong, Zheng,

Samuelson, &Norford, 2020). The study focused on constraint optimization and multi-objective optimization; the investigation provides an innovative perspective on the data age's design progress. After verifying the quality of the non-adaptive solution set, optimizing the converge, uniformity, and extensiveness, analyzing the experimental process, and drawing a multi-objective conclusion, it is determined that additional optimization related to the interior and spatial structure is necessary for artificial intelligence making decisions in the instance of the Library of Highly Cold Lands (Ran, & Dong,2022). Research provided layout boundary or layout space to automatically generate a layout plan. The scene redirection solution has successfully been tested, according to the findings. The study used a redirection algorithm's efficacy which is shown by comparison with the outcomes of uniform scaling (Wu, 2022). The study case simulated two reinforcement learning agents in a cooperative learning setting to discover the ideal 3D layout for the Markov decision process (MDP) formulation. The article examines the tests on a big dataset of actual interior layouts, which includes industrial designs created by qualified designers. The numerical findings suggested model produces layouts of superior quality when compared to the most recent model (Di, & Yu, 2021).

4 Materials and method

Graphic design has been around since the beginning of time. Books, periodicals, packaging, newspapers, banners, emblems, and many more things all benefit from graphic design in some way. Graphic design, topology optimization, our suggested deep reinforcement learning approach, and performance assessment of this graphic design are the primary topics covered in this chapter.

4.1 Graphic design

According to a widely held belief, visual design is the art and skill of giving various words and graphics an orderly, practical, and appealing framework. Both the act (verb) and the product (noun) of visual art are related concepts. A kind of "all design" employed in the creation of different platforms is traditional graphic design. The logical and practical aesthetics that developed in conventional graphic design over the years for media are the foundation for contemporary visual graphic design, which is today employed across multiple fields such as industrial layout, information architecture, message styling, and more. Table 2 displays the types of graphic designs.

Table 2: Types of graphic designs

S.no	Graphic designs types
(i)	Visual identification
(ii)	Promotion and marketing
(iii)	Interface for Users
(iv)	Newspaper
(v)	Packaging
(vi)	Movements
(vii)	Environmental
(viii)	Visual Compositions

Graphics has been known by many different names over the last two centuries, including artistic works, advertising material, digital marketing, graphics, and visuals. This demonstrates how the range of methods used to convey information has broadened beyond traditional visual arts. The 2D graphic arts include book arts, calligraphy, lithography, cinematography, printing, and typography. Applications, experience-based design, interaction methods, user-centered design, and websites are just some of the newer areas that graphic arts have expanded to include. The number of design-related discussions is growing at an astounding rate. There is training and schooling in graphic design all around the globe, at all levels. The figure depicts the graphic model of the building structure in DRL.

Figure 2: Graphic design of building in DRL

4.2 Topology optimization

Topology Optimisation as a construction tool is rarely implemented in the design of buildings. It is usually the result of a laborious procedure necessary to produce results that meet the standards of a designer. Yet, that difficulty shouldn't prevent some builders from trying out these instruments in building design. The density-based approach converts the substance distribution into a finite-element spatial configuration. By constructing discrete elements of varying densities, the finite element method is developed. Mesh is used to represent density spatially in the well-established SIMP method, yielding an optimized layout with spaced boundary conditions. So, it takes a lot of work in post-processing to make a smooth CAD model, and that might reduce the accuracy of the geometry near the border. As the mesh is employed to describe the organizational topology, the variety of design parameters is usually quite huge for 3D design, and many mature optimization strategies are not appropriate for large-scale problems. In this section, we describe a novel approach to density portrayal that resolves those particular issues by using a feed-forward neural network. A high-fidelity feed-forward neural network can be used to illustrate a complex shape, ensuring a smooth surface throughout. Thus, a deep feedforward network is a natural choice for representing the density field in the design domain. In Figure 3, we see a contrast of three feedforward neural networks, each having three hidden units and a unique set of neurons in each of those levels and Figure 4 displays the outcomes of the training.

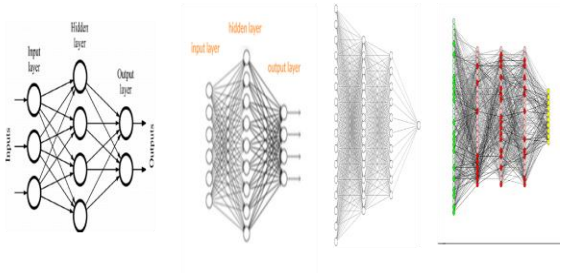

Figure 3: Feed-forward neural network design structure

Figure 4: Outcome of the building training function

A properly justified density field is one in which the limits of the component densities fall within the interval [0, 1]. A sampling distribution in the design domain is defined by a deep feedforward network in which the input for the system is all the point dimensions. The density value at that location is what you get as an answer. The following mapping function $\mathcal{N}$is used to control the output density to stay within the range [0, 1]:

$$\mathcal{N} = \frac{(\tanh(\beta y)+1)}{2}(\beta = 0.5) \quad (1)$$

The density field may be expressed mathematically as:

$$\emptyset(y,x) = \mathcal{M}(\mathbb{N}(y,x,\theta))(2Dproblem) \quad (2)$$

$$\emptyset(y,x,h) = \mathcal{M}(\mathbb{N}(y,x,h,\theta))(3Dproblem) \quad (3)$$

Where $\mathbb{N}$ represents feedforward networks and stands for a free-form parameter. Several discrete layers make up a deep-layered network's topology. Networks with Fhidden layers may be represented as, where $z^{(1)}$ represents the output of the corresponding hidden layer.

$$\mathbb{N}(y,x,h,\theta) =$$

$$\mathbb{N}(e^{(F+q)}(\quad ^{(ZF)}(e^{(F)}\left(\ldots . z^{(1)}\left(e^{(1)}(y,x,h)\right)\right) \quad (4)$$

Where the linear process $(\ )$ is written as,

$$e^{(f)}(y) = U^{(f)}y + p^{(f)} \quad (5)$$

4.3 Minimum compliance

Topology optimization using a compliance-minimizing formulation is developed with deep reinforcement learning (DRL). In the space of design, a DNN represents the density field. Hence, the TO will repeatedly update the network configuration in the design domain to improve the concentration field until the component arrangement provides optimal stiffness performance. During

optimization, the density field in the design domain is changed by adjusting the connection weights in a feedforward fashion. This allows us to formulate the optimization issue as:

$$\begin{cases} Find: \theta \\ Min: V(w,\Phi) = \frac{1}{2}\int_{\Omega} \varepsilon(w)^{D} T\big(\Phi(\theta)\big)\varepsilon(w) t\Omega \\ s.t: \left\{\frac{1}{(\Omega)}\int_{\Omega} \Phi(\theta) t\Omega - \mathrm{C}_{prescribe \le 0}\right. \end{cases} \tag{6}$$

Whereθthe feedforward is network parameters and V is the architectural compliance goal function. The relative densityΦ in the world of design is denoted by, where $\mathrm{C}_{prescribe}$ is the proportion of the volume that must conform to the design. The finite element framework uses the unknown velocity field( ), the pressure (ε), and the elastic matrix (T) to represent these quantities.

4.4 The lower limit of stress compliance

While optimizing for the least conformance with pressure limitation issue, mises pressure is always employed to gauge local stress and serve as a restriction on the search space. Yet, it is numerically costly to restrict local stress. To estimate the local stress limitation, a p-norm method is used here. Many updated strategies for precise local stress regulation have been put forward in recent years. To keep things simple, we use a tried-and-true technique to put a cap on the local stress created by von Mises. In this approach, the constraint is formulated using the p-norm measure PN. Thus, the issue presented in Section 3.2 may be restated *as:*

$$\begin{cases} Find: \theta \\ Min: V(w,\Phi) = \dfrac{1}{2}\displaystyle\int_{\Omega} \varepsilon(w)^{D} T\big(\Phi(\theta)\big)\varepsilon(w) t\Omega \\ S.t: \begin{cases} \left\{\dfrac{1}{(\Omega)}\displaystyle\int_{\Omega} \Phi(\theta) t\Omega - C_{prescribe \le 0}\right. \\ \sigma_{PN} = \left(\displaystyle\sum_{a=1}^{N} (c_a \sigma_a^{cN})^{b}\right)^{\frac{q}{B}} \le \overline{\sigma_{PN}} \left(\left(\displaystyle\sum_{a=1}^{N} (c_a \sigma_a^{cN})^{b}\right)^{\frac{q}{B}} - \overline{\sigma_{PN}} < 0\right) \end{cases} \end{cases} \tag{7}$$

Where, $\sigma_a^c$is the mises pressure on a component, b is the b-norm parameter, $\overline{\sigma_{PN}}$ is the b-norm measure, and $\overline{\sigma_{PN}}$ is the global stress bound. A solid volume of the element is s. The algorithm's performance and accuracy as an estimate of the maximum stress values are both affected by the number you choose for b. All pressure numerical experiments in this work use b = 10.

4.5 Sensitivity testing for layouts

The objective's responsiveness to the model parameters, i.e., the strengths of the feed-forward network, is required for gradient-based optimization. The chain rule will be used to calculate the objective stored procedure sensitivity. You may calculate the density field sensitivity using the adjoint approach.

$$\frac{\partial V}{\partial \emptyset} = \lambda^{D} \frac{\partial R}{\partial \emptyset} w \tag{8}$$

Whereλ^{D} is the constructed stiffness matrix and is the conjugate gradient vector obtained from the conjugate gradient equation $R = -f$. Using the chain rule, we can write down how sensitive objective $V$is is to changes in design variable w.

$$\frac{\partial V}{\partial u} = \frac{\partial V}{\partial \emptyset} \cdot \frac{\partial \emptyset}{\partial u} \tag{9}$$

for $\emptyset$, where $N(M)$is an expression of the density field. The algorithmic differentiation method used in the free program CasADi makes it simple to get the sensitivity of $N(M)$ about the network weights w. In a similar vein, the following derivation using the chain rule may be used to do a risk assessment of the p-norm stress:

$$\frac{\partial \sigma_{PN}}{\partial u} = \frac{\partial \sigma_{PN}}{\partial \emptyset} \cdot \frac{\partial \emptyset}{\partial u} \tag{10}$$

Where, one may find the adjoint technique of $\frac{\partial \sigma_{PN}}{\partial \emptyset}$quantitative susceptibility deduction.

4.6 Deep reinforcement learning

The MDP, the central formalism in RL, has been presented, and some of the difficulties in the field have been touched on. The following discussion will categorize RL technologies into their respective groups. Both value-function-based and policy-search-based techniques may be used to address RL issues. The actor-critic method combines critical values and strategy search into a single strategy. We would then describe these methods, along with some other tools, for addressing RL issues.

4.7 Function of value

Value-function-based approaches, attempt to calculate the monetary benefit (or another measure of value) of being in a certain condition. The predicted return from beginning in state s and continuing to follow is denoted by the state-value function $X^{\pi}(t)$.

$$X^{\pi}(t) = \mathbb{E}[Q]t, \pi] \quad (11)$$

Both the optimum policy$\pi^{*}$and the ideal state-value function $X^{*}(s)$may be expressed in terms of one another.

$$X^{\pi}(t) = \max_{\pi} X^{\pi}(t)\ \forall t\ \epsilon T. \quad (12)$$

Knowledge of $X^{t}(Q)$the best policy might be retrieved by determining the course of action that maximizes the function's value at state t_s among the potential outcomes

$$\mathbb{E}_{t_{s+1}\sim\tau(t_{s+1}|t_s,b)}[X^{*}(t_{s+1})]. \quad (13)$$

The transitional dynamics T are not accessible in the RL setup. As a result, we create a different function referred to as the state-action value or quality value $X^{\pi}(t,)$ which is similar to X^{π}, with the exception that a is given as the first action and is only applied after the subsequent state:

$$P^{\pi}(t,b) = \mathbb{E}[Q]t, b, \pi] \quad (14)$$

By selecting an aggressive at each stage $(,b)$, one may determine the optimum policy given $P^{\pi}(t,b)$arg$ain\ P^{\pi}(t,b)$. According to this rule, we can also determine $X^{\pi}(t)$by maximizing $P^{\pi}(t,b)$: $P^{\pi}(t,b) = max_b P^{\pi}(t,b)$.

4.8 Dynamic programming

To learn P^{π}, we make use of the Markov property and formulate the variable as a Bellman equation, that has the recursive form:

$$P^{\pi}(t_s, b_s) = \mathbb{E}_{t_{s+1}}\left[q_{s+1} + \gamma R^{\pi}\left((t_{s+1}, \pi)t_{s+1}\right)\right] \quad (15)$$

In other words, we may utilize the present values of our approximation of P^{π} to improve it. This suggests that P^{π}can be improved through bootstrapping. This is the cornerstone of the SARSA algorithm and Q-learning.

$$P^{\pi}(t_s, b_s) \leftarrow P^{\pi}(t_s, b_s) + \alpha\delta, \quad (16)$$

Whereα is the learning rate and $\delta = Z - P^{\pi}(t_s,\ _{the\ bas})$is of the temporal difference error; Y is the goal in this case, much as in a typical regression issue. By employing transitions produced by the behavioral policy (the policy derived from), SARSA, an on-policy training algorithm, is utilized to enhance the approximation ofP^{π}, which has the effect of establishing$Z = q_s + \gamma R^{\pi}(t_{s+1}, b_{s+1})$. Q-learning is against policy since R^{π} is modified by transitioning that is not always produced by the derived policy. As an alternative, Q-learning employs $Z = q_s + \gamma = max_b P^{\pi}(t_{s+1}, b_{s+1})$, which closely resembles *.

We employ generalized policy repetition, which comprises policy evaluation and enhancement, to determine P^{*} from an arbitraryP^{π}. Minimizing TD inaccuracies from the trajectory encountered while following the policy is one way in which policy assessment helps to enhance the estimation of the functional form. By making greedy decisions based on the revised functional form, the policy can be made more effective as estimation accuracy rises. Generalized policy iteration allows these steps to be interleaved, rather than performed sequentially to obtain an optimal (as in policy iteration), speeding up the process.

4.9 Sampling

Instead of utilizing optimization techniques to bootstrapping value functions, Monte Carlo approaches use the average return from numerous policy rollouts to predict the anticipated return from a state. This means that contrary to popular belief, pure Carlo techniques are applicable in non-Markovian settings. Nevertheless, they are limited to serial MDPs, since the rollout must end before the return can be determined. To get the most out of both approaches, the

$SC(\lambda)$algorithm combines TD learning with Monte Carlo policy assessment. The $SC(\lambda)$ functions as an interpolation between Carlo computation and ramping in same way that the present value does.

Learning the benefit function $B^{\pi}(t,b)$ is a key component of another effective value approach. Provides relative values and experimental as opposed to creating utter impossibility values asP^{π}, does. Understanding relative values is similar to lowering the threshold or median level of a signal; intuitively, it is simpler to understand that one course of action will have better results than another than to understand the exact return from that course of action. Via the straightforward equation, $A\pi = Q\pi - V\,\pi$ reflects a relative benefit of actions. It is also closely connected to the baseline variability reduction approach used in diffusion policy search methods. Several modern DRL algorithms have used the concept of advantage updates.

4.10 Policy search

The search for the best policy can be done independently of any model of the value function. To maximize the anticipated return $E[R|\theta]$most people choose a parameterized strategy whose parameters may be optimized in either a horizontal stripe or horizontal stripe fashion. Both gradient-free and gradient-based techniques have been used effectively to train neural network models that encode policies. While diffusion optimization has shown promise for covering cheap parameter spaces, most DRL techniques still favor diffusion training since it is more specimens when dealing with policies that have many characteristics.

4.11 Policy gradients

An efficient learning indication of how to fine-tune a parameterized policy may be obtained from gradients. But to calculate the anticipated return, we need to take an average across conceivable paths that the present policy parameterization may provide. This takes average calls for either predetermined (via linearization, for example) or simulated annealing (via sampling) approximations. Only in a prototype system, where the fundamental changeover mechanisms can be modeled, can predictable approaches be used. For the most part, model-free RL settings use a Carlo calculation to determine the anticipated return. This Carlo estimation presents a problem for diffusion learning because gradients do not propagate through random specimens of a probability function. As a result, we use a scoring function or posterior probability estimator (known as the REINFORCE rule in RL) as an estimate of the gradient. The latter name is evocative, as maximizing the log-likelihood is a common method for supervised learning that is used in conjunction with the estimator. The log-likelihood of the sampled action is increased by the estimator's gradient ascent, which is graded by the return. Calculating the gradient of an expectancy over a linear function of a random vector about parameters may be formalized using the REINFORCE ruleθ.

$$\nabla_{\theta}\mathbb{E}_{V}[e(V;\theta)] = \mathbb{E}_{V}[e(V;\theta)\nabla_{\theta}\log o\,(Y)]. \qquad (17)$$

Because this calculation is based on the actual results of trajectories, the resultant gradients are very inconsistent. A more manageable variance may be achieved by including unbiased estimates with lower levels of background noise. The standard approach involves deducting a baseline, which implies putting more emphasis on positive updates than purely financial ones. The most elementary foundation is the average annual return across several events, although there are numerous more possibilities.

4.12 Actor-critic methods

When value features are combined with explicit consideration of the policy, we get actor-critic approaches. The "critic" (value function) provides the "actor" (policy) with constructive criticism that helps it improve. They achieve this by balancing the benefits of reducing the variation of policy grades with the drawbacks of introducing bias when using value function approaches.

Policy gradients in actor-critic approaches are derived from the value function, just as they are in others' development; the key distinction is that actor-critic approaches employ a learned value function. As a result, we will go over actor-critic techniques as a special case of gradient descent methods later on.

5. Results and discussion

This section examines the existing methods like MDP (Ran & Dong, 2022), VR (Wu, 2022), and AI (Di & Yu, 2021) with time consumption, accuracy prediction, precision value, and the recall factor by comparing with our recommended strategy. Python 3.7 is used to implement the models for accurate selections. TensorFlow 2.0.0 is used to implement the value neural network. For simulations, we employed a GNU/Linux server equipped with a 64-bit Intel Xeon Gold CPU executing at 2.10GHz.

5.1 Computation time

A computer operation's "computation time," often known as its "running time," is the amount of time needed to finish it. The quantity of rule implementations will have an impact on how long it takes to finish a computation, which may be seen as a collection of rule applications. With a logic-gate-based quantum computer, the number of unitary transformations is

directly proportional to the time required to complete a single "quantum parallel" calculation.

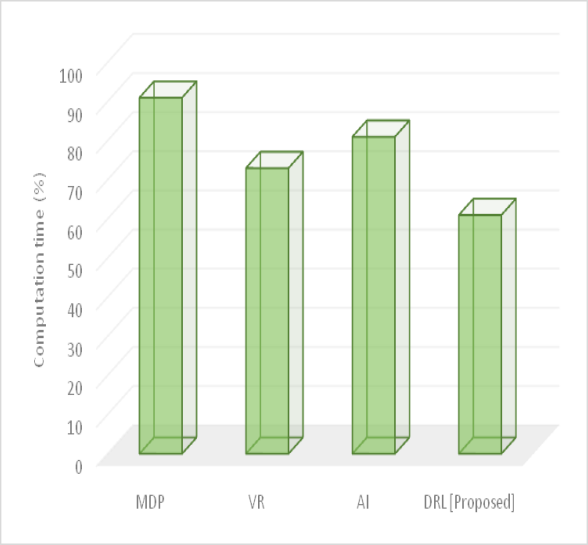

Figure 5: The computation time of the proposed and existing system

Figure 5 and Table 3 shows the computation time for proposed method. The computation time requires the DRL framework to analyze and produce optimal design configurations in an optimization technique for graphic design. For actual time applicability and easy incorporation into a graphic design process, efficient calculation time is essential for timely and flexible design optimization. Standard methods that include VR and MDP take 91% and 73% of the time. AI has an 81%-time utilization rate. The method that has been proposed requires only 61% of the computing time, which is a significant reduction.

Table 3: Comparison of computation time

Methods	Computation time (%)
MDP	91
VR	73
AI	81
DRL [Proposed]	61

5.2 Accuracy

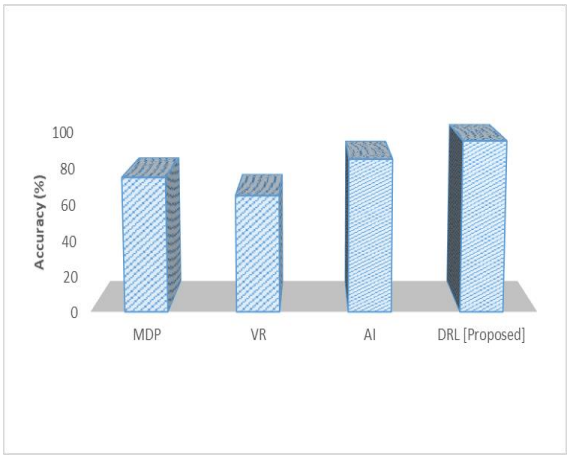

Figure 6: Accuracy of proposed and existing method

The accuracy of the suggested technique is seen in Figure 6. It is possible to think of a device's accuracy as how closely its estimations of a quantity match the value that matches that number.

$$Accuracy = (Truepositives + TrueNegatives)/ (Truepositives + Truenegatives + Falsepositives + Falsenegatives) = (TP + TN)/(TP + TN + FP + FN) \quad (18)$$

Accuracy measures how well the model produces designs that meet predetermined standards, guaranteeing the efficiency of the optimization procedure. The capability of model to apply DRL methods to produce attractive and functionally successful graphic designs is demonstrated by the high metric accuracy obtained. Conventional methods, such as VR and MDP, yield 65% and 75% accuracy. Accuracy is increased to 85% when AI is used. The proposal provides the most effective 95% accuracy rate, demonstrating its effectiveness in improved graphic design processes. Table 4 displays the accuracy of the suggested strategy.

Table 4. Comparison of accuracy

Methods	Accuracy (%)
MDP	75
VR	65
AI	85
DRL [Proposed]	95

5.3 Precision

Precision or positive predictive value is the percentage of pertinent concepts among recovered occurrences. It can imply that the standard for quality is accuracy. Precision is the extent to which the same results are achieved from the same measurements carried out under the same conditions.

Reproducibility is the variance that happens when the same technique is applied over extended times by different instruments and operators.

When every attempt is made to maintain a process, repeatability is the variance that occurs when the same equipment and operator are used and the same short amount of time is given to each repetition.

$$Precision = Truepositives/(Truepositives + Falsepositives) = TP/(TP + FP) \quad (19)$$

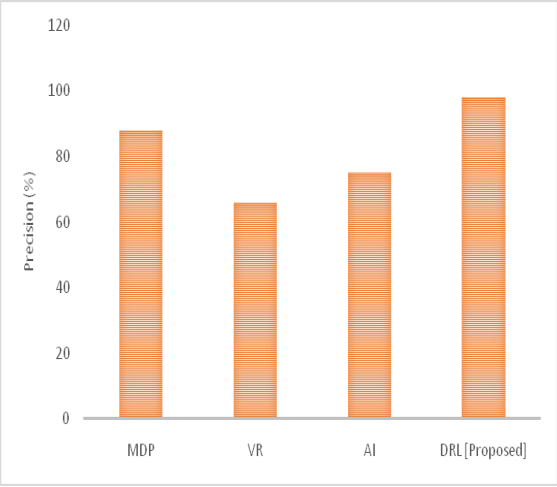

Figure 7: The precision of the proposed and existing method

The precision for the suggested system is shown in Figure 7. The precision is essential for assuring that the algorithm navigates the design space efficiently and generates visually appealing graphics. It displays the model's ability to optimize parameters for design to satisfy predetermined standards and make delicate adjustments, which increases efficiency in graphic design activities. Using a 98% precision rate, the proposed method showed outcomes. Compared with various methods, it performed better at 88%, 75%, and 66% in VR. The research objectives outcomes illustrate determining whether the DRL method succeeds in relation to obtaining higher precision. In Table 5, the suggested approach is shown.

Table 5: Comparison of precision

Methods	Precision (%)
MDP	88
VR	66
AI	75
DRL [Proposed]	98

5.4 Recall

The ability of the model to identify every significant sample in a set of data is referred to as recall. According to statistics, it is defined as the percentage of the TPs multiplied by the sum of TPs and FNs. Utilizing the formula, the recall is calculated.

$$\text{Recall} = \frac{\text{FN}}{\text{FN+TP}} \quad (20)$$

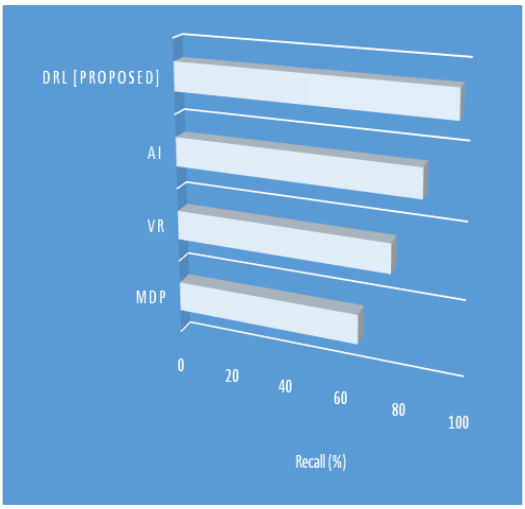

Figure 8: Recall of proposed and existing method

Comparative data for the recall metrics are shown in Figure 8. The Recall is an important component that ensures the models maintain important data and apply it to the design process, improving the efficacy and efficiency of the optimization process to produce elegant designs. With a recall of 77% VR, MDP obtains a recall rate of 66%. AI produces an 87% recall rate. The proposed exceeds other

methods with a 98% recall rate, demonstrating its effectiveness in the specific research environment. Table 6 depicts the comparison of recall

Table 6: Comparison of recall

Methods	Recall (%)
MDP	66
VR	77
AI	87
DRL [Proposed]	98

6 Discussion

Interpretability and clarification issues with DL (Zhou, Lee, Diao, Shi, Balyen, &Peto, et al, (2019)) models can prevent them from being used in domains where it is essential for explaining the decision-making process. Its application in areas with dense datasets is limited as it frequently requires substantial volumes of data with labels for efficient training, Specific knowledge can fail to identify complex patterns in data, which is the foundation of ML (Cioffi, Travaglioni, Piscitelli, Petrillo,& De Felice, et al, (2020)) methods. Complex and non-linear interactions can be difficult for the models to manage, which could result in inadequate performance on assignments where techniques for deep learning work efficiently. RL (Wang, Tang, Huang, Chen, Zhang, & Huang, (2020)) has the potential to be technically expensive and lengthy to train. Limitations include exploration-exploitation compromises, scarce reward scenarios that can cause RL models to fail and the Performance of DDPG (Bouhamed, Ghazzai, Besbes, &Massoud, (2020)) can be hindered by sensitivity to variables and training issues with stability. It could struggle with the issue of highly dimensional action spaces. When applying DDPG to intricate optimization jobs, it must be carefully adjusted and its limits need to be considered perspective in various instances. Deep Reinforcement Learning (DRL) enables the model to learn specific correlations between design elements. It provides numerous benefits in graphic design optimization. Its capacity for iterative adaptation and optimization improves the effectiveness of the method of graphic design by providing relevant information and automating complex design selections for increased innovation and efficiency.

7. Conclusion

To aid in the process of navigating graphic design files, we proposed DRL framework. The most advanced DRL techniques are often used in artificial settings where the distribution of pictures does not correspond to that of natural scenes. This is an important step in achieving more lifelike environments. Because of the rapid proliferation of generative design tools, it is now possible to augment traditional shape-finding procedures with technological answers. Our findings highlight the potential for using topological optimization techniques in the built environment. Some key takeaways are as follows:

(a) As contrasted with the conventional voxel-based optimization technique, when a neural network is used to model the density fields, the amount of architectural parameters is significantly decreased.

(b) As the topology is represented implicitly, the resulting layout does not have a staggered border.

In the long run, this paper's approach offers a fresh chance to combine deep learning with topology optimization. More advanced and robust deep-learning models have been presented in recent years. This paper's proposed approach is a hybrid of deep learning and topology optimization. More deep learning models, like CNN and GAN, will be used to represent the density field in upcoming research.

References

[1] Bichu, Y. M., Hansa, I., Bichu, A. Y., Premjani, P., Flores-Mir, C., &Vaid, N. R. (2021). Applications of artificial intelligence and machine learning in orthodontics: a scoping review. *Progress in Orthodontics*, *22*(1), 18. https://doi.org/10.1186/s40510-021-00361-9

[2] Bouhamed, O., Ghazzai, H., Besbes, H., &Massoud, Y. (2020). Autonomous UAV navigation: A DDPG-based deep reinforcement learning approach. *2020 IEEE International Symposium on Circuits and Systems (ISCAS)*.

[3] Brown, N., Garland, A. P., Fadel, G. M., & Li, G. (2022). Deep reinforcement learning for engineering design through topology optimization of elementally discretized design domains. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4010395

[4] Chen, Y., Tong, Z., Zheng, Y., Samuelson, H., &Norford, L. (2020). Transfer learning with deep neural networks for model predictive control of HVAC and natural ventilation in smart buildings. *Journal of Cleaner Production*, *254*(119866), 119866. https://doi.org/10.1016/j.jclepro.2019.119866

[5] Cioffi, R., Travaglioni, M., Piscitelli, G., Petrillo, A., & De Felice, F. (2020). Artificial intelligence and

machine learning applications in smart production: Progress, trends, and directions. *Sustainability*, *12*(2), 492. https://doi.org/10.3390/su12020492

[6] Das, N., Bechtle, S., Davchev, T., Jayaraman, D., Rai, A., & Meier, F. (2021). Model-based inverse reinforcement learning from visual demonstrations. *In Conference on Robot Learning* (pp. 1930-1942). PMLR.

[7] Deng, Z., & Chen, Q. (2021). Reinforcement learning of occupant behavior model for cross-building transfer learning to various HVAC control systems. *Energy and Buildings*, *238*(110860), 110860. https://doi.org/10.1016/j.enbuild.2021.110860

[8] Di, X. & Yu, P., (2021). Multi-agent reinforcement learning of 3d furniture layout simulation in indoor graphics scenes. *arXiv preprint arXiv*:2102.0937.

[9] Ding, X., Du, W., &Cerpa, A. E. (2020). Mb2c: Model-based deep reinforcement learning for multi-zone building control. In *Proceedings of the 7th ACM international conference on Systems for energy-efficient buildings, cities, and Transportation* (pp. 50–59).

[10] Li, H., Zhu, J., Zhou, Y., Feng, Q., & Feng, D. (2022). Charging station management strategy for returns maximization via improved TD3 deep reinforcement learning. *International Transactions on Electrical Energy Systems*, *2022*, 1–14. https://doi.org/10.1155/2022/6854620

[11] Lin, Y., McPhee, J., & Azad, N. L. (2019). Comparison of Deep Reinforcement Learning and Model Predictive Control for Adaptive Cruise Control. In *arXiv [eess.SY]*. http://arxiv.org/abs/1910.12047

[12] Luong, M., & Pham, C. (2021). Incremental learning for autonomous navigation of mobile robots based on deep reinforcement learning. *Journal of Intelligent & Robotic Systems*, *101*(1). https://doi.org/10.1007/s10846-020-01262-5

[13] Predić, B., Manić, D., Saračević, M., Karabašević, D., &Stanujkić, D. (2022). Automatic image caption generation based on some machine learning algorithms. *Mathematical Problems in Engineering*.

[14] Ran, M., & Dong, J. (2022). *A Multiobjective Optimization Algorithm for Building Interior Design and Spatial Structure Optimization. Mobile Information Systems*.

[15] Shen, Z., Wang, Y., Wu, D., Yang, X., & Dong, B. (2020). Learning to scan: A deep Reinforcement Learning approach for personalized scanning in CT imaging. In *arXiv [physics.med-ph]*. http://arxiv.org/abs/2006.02420

[16] Tao, H., Al-Sulttani, A. O., Salih Ameen, A. M., Ali, Z. H., Al-Ansari, N., Salih, S. Q., & Mostafa, R. R. (2020). Training and testing data division influence on hybrid machine learning model process: Application of river flow forecasting. *Complexity*, *2020*, 1–22. https://doi.org/10.1155/2020/8844367

[17] Tapeh, A. T. G., &Naser, M. Z. (2022). *M*achine Learning, and Deep Learning in Structural Engineering: A Scientometrics Review of Trends and Best Practices. *Archives of Computational Methods in Engineering*. 1–45.

[18] Wang, T., Tang, Y., Huang, Y., Chen, X., Zhang, S., & Huang, H. (2020). Automatic adjustment method of power flow calculation convergence for large-scale power grid based on knowledge experience and deep reinforcement learning. *2020 IEEE 4th Conference on Energy Internet and Energy System Integration (EI2)*.

[19] Yamaguchi, T., Nagahama, S., Ichikawa, Y., & Takadama, K. (2019). Model-based multi-objective reinforcement learning with unknown weights. In Human Interface and the Management of Information. Information in Intelligent Systems: Thematic Area, HIMI 2019, Held as Part of the 21st HCI International Conference, *HCII* 2019, Orlando, FL, USA, July 26-31, 2019, *Proceedings*, Part II 21 (pp. 311-321). *Springer International Publishing*. DOI: https://doi.org/10.1007/978-3-030-22649-7_25

[20] Wu, W., & Feng, Y. (2022). Interior space design and automatic layout method based on CNN. *Mathematical Problems in Engineering*, *2022*, 1–14. https://doi.org/10.1155/2022/8006069

[21] Wu, Y. (2022). Architectural interior design and space layout optimization method based on VR and 5G technology. *Journal of Sensors*, *2022*, 1–10. https://doi.org/10.1155/2022/7396816

[22] Xu, N., Zhang, H., Liu, A.A., Nie, W., Su, Y., Nie, J. & Zhang, Y., 2019. Multi-level policy and reward-based deep reinforcement learning

framework for image captioning. *IEEE Transactions on Multimedia*, 22(5), pp.1372-1383.

[23] Zhang, X., Chen, Y., Bernstein, A., Chintala, R., Graf, P., Jin, X., &Biagioni, D. (2022). Two-stage reinforcement learning policy search for grid-interactive building control. *IEEE Transactions on Smart Grid*, *13*(3), 1976–1987. https://doi.org/10.1109/tsg.2022.3141625

[24] Zhang, X., Chintala, R., Bernstein, A., Graf, P., &Jin, X. (2020). Grid-interactive multi-zone building control using reinforcement learning with global-local policy search. In *arXiv [eess.SY]*. http://arxiv.org/abs/2010.06718

[25] Zhong, X., Zhang, Z., Zhang, R., & Zhang, C. (2022). End-to-end deep reinforcement learning control for HVAC systems in office buildings. *Designs*, *6*(3), 52. https://doi.org/10.3390/designs6030052

[26] Zhou, Y., Lee, W. J., Diao, R., Shi, D., Balyen, L., &Peto, T. (2019). Promising artificial intelligence-machine learning-deep learning algorithms in ophthalmology. *Journal of Modern Power Systems and Clean Energy*, *10*(5), 264–272.

[27] Zhu, M., Wang, Y., Pu, Z., Hu, J., Wang, X., &Ke, R. (2019). Safe, efficient, and comfortable velocity control based on reinforcement learning for autonomous driving. In *arXiv [cs.LG]*. http://arxiv.org/abs/1902.00089