Reinforcement Learning with Transfer Learning for Cross-Context Building Design Optimization

Anonymous Authors[1]

Abstract

Building design optimization often relies on physics-based simulation tools, which have a high computational cost. Surrogate-assisted optimization offers a more efficient alternative by approximating simulation outputs and is typically combined with conventional optimization algorithms such as genetic algorithms or particle swarm optimization. These optimization algorithms initiate their search without any prior knowledge, requiring optimization from scratch for each new building or weather scenario. This study proposes a Reinforcement Learning (RL) approach that incorporates transfer learning through actor–critic policy reuse, enabling adaptation to new weather conditions and building types. Experimental results demonstrate improved sample efficiency, faster convergence, and reduced training variability. These findings highlight the promise of RL-based transfer learning for scalable and sustainable building design optimization.

1. Introduction

Buildings are among the largest consumers of energy globally and account for a substantial share of greenhouse gas emissions, making them a critical sector in climate change mitigation efforts (Programme, 2023). Enhancing the energy performance of buildings through improved design strategies is an effective means of reducing long-term emissions. Early design-stage decisions—such as insulation levels, glazing types, and orientation—can significantly reduce energy demand over the building's lifetime and enhance sustainability (Chen et al., 2018; Homod et al., 2014; Ramessur & Gooroochurn, 2021; Huang et al., 2015; Zhang et al., 2018; Ferreira et al., 2012; Dey et al., 2020).

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

However, optimizing building designs using high-fidelity simulation tools such EnergyPlus and TRNSYS remains computationally expensive and time-consuming. Surrogate models offer a more efficient alternative by approximating simulation outputs through data-driven learning (Asadi et al., 2014; Didwania et al., 2023; Li et al., 2017; Yu & Leng, 2021), but these models are typically task-specific. As a result, both the surrogate model training and the optimization must be repeated for each new building or weather file, limiting their scalability.

To improve generalizability, recent studies have explored surrogate models trained across multiple locations by incorporating weather features or categorical location identifiers (Kerdan & Gálvez, 2020; Zheng et al., 2024; Westermann et al., 2020). However, traditional optimization algorithms such as Genetic Algorithms (GAs), Particle Swarm Optimization (PSO), and Grey Wolf Optimization (GWO) still approach each problem independently, discarding prior search knowledge and requiring fresh exploration for every new task (Ma et al., 2020; Zheng et al., 2022). Although transfer learning techniques have been proposed in the context of evolutionary optimization (Tan et al., 2023), they remain largely unexplored in surrogate-assisted building design workflows. Transferring knowledge from previous tasks has the potential to improve convergence speed and accuracy when applied to new building configurations or climate scenarios.

Recently, Pan et al. (2024) demonstrated that RL agents can learn building design strategies by interacting with surrogate models, offering greater adaptability and sample efficiency than conventional optimization algorithms. However, their work primarily focused on single-task optimization and did not explore the transferability of learned policies across different building types or climates—a key requirement for scalable design automation.

To address this, we present an RL-based transfer learning for building design optimization, where actor–critic policies trained in one context are fine-tuned in new surrogate environments representing different weather conditions or building configurations. Our experiments show that policy transfer accelerates convergence, improves training stability, and reduces the overall cost of optimization for new

tasks. These results highlight the potential of RL-based transfer learning to enable flexible, efficient, and sustainable building design optimization across diverse scenarios.

2. Related Work

A review of surrogate-assisted optimization studies (Manmatharasan et al., 2025) showed that traditional metaheuristic algorithms such as GAs, PSO, and GWO are the most commonly used optimization techniques. These traditional algorithms start their searches without initial information and do not use knowledge from any previous tasks. Despite the advancements in surrogate-assisted optimization workflow, most optimization tasks are still approached in isolation—each new design scenario or location requiring a new optimization run. This practice is inefficient, particularly when the tasks share similar characteristics. Transferring knowledge from previous optimization problems could improve performance in new tasks by reducing both search time and simulation costs. While some studies (Jiang et al., 2018; Mweshi & Pillay, 2024) have used transfer mechanisms with the conventional evolutionary frameworks, these methods are not integrated into surrogate-assisted settings.

Recently, RL has gained attention for its ability to provide adaptive and data-driven solutions in complex optimization tasks. In contrast to traditional optimizers, RL offers native support for transfer learning through mechanisms such as policy reuse and fine-tuning. While the already mentioned work of Pan et al.2024 employed RL, they did not examine the possibility of transfer among different buildings or weather conditions. Although RL transferability remains untapped in surrogate-assisted building design, several studies outside this domain have demonstrated the effectiveness of RL transferability. Parisotto et al.(Parisotto et al., 2016) introduced policy distillation in the game environments, where a student agent accelerates convergence by imitating a pre-trained teacher policy. Wang et al.(Wang et al., 2022) applied policy transfer to well placement optimization in oil reservoirs, using autoencoders to extract latent environmental representations that enabled policy transfer. Similarly, Zhang et al. (Zhang et al., 2024) proposed a safety-critical RL approach for robotic control tasks, allowing pre-trained policies to be deployed in high-risk, resource-constrained environments while preserving safety.

This highlights a clear research gap in incorporating transfer learning into the optimization process itself. RL, with its proven ability to transfer policies across tasks, offers a promising pathway to improve both the accuracy and efficiency of surrogate-assisted building design. Therefore, our approach leverages actor–critic policy transfer within an RL framework to enable knowledge-transferable building design optimization across diverse climates and building types.

3. Methodology

3.1. Reinforcement Learning Framework

The proposed RL–based surrogate optimization approach comprises two main components: a surrogate environment and an RL agent, as illustrated in Figure 1. RL is a natural fit for scalable building design optimization due to its capacity for learning through interactions and its compatibility with transfer learning. In this work, we adopt the Deep Deterministic Policy Gradient (DDPG) algorithm, which supports continuous action spaces and is therefore well-suited for optimizing real-valued building parameters.

The surrogate environment is constructed using a surrogate model trained on EnergyPlus simulations. Given a vector of building design variables B (e.g., insulation levels, glazing properties), the surrogate model predicts key performance metrics such as annual energy consumption (Y_1) and thermal comfort measured by predicted mean vote (Y_2). These predicted values serve as the environment **state**, and are also used to compute the **reward**, which reflects the deviation from the design objectives.

The RL agent follows an actor–critic architecture. The actor network (π) proposes design configurations (**actions**), while the critic network (Q) estimates their Q-value. To improve training stability, we employ **target networks** Q_{target} and π_{target}, which are delayed copies of the main critic and actor networks. These target networks are updated incrementally using soft updates governed by a rate τ, as follows:

$$\theta_{\pi_{\text{target}}} \leftarrow \tau\theta_{\pi} + (1-\tau)\theta_{\pi_{\text{target}}} \tag{1}$$

$$\theta_{Q_{\text{target}}} \leftarrow \tau\theta_{Q} + (1-\tau)\theta_{Q_{\text{target}}} \tag{2}$$

where θ_π and θ_Q denote the weights of the main actor and critic networks, and $\theta_{\pi_{\text{target}}}$ and $\theta_{Q_{\text{target}}}$ are the weights of the corresponding target networks. Both the actor and critic networks are implemented as fully connected neural networks with ReLU activations. The architecture for each network, including the number of hidden layers and the number of units per layer, is determined through hyperparameter tuning.

The critic is trained by minimizing the Temporal-Difference (TD) error between its predicted Q-value and a target value computed using the target networks. The target value t is defined as:

$$t = r + \gamma Q_{\text{target}}(s', \pi_{\text{target}}(s')) \tag{3}$$

Here, r is the immediate reward, γ is the discount factor, and s' is the next state. The critic loss is the mean squared error between this target and the predicted Q-value:

$$\mathcal{L}_{\text{critic}} = \mathbb{E}(s, a, r, s') \left[\left(t - Q(s, a)\right)^2 \right] \tag{4}$$

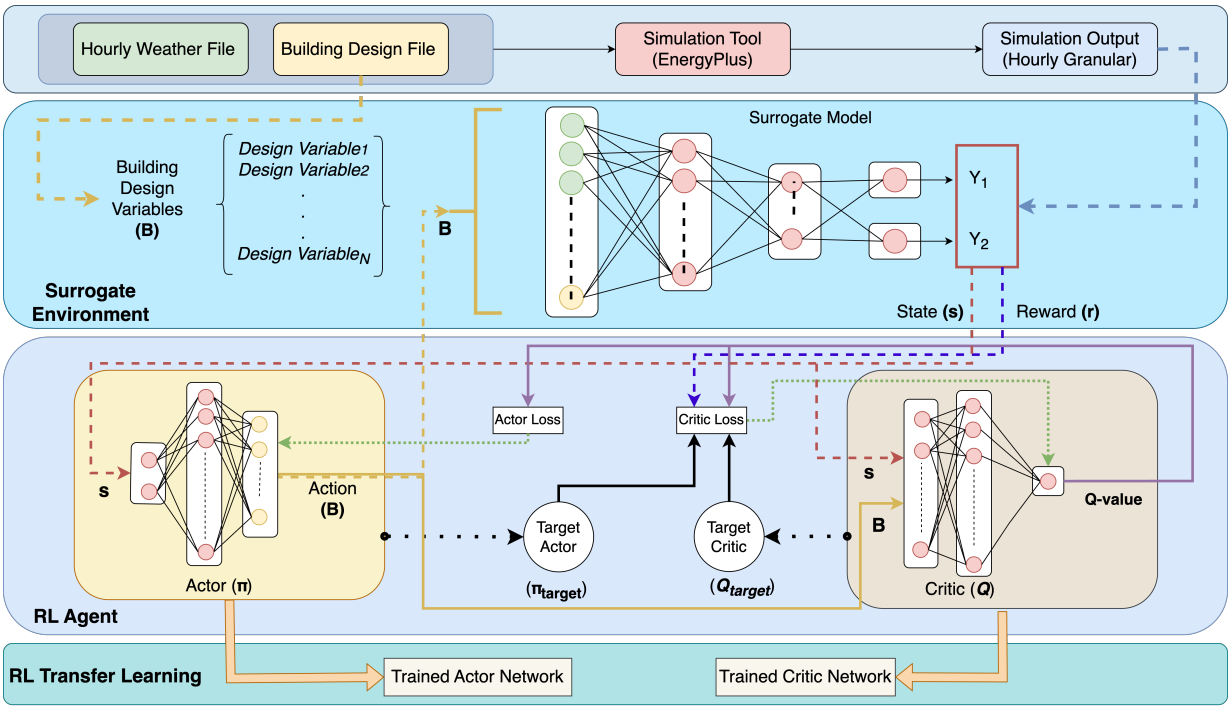


Figure 1. RL–based surrogate optimization framework. The actor–critic agent interacts with a surrogate environment trained on EnergyPlus outputs.

where s and a represent the current state and action. This setup encourages the critic to match its predictions to more stable target estimates generated by the delayed target network. To improve learning stability and sample efficiency, a replay buffer is used to store past transitions (s, a, r, s'). During training, mini-batches are randomly sampled from this buffer to update the actor and critic networks.

The actor is updated to maximize the Q-value predicted by the critic, leading to better-performing design proposals:

$$\mathcal{L}_{\text{actor}} = -\mathbb{E}s\left[Q(s, \pi(s))\right] \tag{5}$$

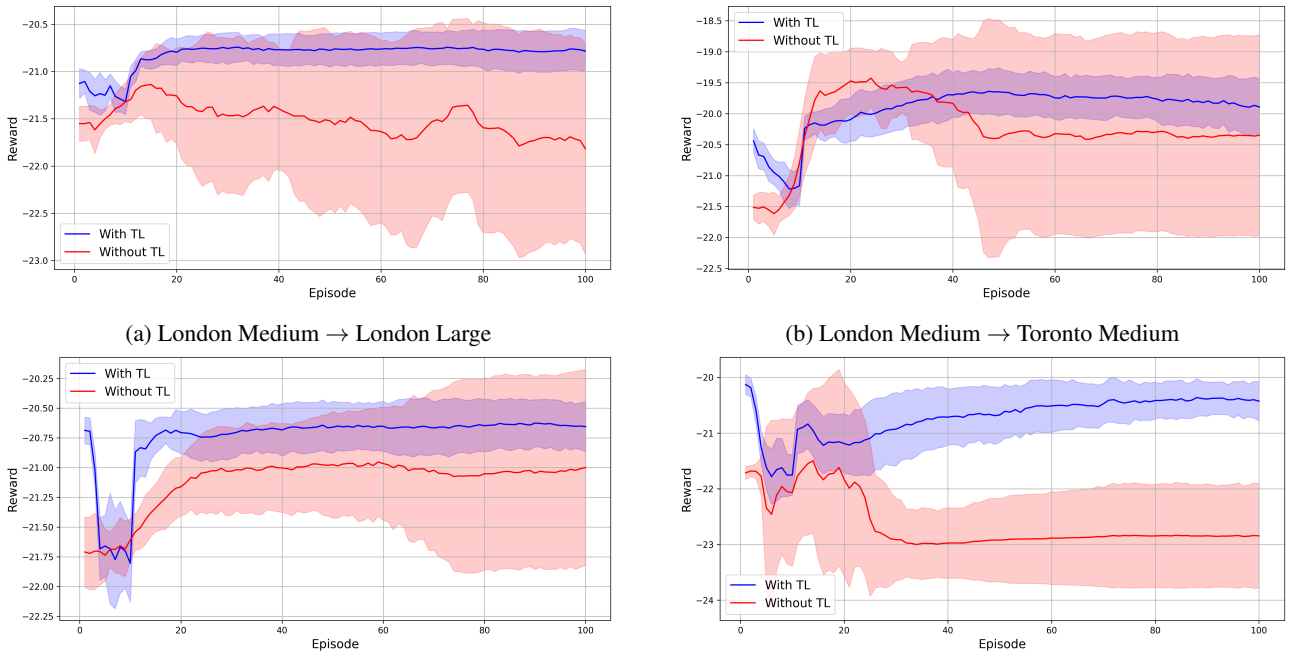
To encourage exploration in the continuous action space, Gaussian noise is added to the actions during training. The noise standard deviation gradually decayed over time to shift the agent from exploration to exploitation. Each RL training session is run for a fixed number of episodes (e.g., 100). Through iterative interaction with the surrogate environment, the actor learns a policy $\pi(s)$ that proposes high-performing building configurations. The learned actor and critic networks, which together encode optimization knowledge, are reused when transferring the RL agent to a new scenario, allowing the agent to begin learning with prior experience.

3.2. Transfer Learning for Cross-Context Adaptation

To enhance sample efficiency and reduce retraining time when transitioning to new design contexts, we adopt a transfer learning approach that reuses both the actor and critic networks. Specifically, actor–critic models trained on a source task are transferred to a new target task, where the building or weather file differs. This allows the RL agent to leverage both policy knowledge and value estimations learned previously.

In the surrogate environment, a new surrogate model is trained on the target building to represent the updated environment. The reward structure for the RL agent remains consistent across tasks, but values are now calculated using the new surrogate model. To facilitate smooth transfer across different building types with varying energy scales, energy consumption values are normalized by dividing each value by the mean energy consumption of its respective building type. This scaling ensures that performance metrics remain comparable, allowing the transferred actor to interpret feedback consistently and continue generating plausible design configurations aligned with the original objective.

Fine-tuning is performed over a number of episodes using the DDPG algorithm. During this phase, both the transferred actor and critic networks are updated from their initial knowledge using newly collected experience in the target environment. This setup accelerates convergence and improves early-stage learning stability by leveraging prior optimization knowledge embedded in both, while still allowing for adaptation to the specifics of the new task.



(a) London Medium → London Large

(b) London Medium → Toronto Medium

(c) Toronto Medium → Toronto Large

(d) Toronto Medium → London Medium

Figure 2. Transfer learning scenarios explored in this study. Each plot compares reward trajectories of models trained from scratch (red) vs. models initialized with pretrained actors (blue).

4. Results

This section presents the results of RL-based optimization experiments designed to evaluate the scalability and transferability of learned policies across different building configurations and climates. Two building types were selected from the EnergyPlus example library: the Medium Office and the Large Office reference model. Each building was simulated under two distinct climate conditions, Toronto and London (Canada), resulting in four unique optimization scenarios. For each scenario, an actor–critic RL agent was trained from scratch to learn optimal building design parameters.

To evaluate the effectiveness of transfer learning and assess whether knowledge from previous optimization episodes could improve convergence speed, stability, and early-stage performance in new but related environments, we conducted experiments in which training was initialized from four pretrained actor–critic models derived from related source tasks and subsequently fine-tuned for the new target tasks. Each transfer scenario was repeated over 20 independent trials to account for variability in training outcomes.

As illustrated in Figure 2, transfer learning consistently provided a clear advantage across all scenarios. Models initialized with pretrained actor-critic networks (blue curves) achieved higher average rewards and exhibited lower variance compared to those trained from scratch (red curves). This indicated that prior optimization knowledge is successfully leveraged to accelerate learning and improve early-stage design exploration. These benefits were most pronounced in the initial episodes, suggesting that transfer learning enhances sample efficiency and training stability.

5. Conclusion

This work proposed an RL–based transfer learning for building design optimization using surrogate models. By reusing actor–critic networks across different weather files and building types, the approach improves convergence and training stability while reducing the need for retraining from scratch. These benefits were most evident when tasks shared similar characteristics, supported by normalization strategies that aligned performance metrics across scenarios. The results highlight the potential of RL-based transfer learning to enable scalable and sample-efficient design workflows. Future work will consider advanced transfer learning strategies to enhance generalization across varied task settings and application domains.

Impact Statement

This research contributes to scalable and efficient building design optimization, supporting global efforts toward sustainable and low-carbon architecture. By reducing simulation and retraining costs, the proposed method accelerates energy-efficient design practices across diverse weather files

and buildings.

References

Asadi, E., da Silva, M. G., Antunes, C. H., Dias, L., and Glicksman, L. Multi-objective optimization for building retrofit: A model using genetic algorithm and artificial neural network and an application. *Energy and Buildings*, 81:444–456, 2014. doi: 10.1016/j.enbuild.2014.06.009.

Chen, Y., Norford, L. K., Samuelson, H. W., and Malkawi, A. Optimal control of HVAC and window systems for natural ventilation through reinforcement learning. *Energy and Buildings*, 169:195–205, 2018. doi: 10.1016/j.enbuild.2018.03.051.

Dey, M., Rana, S. P., and Dudley, S. A case study based approach for remote fault detection using multi-level machine learning in a smart building. *Smart Cities*, 3:401–419, 2020. doi: 10.3390/smartcities3020021.

Didwania, S. K., Reddy, T. A., and Addison, M. S. Synergizing design of building energy performance using parametric analysis, dynamic visualization, and neural network modeling. *Journal of Architectural Engineering*, 29, 2023. doi: 10.1061/JAEIED.AEENG-1521.

Ferreira, P. M., Ruano, A. E., Silva, S., and Conceição, E. Z. Neural networks based predictive control for thermal comfort and energy savings in public buildings. *Energy and Buildings*, 55:238–251, 2012. doi: 10.1016/j.enbuild.2012.08.002.

Homod, R. Z., Sahari, K. S. M., and Almurib, H. A. Energy saving by integrated control of natural ventilation and HVAC systems using model guide for comparison. *Renewable Energy*, 71:639–650, 2014. doi: 10.1016/j.renene.2014.06.015.

Huang, H., Chen, L., and Hu, E. A new model predictive control scheme for energy and cost savings in commercial buildings: An airport terminal building case study. *Building and Environment*, 89:203–216, 2015. doi: 10.1016/j.buildenv.2015.01.037.

Jiang, M., Huang, Z., Qiu, L., Huang, W., and Yen, G. G. Transfer learning-based dynamic multiobjective optimization algorithms. *IEEE Transactions on Evolutionary Computation*, 22:501–514, 2018. doi: 10.1109/TEVC.2017.2771451.

Kerdan, I. G. and Gálvez, D. M. Artificial neural network structure optimisation for accurately prediction of exergy, comfort and life cycle cost performance of a low energy building. *Applied Energy*, 280:115862, 2020. doi: 10.1016/j.apenergy.2020.115862.

Li, K., Pan, L., Xue, W., Jiang, H., and Mao, H. Multi-objective optimization for energy performance improvement of residential buildings: A comparative study. *Energies*, 10:245, 2017. doi: 10.3390/en10020245.

Ma, X., Chen, Q., Yu, Y., Sun, Y., Ma, L., and Zhu, Z. A two-level transfer learning algorithm for evolutionary multitasking. *Frontiers in Neuroscience*, 13, 1 2020. ISSN 1662-453X. doi: 10.3389/fnins.2019.01408.

Manmatharasan, P., Bitsuamlak, G., and Grolinger, K. AI-driven design optimization for sustainable buildings: A systematic review. *Energy and Buildings*, 332:115440, 2025. doi: 10.1016/j.enbuild.2025.115440.

Mweshi, G. and Pillay, N. Improving the performance of genetic algorithms for combinatorial optimization using machine learning for knowledge transfer. In *Proceedings of the 16th International Joint Conference on Computational Intelligence*, pp. 363–374. SCITEPRESS - Science and Technology Publications, 2024. ISBN 978-989-758-721-4. doi: 10.5220/0013084200003837.

Pan, Y., Shen, Y., Qin, J., and Zhang, L. Deep reinforcement learning for multi-objective optimization in BIM-based green building design. *Automation in Construction*, 166: 105598, 2024. doi: 10.1016/j.autcon.2024.105598.

Parisotto, E., Ba, J. L., and Salakhutdinov, R. Actor-mimic: Deep multitask and transfer reinforcement learning, 2016. URL https://arxiv.org/abs/1511.06342.

Programme, U. N. E. Building materials and the climate: Constructing a new future, 2023. URL https://wedocs.unep.org/20.500.11822/43293.

Ramessur, N. and Gooroochurn, M. Automated passive measures: the next step in reducing the carbon footprint of our buildings. In *2021 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, Brisbane, Australia, December 2021.

Tan, Z., Luo, L., and Zhong, J. Knowledge transfer in evolutionary multi-task optimization: A survey. *Applied Soft Computing*, 138:110182, 2023. doi: 10.1016/j.asoc.2023.110182.

Wang, Z., Zhang, K., Zhang, J., Chen, G., Ma, X., Xin, G., Kang, J., Zhao, H., and Yang, Y. Deep reinforcement learning and adaptive policy transfer for generalizable well control optimization. *Journal of Petroleum Science and Engineering*, 217:110868, 2022. doi: 10.1016/j.petrol.2022.110868.

Westermann, P., Welzel, M., and Evins, R. Using a deep temporal convolutional network as a building energy surrogate model that spans multiple climate zones. *Applied Energy*, 278:115563, 2020. doi: 10.1016/j.apenergy.2020.115563.

Yu, F. and Leng, J. Quantitative effects of glass roof system parameters on energy and daylighting performances: A bi-objective optimal design using response surface methodology. *Indoor and Built Environment*, 30:1268–1285, 2021. doi: 10.1177/1420326X20941220.

Zhang, Q., Wu, C., Tian, H., Gao, Y., Yao, W., and Wu, L. Safety reinforcement learning control via transfer learning. *Automatica*, 166:111714, 2024. doi: 10.1016/j.automatica.2024.111714.

Zhang, W., Liu, F., and Fan, R. Improved thermal comfort modeling for smart buildings: A data analytics study. *International Journal of Electrical Power and Energy Systems*, 103:634–643, 2018. doi: 10.1016/j.ijepes.2018.06.026.

Zheng, J., Zhang, B., Zou, J., Yang, S., and Hu, Y. A dynamic multi-objective evolutionary algorithm based on niche prediction strategy. *SSRN Electronic Journal*, 2022. doi: 10.2139/ssrn.4280056.

Zheng, Z., Zhou, J., Yang, Y., Xu, F., and Liu, H. Economic optimization of exterior wall insulation in chinese office buildings by coupling artificial neural network and genetic algorithm. *Thermal Science and Engineering Progress*, 50:102582, 2024. doi: 10.1016/j.tsep.2024.102582.