DEEP LEARNING-BASED SURROGATE MODELING FOR PERFORMANCE-DRIVEN GENERATIVE DESIGN SYSTEMS

SHERMEEN YOUSIF1 and DANIEL BOLOJAN2
1,2Florida Atlantic University.
1syousif@fau.edu, 0000-0003-4214-1023
2dbolojan@fau.edu, 0000-0003-2060-367X

Abstract. Within the context of recent research to augment the design process with artificial intelligence (AI), this work contributes by introducing a new method. The proposed method automates the design environmental performance evaluation by developing a deep learning-based surrogate model to inform the early design stages. The project is aimed at automating performative design aspects, enabling designers to focus on creative design space exploration while retrieving real-time predictions of environmental metrics of evolving design options in generative systems. This shift from a simulation-based to a prediction-based approach liberates designers from having to conduct simulation and optimization procedures and allows for their native design abilities to be augmented. When introduced into design systems, AI strategies can improve existing protocols, and enable attaining environmentally conscious designs and achieve UN Sustainable Development Goal 11.

Keywords. Deep Learning; Artificial Intelligence; Surrogate Modeling; Automating Building Performance Simulation; Generative Design Systems; SDG 11.

1. Introduction

Generative design has already been established as a commonly used method to computationally generate and examine a large set of designs and often lends itself to designers’ exploration and aesthetic evaluation. Building performance simulation (BPS) is an established method, required to target building design resiliency and adaptability to rising environmental problems (Attia, 2011). BPS has been pursued by modelers and engineers, assisting in attaining successful environmental design. Also, for optimization, it is important to couple BPS with parametric generative systems, to evaluate all design options. However, BPS remains poorly integrated into generative design processes (Danhaive & Mueller, 2021). More than before, environmental performance simulation, at both architectural and urban scales, has become essential to achieve resilience and adaptability to rising environmental problems as targeted in the UN Sustainable Development Goals 11. Within this goal, re-thinking the design process and prototyping new approaches to (facilitate) achieving high environmental performance in the built environment became important, which motivated this
Simulation is still perceived as a complicated process that requires expert knowledge. Studies show that non-experts’ simulation studies contain so many errors and shortcomings that the value of these studies in the design process becomes questionable (Reinhart & Wienold, 2011). For lay architects to perform BPS, the task requires a level of competence and technical abilities that they often do not possess. Another problem with conducting BPS studies is the lack of their seamless integration into generative design frameworks and associated issues of designer interaction. Also, there is a need to compensate for the normal idle computation time in the generative design process to support parallelized data exploration and synthesis while retaining close human-in-the-loop engagement (Urban Davis et al., 2021). More importantly, BPS is still computationally expensive and takes a long simulation time. Daylight simulation is time-consuming due to ray-tracing and requires especially an extended time for annual studies. The computation time issue becomes particularly problematic in generative parametric systems, where thousands of design alternatives emerge iteratively and require instantaneous feedback. Within BPS, energy modeling and daylight analysis are often targeted. Daylight performance has a direct impact on the building form, i.e., daylight is a form giver (Reinhart, 2019). Compact forms have fewer surfaces and thus fewer opportunities for daylighting, while extended forms perform better due to their higher surface-to-volume ratio, allowing more opportunities for window surfaces (Caldas, 2001). This means designers need to know about each design’s daylight performance to make informed decisions in regard to morphology and design revisions early in the design process.

The recent introduction of AI has leveraged design protocols, marking the transition to the second generation of generative systems (Chen & Stouffs, 2021). Since AI can analyze, learn, and synthesize data, it can aid designers in making successful decisions by enabling the prediction of environmental parameters of their designs. Yet, developing accurate, predictive real-time techniques for environmental analysis remains difficult (Rahmani Asl et al., 2017). A solution to achieve high-fidelity real-time prediction of environmental analysis in generative design is to employ deep learning-based surrogate modeling. A surrogate model is "an approximation of the input-output data obtained by the simulation" (Kim & Boukouvala, 2019, p. 2). Deep learning (DL) models are computational models with multiple processing layers that learn representations of data at multiple levels of abstraction. Generative Adversarial Networks (GANs) are DL-based models that allow machines to learn structures and semantics by extracting features from input datasets (Goodfellow et al., 2016), and thus can be used for surrogate modeling.

In this paper, we address the environmental performance of the enclosure system where we target daylight analysis with the use of a deep learning-based surrogate model. The proposed surrogate model was developed in three stages: (1) generative modeling and daylight simulation for data acquisition; (2) DL-model training for building the surrogate model; and (3) assessment and validation. In our previously published Deep-Performance framework (Yousif & Bolojan, 2021), the investigation was limited to single-space floor plans without interior walls, to test the initial prototype of the proposed method. In this paper, we present further development to our method by improving the surrogate model and conducting additional experiments to include
more complex spatial configurations for the input datasets, as well as interior wall partitions and multi-room floor plan layouts. Our improved method shows promising results with regards to accurate predictions of daylight performance for complex floor plan designs with interior wall partitions and multi-space configurations (Figure 1).

Figure 1. Left: The previous dataset with simple single-space layouts; right: the new dataset with complex multi-space floor plan layouts.

2. Background

It is speculated that AI is most likely to have the biggest influence on performance-based aspects, particularly in architectural practice and urban design, where data-driven approaches and performance-informed design are becoming increasingly important (Leach, 2021). Reviewing existing literature, performative AI has seen an exponential increase in architectural research addressing multiple building performance aspects. An important reference for AI research in design is the City Intelligence Lab in Austria, offering a digital platform that utilizes AI technologies for urban planning workflows and processes. The lab examines technologies that involve generative design and AI solutions for data-driven design (2019). In design practice, the XKool firm investigates how deep learning can be used to construct numerous environmental scenarios, based on training datasets that combine geographical variables and building codes (2019). Another pioneer is Spacemaker, a design firm that offers AI-driven design and planning solutions for early design stages in pursuing sustainable built environments (2020). In approximating building energy modeling, the work of (Singaravel et al.) uses a method of component-based machine learning for mimicking BPS (2018). Papadopoulos et al. (2018) employ machine learning techniques combined with genetic algorithm-based optimization to offer energy use evaluations of building designs. Research is also expanding in using AI for automating daylight analysis (i.e., Ngarambe et al., 2020; Shaghaghian & Yan, 2019) Despite such progress in performative AI, these daylight-related approaches represent undergoing experimentation and do not yet offer validated methods for real-time daylight performance prediction in generative systems, which has motivated this project. We extend the state-of-the-art by introducing a new deep learning-based surrogate model that predicts daylight performance of floor plan designs with high accuracy.

Before describing the research methodology, we offer here definitions and explanations of the implemented methods. Surrogate models are prediction models that seek to approximate the output of simulation models as closely as possible and can offer compact and instantaneous performance information instead of simulation (Forrester et al., 2008). GANs are techniques for training a machine to perform complex tasks in a generative process measured against a set of training images (Goodfellow et al., 2016; Leach, 2021). A sub-class of GANs are the conditional or
"supervised" models such as Pix2Pix. In this image-to-image translation model, synthesizing images is based on labeling or pairing corresponding datasets. Reconstructing and producing images occurs based on the labels of one image of the pair (Isola et al., 2017). We have adopted the Pix2Pix model and further improved it, formulating a revised strategy to be applied to building designs, and their corresponding datasets. To verify and validate the accuracy of the DL prediction results, we implemented two quantified metrics, the Structural Similarity Index Model (SSIM), and Perceptual Similarity (PS). SSIM is a method for assessing AI-synthesized image quality that involves collecting structural data and evaluating the degradation of that data for the images in question (Wang et al., 2004). PS is another method for evaluating deep learning-generated images in a manner comparable to how humans make perceptual decisions (Zhang et al., 2018).

3. Research Methods

As explained above, this study was aimed at developing a new framework that incorporates an accurate approximation method, a surrogate model for predicting daylight studies of design options in generative design protocols. The methodology included experimentation with the surrogate model techniques, developing the model to be integrated into a performance-driven generative framework, prototyping the overall framework, application, and testing. In the development phase (authors’ framework), the prototype was formulated into three tasks, as illustrated on the left side of Figure 2. First, (1) dataset acquisition was pursued using a parametric system with daylight simulation integrated, (2) the DL-based model was trained for prediction of daylight performance, and (3) assessment and validation studies were conducted, comparing prediction with actual simulations. For the system users (designers), in the application phase, the system becomes a two-process workflow that consists of a generative process (with floor plan design options) and a real-time daylight performance prediction offered by our trained model, as shown in the right part of Figure 2. The framework prototype comprises a package of algorithms that is under development and will be released as a Grasshopper add-on.

![Figure 2. The workflow of the proposed DP framework, real-time daylight performance evaluation predictions using deep learning. The left side is the authors’ workflow, and the right side is the users’ workflow.](image)

3.1. SURROGATE MODEL DEVELOPMENT

In our previous test-case application (Yousif & Bolojan, 2021), the method was applied
DEEP LEARNING-BASED SURROGATE MODELING
FOR PERFORMANCE-DRIVEN GENERATIVE DESIGN
SYSTEMS

to a sample of simplified floor plan layouts. To improve the prototype and apply the
DL model to real (existing) floor plan designs, this experiment was conducted. The
framework was developed using different methods for the required tasks. Dataset
generation was done using the Rhino/Grasshopper® environment for parametric
modeling and LadyBug® and HoneyBee® plugins (Roudsari, Pak, and Smith 2013)
were used for daylight simulation. PyTorch® and Tensorflow® deep learning
packages were used to develop and train the surrogate model. Validation was
performed using machine learning algorithms such as Structural Similarity Index
Metric (SSIM) and Perceptual Similarity (PS).

3.1.1. Dataset Generation

To generate the required data for training, we defined a parametric model to create
1815-floor plan design options that represent typological design layouts (Figure 4), and
to perform daylight simulation for those generated designs. In the testing phase, the
"CubiCasa" (Kalervo et al., 2019) dataset was used. Each design option was subjected
to five annual daylight simulation metrics: (1) spatial Daylight Autonomy (sDA), (2)
Direct Light Access (DLA), and (3) Useful Daylight Illuminance ranges of UDL10-
100, (4) UDL100-2000, and (5) UDL2000_more in lux. sDA is an annual metric that
quantifies the ratio of the area within a space for which the daylight autonomy exceeds
a specified value (IES, 2020). DLA measures the space’s access to direct daylight for
a specific duration, while UDLI is a ratio of time across a daylighting study period
where the illuminance at a point is between certain minimum and maximum levels.
Each daylight simulation was conducted using a grid resolution of 0.2m (the distance
between the test points/daylight sensors) and at a typically used working plane height
of 0.76m. As shown in Figures 4, the floor plans were annotated by defining three
major classes/labels. The light gray color represents the floor area class, the black label
is for the wall class, and the yellow color signifies the window class. In this instance,
the yellow color saturation determines the window height (in this instance, we used a
uniform ceiling-to-floor height for all windows), and the gray shading denotes the
program of the space, which was an open office in this experiment. Future
developments and additional test-case applications will use different colors to define
program activities, such as office, retail, as well as considering shading devices.

Figure 3. Topological Data Analysis. Dataset augmentation is represented through a U-Map
algorithm. The top row represents the original dataset distribution, while the bottom row illustrates
the augmented data distribution.
In the dataset management phase, 1815 floor plan samples were augmented by applying various transformation procedures (preserving the north orientation) to the parametric model prior to running the corresponding daylight simulation, increasing the dataset to 5400 pairs. Daylight metric simulations were carried out with the newly augmented dataset and paired with their corresponding floor plan layouts. For training, the data was structured as follows: 85% of the paired images were used for training and 15% were used as testing sets. Thus, we used 4590 pairs for training and another 810 for testing. Five Pix2Pix models were trained, with each model corresponding to one daylight metric. The dataset floor plans comprised different complexities, multiple/single spaces, and areas varying between 50 and 200 square meters. Data augmentation was a crucial task before model training. Normally, the performance of the deep learning model depends on two aspects: the neural network model’s architecture, and the quality and quantity of data. Even when using the proper model for the intended task, the results can be unsatisfactory. The reason for this can be attributed to the dataset quality and diversity; therefore, understanding dataset topology is crucial. We used a U-Map algorithm to reduce the dimensionality of the features’ manifold and to understand and analyze the topology of the dataset. As the upper row of Figure 3 shows, the initial dataset of 1815 paired images had a very limited (almost linear) distribution and limited variation (diversity). Therefore, domain-specific augmentation techniques were applied to create new and different data points (training samples), resulting in a more diverse and enhanced distribution of the dataset, as shown in the lower row of the figure.

Figure 4. A sample of 4 design options was evaluated against 5 daylight simulation metrics of sDA, DLA, UDLI100, UDLI100-2000, and UDLI2000_more.

3.1.2. Model Training

For training the model, the network architecture used is based on the Pix2Pix model, yet we improved the networks’ filters to create feature maps that summarize the identified features in the input data (Figure 4). All five models were trained for a total of 200 epochs, which required approximately 24 hours of training. For the first half of the training, the Adam optimizer was utilized with a learning rate of 0.0005, which was then decreased linearly to 0.0000 throughout the remaining iterations. For both generator and discriminator networks, the optimizer momentum term was set to 0.65. The resolution was set to 1024x1024 pixels for both the input and output. Due to GPU training device memory constraints, a batch size of 40 was used, and we normalized the input layer with batch normalization (Ioffe & Szegedy, 2015). For the discriminator network, the basic network architecture of 70x70 PatchGAN was used, and the generator network used was a RESNET 9blocks network architecture. To reduce the model’s oscillation, we used a pool size of 50 (Shrivastava et al., 2017). Each network’s layer is followed by instance normalization and a ReLU layer.
3.1.3. Results and Validation

In the testing phase, when new floor plan designs were introduced to the trained model, it predicted the daylight simulation successfully. The accuracy of the synthetic (predicted) daylight meshes was assessed against the HoneyBee®-based (actual) simulation results. In one assessment study, quantifying the accuracy comparison between the surrogate model and the simulation model results. The results were evaluated by quantifying the ratio of the area of the space for which the daylight metric exceeds or lies within a specified value. This represents the same method used by the HoneyBee® add-on to calculate the daylight metrics. The comparison results showed that the network predicted the sDA metric with an average prediction accuracy of 89% when presented with a new batch (unseen by the trained model) of 15-floor plan alternatives of real floor plans, the CubiCasa dataset, (Figure 5). The figure demonstrates a comparison between the simulation sDA values and meshes performed using HoneyBee® and the DP-Prediction of sDA values and meshes.

<table>
<thead>
<tr>
<th>Simulation sDA</th>
<th>Prediction sDA</th>
</tr>
</thead>
<tbody>
<tr>
<td>89.9</td>
<td>84.4</td>
</tr>
<tr>
<td>17.2</td>
<td>12.0</td>
</tr>
<tr>
<td>48.9</td>
<td>39.4</td>
</tr>
<tr>
<td>95.0</td>
<td>87.6</td>
</tr>
<tr>
<td>28.9</td>
<td>22.3</td>
</tr>
<tr>
<td>31.0</td>
<td>26.0</td>
</tr>
<tr>
<td>48.2</td>
<td>40.3</td>
</tr>
<tr>
<td>50.6</td>
<td>45.0</td>
</tr>
<tr>
<td>96.2</td>
<td>94.1</td>
</tr>
<tr>
<td>47.1</td>
<td>50.6</td>
</tr>
<tr>
<td>85.8</td>
<td>84.5</td>
</tr>
<tr>
<td>96.2</td>
<td>92.1</td>
</tr>
<tr>
<td>22.3</td>
<td>29.2</td>
</tr>
<tr>
<td>85.0</td>
<td>95.1</td>
</tr>
<tr>
<td>94.6</td>
<td>94.6</td>
</tr>
<tr>
<td>94.5</td>
<td>94.5</td>
</tr>
</tbody>
</table>

Figure 5. Prediction results of testing the trained model with the (CubiCasa) dataset- comparison between the simulation sDA values and DP-Prediction of sDA values.

It is important to note here that one advantage of the Pix2Pix model is that it is trained with images that represent daylight meshes (grids), which can be easily quantified and measured post-training. Quantitative assessment methods were followed, assessing the accuracy of the prediction results of the surrogate model. The Structural Similarity Index Measure, or SSIM, was the first validation study we conducted. In this comparison, the predicted daylight simulation results were evaluated against the HoneyBee® simulation, and the results of testing (four floor plans of the CubiCasa dataset) show an average similarity value of 0.95 with an overall similarity value range of 0.91-0.99. For the second assessment method, perceptual similarity, the lowest value retrieved was 0.93 and the highest value was 0.97. Both assessments were performed using the sDA metric (Figure 6).

Figure 6. Four floor plans of the testing dataset (CubiCasa) with synthetic-real pairs were assessed by the SSIM (upper row) and PS measure (lower row) for the sDA metric.
4. Discussions

Our system, as presented, can predict daylighting with high accuracy. More significantly, this test-case application has shown promise in automating other environmental performance goals. Automatic prediction is possible with such a surrogate model, which is useful for decision-making at the early stages of design. The prediction model is to be injected into design generation, as depicted in Figure 7. It will allow designers to explore a wide range of design choices while assessing design performance in real-time throughout the inference phase. The significance of retrieving predicted daylight analysis is its impact on morphology and associated design decisions in design development. Automation of performative aspects accelerates and improves design decision-making, allowing for a faster feedback loop between design decision and environmental evaluation.

The proposed method is 600 times faster than the typical annual daylight simulation performed by HoneyBee® and required for the floor plans under consideration. Compared to the HoneyBee® daylight simulation results that took 3 minutes for each simulation run, our surrogate model was able to offer a comparable accuracy of 90%, taking less than 0.3 seconds for each prediction. In a generative design process, when exploring 6000 design options, it would take 3*6000 = 18000 minutes, equivalent to 12.5 days to retrieve daylight simulation results using HoneyBee® or Diva, in contrast to 0.3*6000 = 1800 seconds, equivalent to 30 minutes using the surrogate model. Besides saving computation time, our method offers a pre-trained model that can be used by designers into generative protocols for instant feedback on daylight performance. This way, any designer can have access to environmental performance of their designs.

5. Conclusions and Future Work

Presented here is research that contributes to the transition from simulation to prediction-based performance evaluation, using surrogate modeling. An approach was developed to provide real-time daylight performance predictions of high fidelity to enhance generative design methods. The findings suggest that deep learning approaches might be used to automate additional building performance measures. The significance of this research is to enable systematic performance-driven design space navigation by injecting trained models into a design process driven by designers’ creative exploration. The ultimate objective is to enable environmentally efficient and affordable design of the built environment using data-driven approaches. This goal is aligned with SDG 11 to make cities and human settlements resilient and sustainable. In order to achieve sustainability, design processes should be performance-driven and involve environmental feedback of design options and enable identifying
environmental consequences of design decisions.

For future work, in further developing the model, we aim to use labeling the floor plans according to program activities to achieve an accurate simulation for realistic floor plans with multiple-program activities. Also, the next step is to add parameters for window heights and shading devices (and their dimensions), encoding this information into the floorplan labels. In addition, more dynamic and angled floor plans will be pursued in future applications. Another future work involves facilitating optimization by filtering the optimum layout/s in terms of their daylight performance. The framework will also be improved to achieve an articulated and searchable design space when navigating through thousands of design options. Also, additional AI models will be required to sort out successful design alternatives with higher environmental performance. Further improvements will be focused on formulating a designer-friendly interface that will be assessed by lay architects in empirical studies. To achieve computation efficiency, a cloud-based interface integrating our method will be pursued.

References


