A RAPID PREDICTION MODEL FOR VIEW-BASED GLARE PERFORMANCE WITH MULTIMODAL GENERATIVE ADVERSARIAL NETWORKS

XIAOQIAN LI[1], ZHEN HAN[2], GANG LIU[3] and RUDI STOUFFS[4]

[1,2,3] *School of Architecture, Tianjin University.*
[1,2,4] *Department of Architecture, National University of Singapore.*
[1]*lixiaoqian_95@tju.edu.cn, 0009-0009-0160-6817*
[2]*zhen.han@u.nus.edu, 0000-0002-9547-9063*
[3]*liug@tju.edu.cn, 0000-0002-7864-7846*
[4]*stouffs@nus.edu.sg, 0000-0002-4200-5833*

Abstract. Machine learning-based glare prediction has greatly improved the efficiency of performance feedback. However, its limited generalizability and the absence of intuitive predictive indicators have constrained its practical application. In response, this study proposes a prediction model for luminance distribution images based on the multimodal learning approach. This model focuses on objects within the field of view, integrating spatial and material features through images. It also employs semantic feature mapping and multimodal data integration to flexibly represent building information, removing limitations on model validity imposed by changes in design scenarios. Additionally, the study proposes a multimodal Generative Adversarial Network tailored for the multimodal inputs. This network is equipped with unique feature fusion and reinforcement blocks, along with advanced up-sampling techniques, to efficiently distill and extract pertinent information from the inputs. The model's efficacy is verified by cases focusing on residential building luminance distribution, with a 97% improvement in computational speed compared to simulation methods. Offering both speed and accuracy, this model provides designers with a rapid, flexible, and intuitive supporting approach for daylight performance optimization design, particularly beneficial in the early design stage.

Keywords. Glare Prediction, Prediction Model, Multimodal Model, Generative Adversarial Networks

1. Introduction

Daylight is crucial for indoor environments; it offsets electric lighting costs and promotes the health and well-being of occupants (Al Horr et al., 2016). However, uncontrolled penetration of daylight into buildings could lead to undesirable luminous environments, impairing vision or causing visual discomfort. Therefore, glare autonomy is a vital part of improving the quality of building indoor environments.

However, during the early stages of building design, assessing glare presents a unique computational challenge because of complex and time-consuming rendering and calculations. This makes it hard for designers to receive real-time performance feedback during the process of selecting or improving design schemes, limiting its application in practical design (Jones et al., 2018). Three trends in accelerating glare evaluation research are 1) Idealizing the light transmission process or screening out focus areas for simplified calculations (Giovannini et al., 2020); 2) Accelerating rendering by tracing multiple primary rays in parallel on a Graphics Processing Unit (Jones, 2019); 3) Implementing rapid feedback with predictive models based on Machine Learning algorithms (Ayoub, 2019). Predictive models are increasingly favored for their cost-effectiveness and rapid processing capabilities.

Some relevant papers are reviewed below (Table 1). Predictions mainly focus on numerical evaluation metrics, such as Daylight Glare Probability (DGP), vertical eye illuminance (E_v), or annual metrics Annual Sunlight Exposure (ASE). However, there is less research on intuitive image-based indicators, such as luminance distribution maps, even though intuitive visual formats are more readily accepted and understood, matching better with the requirements of the design process. Moreover, the generalizability of prediction models is a key focus for users, as it determines models' adaptability to various design scenarios. However, it is a significant limitation in existing prediction models (Ngarambe et al., 2022). Crucial reasons for this are, besides the network structure's inherent attributes, the selection of design elements inputted and their feature representation methods. These are crucial as they dictate whether solutions with significant variations can be effectively translated through a uniform rule. In reviewed research, inputs typically consist of geometry, materials, and environment features, which are converted into numerical variables for the network's inputs (Pierson et al., 2018). However, this approach's adaptability is typically suboptimal, for instance, failing to accommodate diverse spatial forms or material configurations. One potential reason is the reliance on single-modal data, which is inadequate for representing all features comprehensively. In contrast, multimodal data, integrating the advantages of different data types, has proven highly effective in other research areas. This approach, however, is seldom employed in the field of building performance prediction, suggesting an untapped potential for multimodal data to enhance the generalizability of representing building information.

Therefore, to develop a sufficiently generalized and intuitive method for glare prediction, this paper proposes a multimodal feature representation method combined with semantic information mapping and its corresponding network structure. This method employs images to depict features of sub-aspects within the field of view, such as the type, material, and location of surfaces, and combines these with numerical data on environmental conditions and viewpoints. This combination allows for a versatile portrayal of the spatial features, adaptable to different design scenarios. Furthermore, based on Generative Adversarial Networks (GANs), the proposed approach includes operations such as multimodal feature fusion, key feature reinforcement, and refined up-sampling to satisfy the requirements of a network with multimodal inputs. This model enables the rapid generation of image-based glare evaluation from any viewpoint, effectively adjusting to any changes in building details, which provides efficient and effective support for designers.

Table 1. Machine learning-based prediction model for indoor glare prediction

Ref.	Inputs			Output			Algorithm
	Variables			Perfor-mance metrics	Data type		
	geome-try	mate-rial	wea ther		numer-ical	image	
Xie et al. 2023	√	×	×	DGP	√	×	GBRT
Radziszewski et al., 2018	√	×	√	DGP	√	×	ANN
Xie et al. 2021	×	×	√	DGP	√	×	KNN, RF
Liu et al.2020	×	×	√	Iuminance	×	√	DNN
Luo et al. 2022	√	×	√	Ev	√	×	ANN
Lin et al.2021	√	×	√	ASE	√	×	ANN
Nourkojouri et al. 2021	√	√	×	ASE	√	×	ANN
Ayoub et al.2019	√	×	×	ASE	√	×	ANN
Mostafavi et al. 2022	√	×	×	Iuminance	×	√	GAN

2. Methods

2.1. SEMANTIC INFORMATION MAPPING

The essence of semantic information mapping is to utilize pixel brightness and spatial attributes to detail interior surfaces, including their type, material, shape, and location. This mapping is grounded in the imagery captured from a single viewpoint by a perspective camera. The approach narrows the learning focus to the field of view, rather than the entire space, and the conversion to images impervious to spatial form alterations. Subsequently, the view field content is segmented into subregions. Each is tailored to different semantic objects based on surface types. These segments are merged with their respective material properties. This fusion forms the final semantic information matrix, a cohesive representation of spatial and material attributes. Additionally, capturing imagery with a fisheye camera typically requires rendering a 3D model. Due to the fisheye camera's broader range compared to a perspective camera, this study introduces a technique to grid the hemisphere in the viewing direction to reduce computational load, i.e., merging multiple perspective views to simplify fisheye capture. The specific calculation steps are as follows:

Step 1: Construct the 3D building model, identifying and color-coding various surface attributes.

Step 2: Determine the target viewpoint and its direction. Construct a hemisphere centered at the viewpoint and perpendicular to the line of sight. Grid the hemisphere based on precision requirements. Create sub-viewing directions from the viewpoint to the center of each grid, set the camera's focal length, and capture interior images along these viewing directions. The details are illustrated in Figure 1.

Step 3: Based on captured images, construct semantic matrixes that categorize the types of interior surfaces.

Step 4: Identify the material properties of each surface type and map them in situ to

the semantic matrix, creating the image input for this model. It is important to note that transparent materials, such as glass, may cause an overlap of multiple types of surfaces, for instance, seeing another wall through a window. In such cases, their material properties are a blend of multiple materials, as detailed in Table 2.
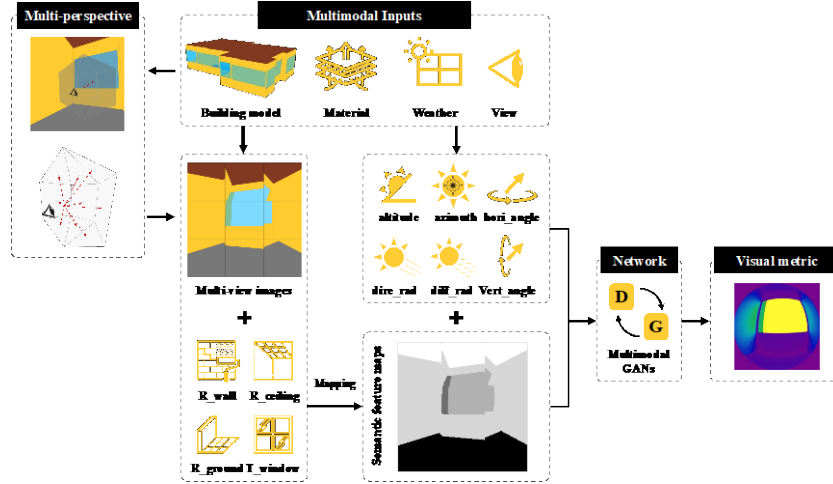

Figure 1. The framework of the prediction model for view-based glare performance.

Table 2. Definitions of surface material properties.

Surface type	Material index
Window+Wall	window_t * wall_r
Window+Window+Wall	window_t * window_t * wall_r
Window+Window+Ceiling	window_t * window_t * celing_r
Window+Window+Floor	window_t * window_t * floor_r

2.2. MULTIMODAL INPUT VARIABLES

In addition to the spatial and material features, environmental conditions and viewpoint information are also key factors influencing glare. The environmental variables are chosen based on existing research. Solar altitude and azimuth angles determine the daylight level, while sky conditions are abstracted to direct normal radiation and diffuse horizontal radiation, reflecting the external daylight transmission. Additionally, the horizontal and vertical angles of the viewpoint are considered to illustrate the relationship between the view direction and the angle of incoming daylight. These six variables, quantified numerically, combine with the semantic matrix to form the multimodal input for the prediction model, as shown in Figure 1.

2.3. MULTIMODAL GENERATIVE ADVERSARIAL NETS

In this study, a novel multimodal GAN (mGAN) is proposed for multimodal inputs. This framework is constructed based on the Pix2Pix model (Isola, 2017), which

consists of a Generator and a Discriminator, receiving multimodal inputs composed of images and vectors. The network structure is illustrated in Figures 2 to 4.

To accommodate multimodal inputs, the mGAN incorporates a vector-based feature encoding and a feature fusion mechanism. Vector features are sequentially input into the network and transitioned from a low-dimensional latent space to a higher spatial dimension through a series of transposed convolution blocks, aligning their dimensions with image features. Simultaneously, image features are exacted with a series of convolution blocks, progressively increasing their channel count, integrating more refined information, and preserving spatial structure. Finally, the Hadamard product is used to fuse the extracted vector and image features.

Beyond the standard convolution and transposed convolution blocks in Pix2Pix, the Generator's up-sampling integrates PixelShuffle blocks with transposed convolution blocks. PixelShuffle blocks upscale by rearranging channel data instead of inserting blank pixels, transforming feature maps from lower resolution with more channels to higher resolution with fewer channels. This approach maintains uniform pixel distribution and spatial consistency during up-sampling, thus avoiding the checkerboard effect often associated with transposed convolution only. This blend of upscaling techniques enables the model to flexibly manage feature representation and information flow across various layers, balancing learning capacity, image quality, and computational efficiency.

Additionally, to solve the considerable dimensional discrepancy between image and vector inputs, the Generator of the mGAN, besides its existing skip connections, has added reinforcement connections. This involves exacting vector-based inputs with various depths and incorporating these multi-level deep features as part of the resources into the corresponding up-sampling location. This strategy is designed to prevent the loss of vector information due to excessive dimensional differences. During the feature extraction process, inception blocks from GoogleNet are used, utilizing convolution kernels of different scales for multi-scale feature refinement, thereby enriching detail.

The standard pix2pix loss function is used in the mGAN. The total loss function is a weighted sum of the adversarial loss and the L1 loss to improve the generalizability and robustness of the network.
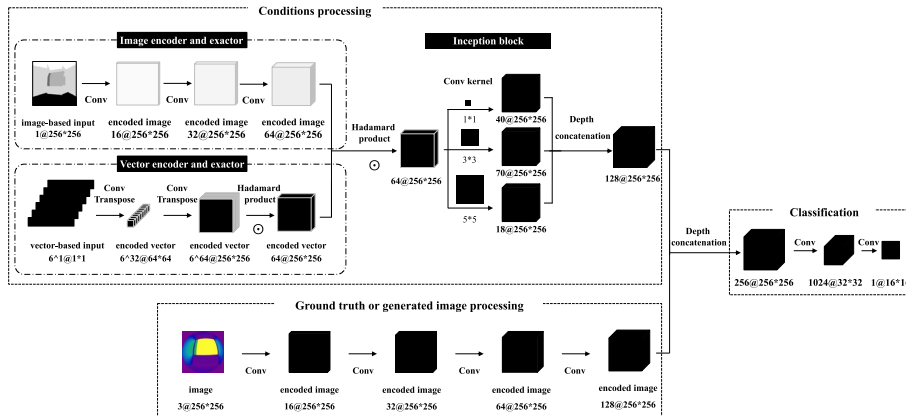

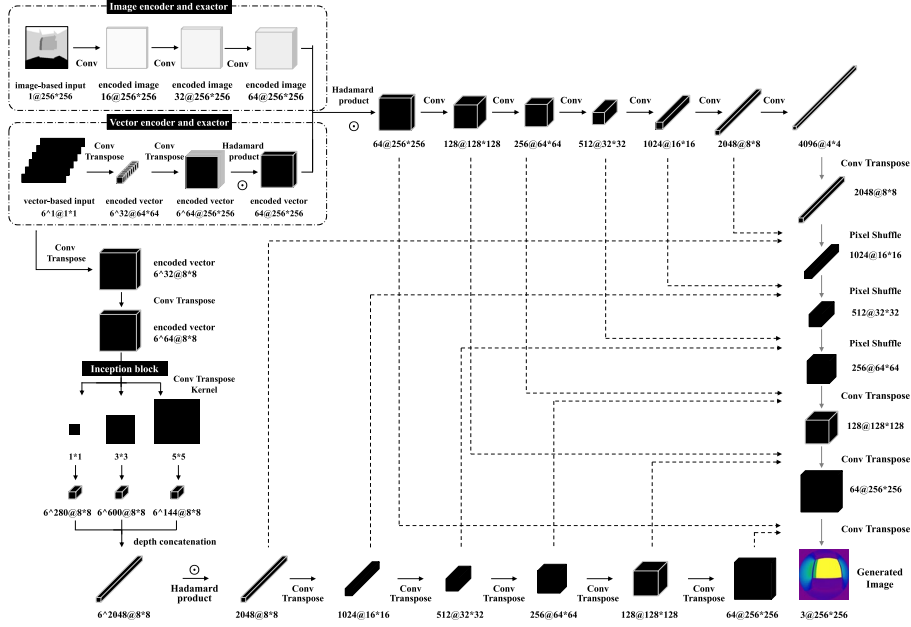*Figure 2. The construction of the Discriminator.*
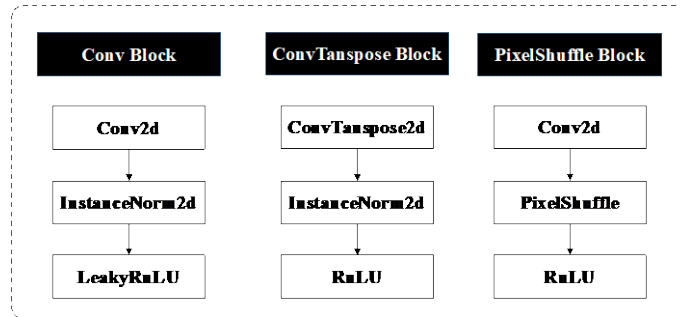
Figure 3. The construction of the Generator.

Figure 4. The construction of key blocks in the mGAN.

3. Evaluation of the prediction model

To test the performance of the proposed glare prediction model, it has been applied to predict the luminance distribution within residential buildings.

The data sources for this study comprise a building form scheme dataset, a material dataset, and a weather dataset. The scheme dataset employed in this study is RPLAN, an Asian residential building floor plan dataset assembled by Wu et al. in 2019. 5000 building models were generated with Grasshopper (single-wall models for daylight simulation). Material attribute parameters are guided by the Chinese Daylighting Design Code, involving the reflectance of the walls, ceiling reflectance, floor reflectance, and the transmittance of windows. The weather data is sourced from twelve cities in China, each belonging to one of five different light climate zones, and

covers the time range from 8:00 AM to 6:00 PM. Data from these multiple sources are randomly selected and integrated to develop a simulation model, which is utilized for generating the learning target with Ladybug.tools. To standardize the visualization mapping relationship, the display lower boundary for cloud maps is set to 0 lux, with an upper boundary of 2000 lux.

The dataset comprises 4,800 samples, allocated in an 8:2 ratio for training and testing purposes. To improve training efficiency, the resolution of the input images has been downscaled to 256x256 pixels. Additionally, numerical features have been normalized to a range between 0 and 1, ensuring a balanced influence on the overall data distribution. The detailed hyperparameter settings are shown in Table 4.

The performance of the prediction model is evaluated in terms of both data accuracy and quality of the images. The accuracy metrics include MAE (Mean Absolute Error) and MAPE (Mean Absolute Percentage Error), while image quality is assessed through SSIM (Structural Similarity Index Measure) and Image Difference Analysis (IDA).

Table 3. Ranges of material, weather and view variables.

Variables of material	Range
Reflection rate of wall (wall_r)	[0.3, 0.7]
Reflection rate of ceiling (celing_r)	[0.5, 0.9]
Reflection rate of floor (floor_r)	[0.1, 0.5]
Transmittances of windows(window_t)	[0.5, 0.8]
Solar azimuth angle(az_sun) Solar altitude angle (al_sun) Direct normal radiation (dir_rad) Diffuse horizontal radiation (dif_rad)	Data from Beijing, Shanghai, Chengdu, Guangzhou, Harbin, Hohhot, Kunming, Lhasa, Shenyang, Urumqi, Xian, Xining
View azimuth angle(az_sun)	[0, 360]
View altitude angle (al_sun)	[0, 360]
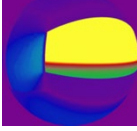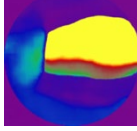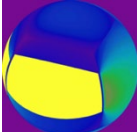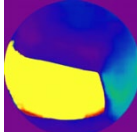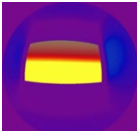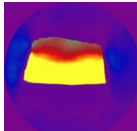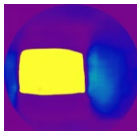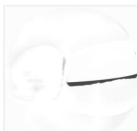
Table 4. Hyperparameters of training.

Parameters	Learning Rate_D	Learning Rate_G	Batch Size	n_epoch	Optimizer
Settings	Initial value: 0.0005 Drop period: 5 Drop rate: 0.3	Initial value: 0.0015 Drop period: 5 Drop rate: 0.3	8	200	Adam

4. Results

The Discriminator converges after around 120 epochs, while the Generator converges after around 140 epochs. Table 5 provides an overview of various cases from this test set, involving 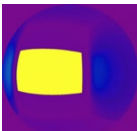the inputs, ground truth, predicted images, and IDA maps. Additionally, Figure 6 illustrates MAE, MAPE, and SSIM for the test set. The training set shows an average MAE of 3.654, an average MAPE of 0.124, and an average SSIM of 0.919. In contrast, the test set has an average MAE of 18.233, an average MAPE of 0.466, and an average SSIM of 0.814.

Table 5. Example results of the test set.

Imgae inputs	Vector inputs		Ground truth	Prediction	IDA
	al_sun:57.231 az_sun:110.561 dir_rad:792	dif_rad:30 az_view:285.109 al_view:89.790			
	al_sun:33.153 az_sun:197.812 dir_rad:753	dif_rad:152 az_view:7.571 al_view:104.526			
	al_sun:55.426 az_sun:110.503 dir_rad:621	dif_rad:200 az_view:20.579 al_view:87.230			
	al_sun:74.489 az_sun:241.523 dir_rad:487	dif_rad:242 az_view:227.846 al_view:99.916			
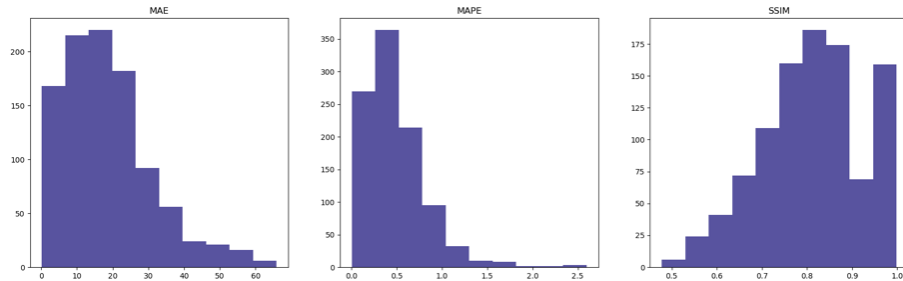
Figure 5. Accuracy indicators of the test set.

5. Discussion

The efficacy of the proposed prediction model has been verified through the case study, and its strengths and limitations deserve further discussion.

1) Time gap with traditional simulation: The proposed model's calculation time involves the time required for capturing the view content, constructing the semantic feature matrix, and generating the predictive image. Constructing the semantic feature matrix is the most time-consuming part. For example, in a 9-viewpoint splice case, the average processing time is about 7.6s, while the other steps take less than 1s. By comparison, simulations for the same dataset range from 276 to 432s, averaging 354s. This model achieves a 97% improvement in speed compared to traditional simulations.

2) Generalizability of the proposed prediction model: This model focuses on

objects within the field of view rather than the entire building. The internal surfaces' types, shapes, and locations are represented as images, thus remaining unaffected by structural changes. Material information, integrated into the images through semantic feature mapping, allows the model to adapt various material setups. The environmental setting is distilled into essential weather parameters, releasing the model from climatic constraints. Overall, this model enhances generalizability in spatial, material, and environmental domains, catering to various design scenarios.

3) Partial sample with poor performance: A few predicted results with a large error. There are two potential reasons. The first is overfitting within mGANs, suggested by the superior performance of the training set over the test set. Despite efforts to enhance model generalizability — like increasing sample size, adding regularization, and using robust layers — overfitting remains a challenge. The second reason might be the partial loss of vector-based feature information during processing. In some low-performance samples, while the surface blocks appear similar, great differences in the luminance distribution of external windows are observed. This could be due to the model's limited ability to learn weather features, attributed to a dimensional gap between vector and image features. Addressing these two issues is crucial for future work in this research.

6. Conclusion

In this paper, a novel rapid predictive model for glare is proposed, designed to quickly visualize glare conditions without limitation from design scenario variations. In this model, building spaces are represented as standardized viewpoint snapshots. Semantic feature mapping is employed to integrate spatial characteristics and material properties. Images containing comprehensive information, along with vectors that reflect environmental conditions, constitute the multimodal input of this model. To fulfill them, mGANs have been developed, focusing on multimodal features. The network integrates feature fusion, feature reinforcement blocks, and advanced upscaling methods, enabling it to effectively eliminate redundant information from multimodal inputs, deeply extract features, and generate predictive images. The model's efficacy is verified through case studies on luminance distribution prediction in residential buildings. Compared to simulation methods, it achieves a 97% improvement in computational speed. Overall, the model provides designers with a faster, more flexible, and intuitive glare visualization method, useful in scenarios such as requiring quick evaluation for the performance of numerous schemes or rapid intuitive understanding of glare conditions. Currently, this model exists as a conceptual prototype. Future research will further refine the overall process and develop related associated tools based on popular design platforms to truly support design practice.

Acknowledgements

References

Al Horr, Y., Arif, M., Kaushik, A., Mazroei, A., Katafygiotou, M., & Elsarrag, E. (2016). Occupant productivity and office indoor environment quality: A review of the literature. Building and environment, 105, 369-389. https://doi.org/10.1016/j.buildenv.2016.06.001

Ayoub, M. (2019). A multivariate regression to predict daylighting and energy consumption of residential buildings within hybrid settlements in hot-desert climates. Indoor and built environment, 28(6), 848-866. https://doi.org/10.1177/1420326X18798164

Giovannini, L., Favoino, F., Verso, V. R. M. L., Serra, V., & Pellegrino, A. (2020). GLANCE (GLare ANnual Classes Evaluation): An approach for a simplified spatial glare evaluation. Building and Environment, 186, 107375. https://doi.org/10.1016/j.buildenv.2020.107375

Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1125-1134).

Jones, N. L. (2019, September). Fast climate-based glare analysis and spatial mapping. In Proceedings of building simulation 2019: 16th Conference of IBPSA.

Jones, N. L., & Reinhart, C. F. (2019). Effects of real-time simulation feedback on design for visual comfort. Journal of Building Performance Simulation, 12(3), 343-361. https://doi.org/10.1080/19401493.2018.1449889

Liu, Y., Colburn, A., & Inanici, M. (2020). Deep neural network approach for annual luminance simulations. Journal of Building Performance Simulation, 13(5), 532-554. https://doi.org/10.1080/19401493.2020.1803404

Luo, Z., Sun, C., Dong, Q., & Qi, X. (2022). Key control variables affecting interior visual comfort for automated louver control in open-plan office--a study using machine learning. Building and Environment, 207, 108565 https://doi.org/10.1016/j.buildenv.2021.108565

Lin, C. H., & Tsay, Y. S. (2021). A metamodel based on intermediary features for daylight performance prediction of facade design. Building and Environment, 206, 108371. https://doi.org/10.1016/j.buildenv.2021.108371

Mostafavi, F., Tahsildoost, M., Zomorodian, Z. S., & Shahrestani, S. S. (2022). An interactive assessment framework for residential space layouts using pix2pix predictive model at the early-stage building design. Smart and Sustainable Built Environment. https://doi.org/10.1108/SASBE-07-2022-0152

Ngarambe, J., Adilkhanova, I., Uwiragiye, B., & Yun, G. Y. (2022). A review on the current usage of machine learning tools for daylighting design and control. Building and Environment, 109507. https://doi.org/10.1016/j.buildenv.2022.109507

Nourkojouri, H., Shafavi, N. S., Tahsildoost, M., & Zomorodian, Z. S. (2021). Development of a machine-learning framework for overall daylight and visual comfort assessment in early design stages. Journal of Daylighting, 8(2), 270-283. https://doi.org/10.15627/jd.2021.21

Pierson, C., Wienold, J., & Bodart, M. (2018). Review of factors influencing discomfort glare perception from daylight. Leukos, 14(3), 111-148. https://doi.org/10.1080/15502724.2018.1428617

Radziszewski, K., & Waczyńska, M. (2018). Machine learning algorithm-based tool and digital framework for substituting daylight simulations in early-stage architectural design evaluation. In Proceedings of the Symposium on Simulation for Architecture and Urban Design (pp. 1-7).

Xie, F., Song, H., & Zhang, H. (2023). Research on Light Comfort of Waiting Hall of High-Speed Railway Station in Cold Region Based on Interpretable Machine Learning. Buildings, 13(4), 1105. https://doi.org/10.3390/buildings13041105

Xie, J., & Sawyer, A. O. (2021). Simulation-assisted data-driven method for glare control with automated shading systems in office buildings. Building and Environment, 196, 107801. https://doi.org/10.1016/j.buildenv.2021.107801

Wu, W., Fu, X. M., Tang, R., Wang, Y., Qi, Y. H., & Liu, L. (2019). Data-driven interior plan generation for residential buildings. ACM Transactions on Graphics (TOG), 38(6), 1-12. https://doi.org/10.1145/3355089.3356556