Enabling Human-Centered Daylighting Operation using Non-Intrusive Luminance Monitoring and Deep Learning

Sichen Lu[1,2*], Dongjun Mah[1,2], Athanasios Tzempelikos[1,2]

[1]Purdue University, Lyles School of Civil Engineering,
550 Stadium Mall Dr., West Lafayette, IN 47906, USA

[2]Center for High Performance Buildings, Ray W. Herrick Laboratories, Purdue University, 140 S. Martin Jischke Dr., West Lafayette, IN 47907, USA

* Corresponding Author: lu448@purdue.edu

ABSTRACT

Luminance monitoring within the occupants' field of view (FOV) is required for assessing visual comfort and overall visual preferences, but it is practically challenging and intrusive. As a result, real-time, human-centered daylighting operation remains a challenge. No studies have determined if it is possible to acquire essential information on how conditions affect preferred luminance distributions, using a camera sensor placed in non-intrusive positions. This paper presents a novel deep-learning based framework method to demonstrate that meaningful features in the visual field can be extracted without invasive measurements or 3-D reconstruction of the occupant FOV. A Conditional Generative Adversarial Network (CGAN), pix2pix is used to transfer information from non-intrusive images to FOV images. Pix2pix takes a condition image (measured from a non-intrusive camera position) and generates a FOV image that is similar to a target image (measured from FOV). Two datasets were collected in an open-plan office with low-cost HDRI cameras installed at two alternate locations (a wall or a monitor), to separately train two pix2pix models with the same target FOV images. The results show that the generated FOV images closely resemble the measured FOV images in terms of pixelwise luminance errors, mean luminance, and structural similarity. The major source of error comes from some bright scenes visible through windows that are absent from non-intrusive images but appear in the FOV. However, more than 85% of the cases in both the training and validation set have low absolute luminance differences, given the inherently high window luminance under sunny conditions. This study is the first proof of concept demonstrating that it is possible to evaluate of visual preferences and enable human-centered daylighting operation without intrusive luminance monitoring, by employing the full potential of HDRI and deep learning techniques.

1. INTRODUCTION

High Dynamic Range Imaging (HDRI) sensors can acquire per-pixel luminance maps that can be efficiently processed to predict general visual comfort metrics but also correlate scene luminance characteristics to subjective responses and extract personal preferences of individuals with respect to the visual environment. Mah and Tzempelikos (2024) proposed a new approach for inferring personal daylight preferences utilizing pixelwise HDRI luminance information from occupants' FOV and deep learning techniques; Convolutional Neural Network (CNN) models trained with luminance and new luminance contrast similarity index maps demonstrated remarkable preference classification accuracy. Evaluation of daylight glare is view-dependent and is based on monitoring luminance distributions from the occupants' FOV as much as possible; however, placing sensors next to occupants' eyes is intrusive and presents practical challenges in daylighting control and operation in buildings.

There are a few recent studies that attempted to use luminance information from a non-intrusive camera (mounted on the ceiling or on side walls) for estimating mean luminance values and Daylight Glare Probability (DGP) observed from FOV, using a linear regression approach or deep learning models (Kruisselbrink et al., 2020; Mentens et al., 2021a & 2021b; Songwa et al., 2021). Alternatively, Kim & Tzempelikos (2021 & 2022) estimated the entire luminance distribution within FOV using photogrammetry techniques. Kruisselbrink et al. (2020) used luminance information from ceiling-mounted cameras to predict mean luminance values of several regions observed from FOV through a linear regression model. Subsequently, a CNN called LumNet was proposed to learn information from HDRI images measured from ceilings for estimating mean desktop luminance (Songwa et al., 2021). Recently, Kim &

Tzempelikos combined Structure-from-Motion and Multi-View-Stereo photogrammetry techniques to project luminance distributions measured from a wall camera to occupants' FOV (2021 & 2022).

Leveraging the prediction power of deep learning techniques, reconstructing luminance distributions within occupants' FOV may not be necessary: theoretically, CNNs trained with images measured from a non-intrusive camera position might still result in high accuracy for classifying visual preferences. To intuitively explain why those non-intrusive images are feasible for predicting visual preferences, it is important to show that luminance distributions measured from non-intrusive viewpoints and FOV are consistent and transferable. An image-to-image translation with Conditional Generative Adversarial Networks (CGANs) may well serve this purpose, and such neural networks may successfully predict pixelwise luminance distributions within FOV based on those measured from a non-intrusive position. Unlike traditional GANs, which generate images from random noise (Goodfellow et al., 2014), CGANs condition on an input image and generate an output image that resembles a target image. Two recent studies have applied pix2pix for daylight performance predictions. Specifically, He et al. (2021) and Li et al. (2024) used CGANs to predict interior illuminance maps based on grayscale grid-based floor maps. Both studies achieved a mean structural similarity index measure (MSSIM) of around 90% between the predicted and target (simulated) illuminance maps.

Although a few studies have explored the potential of using non-intrusive cameras to predict visual information perceived from FOV, none have yet incorporated deep learning to forecast pixelwise luminance values within FOV using such camera setups. This study aims to validate that information from FOV and non-intrusive viewpoints is consistent and transferable via neural networks. Pix2pix, is used to generate luminance maps from FOV based on those captured from non-intrusive camera positions. Two datasets were collected in an open-plan office with low-cost HDRI cameras installed at two alternate locations (a wall or a monitor), to separately train two pix2pix models with the same target FOV images. The images were converted to RGB image for training and validation. The prediction performance of the pix2pix model was evaluated by comparing generated and measured FOV luminance maps using MSSIM and various error metrics across different feature regions: Window, Background, Workplane, and Entire FOV.

2. METHODOLOGY

This study aims to demonstrate that luminance information is transferable between non-intrusive camera positions and occupants' FOV and serves as "proof" to explain why luminance distributions from non-intrusive positions could be used to predict visual preference.

2.1 Pix2pix

Pix2pix (Figure 1) is one of CGANs that takes an image as input and generates an image that is similar to a target image (Isola et al., 2017). Pix2pix has a generator and discriminator which are simultaneously trained and compete against each other to minimize their own loss functions. The generator performs an image-to-image translation where a condition image is transformed to a fake image that is similar to a target image. In this study, the condition images are images measured from a non-intrusive camera position (either from the wall or monitor), the fake (generated) images are predicted FOV images, and the real (target) images are measured FOV images. After training converges, the pix2pix model can be used to predict pixelwise luminance distributions within FOV using images measured from a non-intrusive camera.

2.1.1 Generator

U-net (Ronneberger, 2015) is used as the generator to predict FOV images based on condition images (measured from non-intrusive positions) through a downsampling (Encoder) and upsampling (Decoder) process. The Encoder applies convolutional layers to extract important information from a condition image, where the channel size increases while the size of images reduces; through the Encoder, the condition image (of 3 channels and 256 × 256 pixels, total 196,608 values) is embedded by a compressed spatial representation which consists of 512 channels/values, which is also called an image embedding process. The Decoder uses deconvolutional layers to gradually restore spatial dimensions to produce an output image of the same size as the input image. Skip-connections are applied by directly concatenating outputs from specific layers of the Encoder to outputs of their corresponding layers in the Decoder.

2.1.2 Discriminator

The goal of a discriminator is to correctly identify correct and fake images. Structurally similar to a CNN, the discriminator extracts important information to determine the probability of the input to be real or fake. It receives pairs of images as input: one pair is the real image concatenated to the condition image, and the other pair is the fake image concatenated to the condition image. Both pairs are 6-channel images and are fed to the discriminator separately to evaluate the authenticity of the image given the same condition image. The discriminator uses a PatchGAN

architecture which examines each patch (70 × 70 pixels) within an image and outputs a 2D matrix where each element reflects the probability of each patch being real (=1) or fake (=0) (Isola et al., 2017).

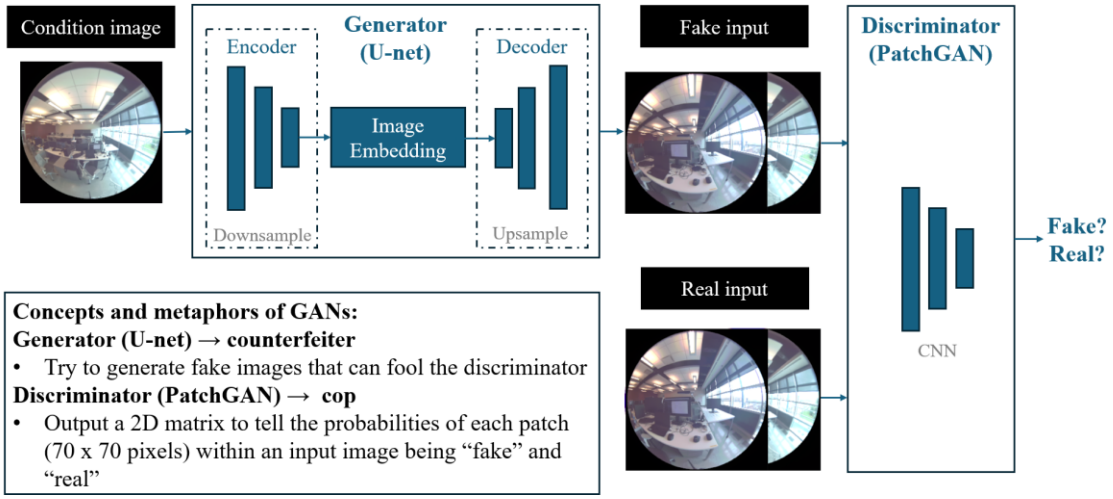

Figure 1: Architecture of pix2pix. Condition image: non-intrusive image; fake input: generated FOV image concatenated with the condition image; real input: measured FOV image concatenated with the condition image.

2.2 Dataset used for model training

In this study, two experimental datasets are created to train separate pix2pix models. Three calibrated Raspberry Pi cameras were placed at different locations in an open plan office with large windows and controllable shades (Figure 5). Each dataset consists of 1,191 pairs of condition images (taken from a specific non-intrusive location) and target images (captured from occupants' FOV). For a consistent comparison, both datasets utilize the same target FOV images. The only difference between the datasets lies in the source of the condition images: one is obtained from the sidewall and the other from the monitor. The two datasets are referred to as the wall dataset and monitor dataset in the following text. Representative images measured from the three camera locations and their corresponding false-color luminance maps are displayed in Figure 2.

HDR images were initially captured in an open office and underwent a series of preprocessing steps to create the datasets. These steps are assigned to prepare the data for effective learning and to ensure smooth convergence of the pix2pix model. The images were collected every 10 minutes, from 9 am to 6 pm over a period of two months, from three different camera positions: two non-intrusive positions (wall and monitor) and one intrusive position (FOV). The original HDR images were first resized to 256 × 256 pixels before converted into 3-channel RGB images through the opencv-python library (OpenCV, 2024). In this study, only images that have RGB values within the standard range of 255 are considered, ensuring that the pix2pix model only trained on scenes with a luminance value less than 45,645 cd/m^2. This study selectively includes images that represent "comfortable scenes" (without excessive luminance levels) to enhance the stability of the learning process. In addition, images captured in dark conditions, such as early morning and late afternoon are excluded due to their different data distributions compared to the typical scenes perceived by occupants. To identify those dark scenes, a mask is applied to the window area of each FOV image; if the maximum luminance detected by the mask is below 100 cd/m^2, images from that timestep are excluded.

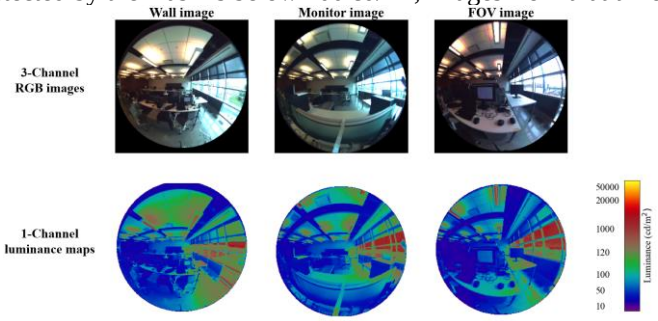

Figure 2 RGB images (top) and their corresponding luminance maps (bottom) measured from 3 camera positions.

2.3. Model training and learning process

2.3.1. Generator and discriminator loss functions and accuracy

The loss functions of the generator and the discriminator are inherently antagonistic and minimized concurrently through updates on weights of the network. The task of the generator is to generate images that are indistinguishable from target images and to fool the discriminator, while the goal of the discriminator is to correctly distinguish between target (real) and generated (fake) images. Therefore, the discriminator and generator follow adversarial training.

The generator uses a structured loss function which panelizes any discrepancy between the structure of the generated image and that of the target one. The loss function of the generator ($Loss_G$) is a weighted sum of two loss functions as shown in equation (1): the L1 loss and a generator adversarial loss (G_{loss}). A higher weight ($\lambda = 100$, equation (1)) is assigned to the L1 loss based on the original pix2pix architecture (Isola et al., 2017). The L1 loss is used to minimize the mean absolute difference on a pixelwise basis between the generated and target image; this loss helps to increase similarity between the generated and targe image, with a focus on achieving accurate reconstruction of pixel values. Additionally, the generator adversarial loss is used to minimize the likelihood of the generated images being recognized as fake by the discriminator; this loss is computed with a cross-entropy loss between the discriminator's output and an array of ones (equation (2)), penalizing predictions that deviate from their true class labels.

$$Loss_G = L1_{loss} \times \lambda + G_{loss} \tag{1}$$

$$G_{loss} = -\frac{1}{N} \sum_{i\_fake\_real=1}^{N} \log\left(p_{i\_fake\_real}\right) \tag{2}$$

where $p_{i\_fake\_real}$ is probability that a patch in a fake image is identified as real by the discriminator. The discriminator aims to correctly classify real and fake images. It uses a cross-entropy loss function, comparing the discriminator output for real input to an array of ones (equation (4)), and its output for fake input to an array of zeros (equation (5)). To balance the training between the generator and the discriminator, a weight of 0.5 is assigned to the loss function of the discriminator, $Loss_D$ (equation (3)). This approach moderates the rate of learning and performance improvements of the discriminator relative to the generator.

$$Loss_D = 0.5 \times (real_{loss} + fake_{loss}) \tag{3}$$

$$real_{loss} = -\frac{1}{N} \sum_{i\_real=1}^{N} \log\left(p_{i\_real}\right) \tag{4}$$

$$fake_{loss} = -\frac{1}{N} \sum_{i\_fake\_fake=1}^{N} \log\left(1 - p_{i\_fake\_fake}\right) \tag{5}$$

where p_{i_real} is probability that a patch in a real image is identified as real by the discriminator, $p_{i\_fake\_fake}$ is probability that a patch in a fake image is identified as fake by the discriminator, and N (=batch size×30×30) is the number of probability output from the discriminator. In the discriminator's binary classification, accuracy for real input is defined as the portion of real input patches correctly identified as real, and similarly for fake input patches. According to PatchGAN, each input image is divided into patches (70×70 pixels), and the discriminator computes a probability for each patch indicating whether it is real (=1) or fake (=0). A classification threshold of 50% is used; if the probability assigned to a patch exceeds this threshold, this patch is classified as real; otherwise, it is classified as fake. The overall accuracy of the discriminator for classifying the entire image is represented by the mean accuracy calculated across all patches (in a 30×30 grid).

2.3.2. Model training parameters, learning process, and convergence

3-channel RGB images are used for training the pix2pix model, and luminance maps are used to evaluate the prediction performance of the model (Figure 2). Each dataset is randomly split into training and validation sets in an 8:2 ratio (953 and 238 pairs of images, respectively), while the two datasets have the same target images in their respective training and validation sets. A batch size of 4 is used for training. Smaller batch sizes can present challenges in terms of training stability and time, but they also provide significant benefits in terms of model robustness and the ability to generalize from limited data. Using a small batch size, the model benefits from more frequent weight updates, which can enhance learning dynamics by preventing premature convergence on a suboptimal solution. Moreover, the increased variability and noise introduced in gradient updates by smaller batches encourage the generator to explore a broader array of weights and potential solutions. Proper data preprocessing (Section 2.2.1) is essential in this context; it helps stabilize the model by enabling it to steadily learn from the patterns in data and adapt to the introduced noise.

Each model is configured to train up to 200 epochs, but an early stop criterion is implemented to terminate training if there are no improvements in the L1 validation loss for 20 consecutive epochs. This approach helps in preventing overfitting issues where the model starts to learn from noise in the training set and ensuring efficient training by not unnecessarily extending the learning process beyond beneficial gains. To observe the learning process and monitoring convergence in the pix2pix model, a few techniques and metrics are used: (1) loss and accuracy monitoring; (2) mean structural similarity image metric (MSSIM) monitoring; (3) visual inspection of generated images.

The Structural Similarity Index Measure (SSIM) is used to assess the similarity between two images (target and generated luminance maps), specifically focusing on their structural aspects rather than just pixel-to-pixel comparisons (Wang et al, 2004). SSIM compares local patterns of pixel intensities that are normalized for luminance and contrast within a local window **x** and **y** of a generated image and target image (11×11 pixels in size) (equation (6)).

$$\boldsymbol{SSIM}(\boldsymbol{x},\boldsymbol{y}) = \frac{(2\mu_x\mu_y + C_1)(2\delta_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\delta_x^2 + \sigma_y^2 + C_2)} \quad (6)$$

where μ_x and μ_y are mean luminance of **x** and **y**, σ_{xy} is the correlation between x and y, σ_x and σ_y are the standard deviation of **x** and **y**, C_1 and C_2 are stability constants determined by the range of the luminance intensity. This comparison is performed across the image using a sliding window, to compute SSIM ranging from 0 to 1, where 1 indicates identical pairs of local patterns, and 0 indicates complete dissimilarity. The mean SSIM (MSSIM) for the entire image is calculated by averaging these local SSIM values, excluding any pixel outside FOV.

The training and validation performance of the generator are presented in terms of the loss functions and MSSIM between the target and generated luminance maps, while only training performance of the discriminator is evaluated. The validation loss and accuracy of the discriminator are not presented, because they might not provide additional useful information beyond the training performance of the discriminator. Essentially, the discriminator is always validating on new, unseen data produced by the generator, making a fixed validation set less meaningful for evaluating the discriminator.

In addition to monitoring the performance metrics, visual inspections of luminance maps (rather than raw 3-channel generated images) are performed to assess improvements and check for quality. For the scope of this study, 3-channel images are converted to luminance maps, where luminance values of each pixel can be estimated with its RGB values following equation (7) (Inanici, 2006). False-color luminance maps of a validation subset which consists of 24 condition and target images and their corresponding generated images are saved if MSSIM exceeds a threshold (=98%). Visual assessment is helpful for checking whether overfitting issues occur and whether the model improves.

$$L = 179 \cdot (R \cdot 0.2126 + G \cdot 0.7152 + B \cdot 0.0722) \quad (7)$$

3. RESULTS

3.1 Learning process of pix2pix

To predict FOV images, the pix2pix model is trained and validated on the wall and monitor dataset separately. Figures 3 and 4 show the training and validation performance of the wall model and monitor model respectively, including (a) loss functions of the generator ($Loss_G$), (b) MSSIM calculated with the target and generated luminance maps, (c) loss functions of the discriminator ($Loss_D$), and (d) classification accuracy of the discriminator.

In the observed training sessions of both models, both the training and validation losses of the generator consistently reduce and stabilize after approximately 50 epochs, while continue to decrease at a diminishing rate. Some hills and bumps are observed in the generator loss, which can be attributed to its composite nature, consisting of the L1 loss and the adversarial loss. The adversarial component is sensitive to the performance of the discriminator, as the discriminator is always evolving and adapting to new input generated by the generator, leading to fluctuations in the overall loss landscape. Both the wall and monitor model can steadily improve validation and training MSSIM and produce images very similar to the target images with a validation MSSIM above 98%. The discriminator's training loss and accuracy provides direct feedback on how challenging it finds to distinguish the generated (fake) image from the target (real) one. The loss functions of the discriminator for correctly recognizing the real (target) and fake (generated) image first increases and oscillates around certain values, where it makes a random guess about whether the input is fake or real. The accuracy of the discriminator reduces to 50%, indicating that it fails to distinguish the fake and real input, which is the objective of the model.

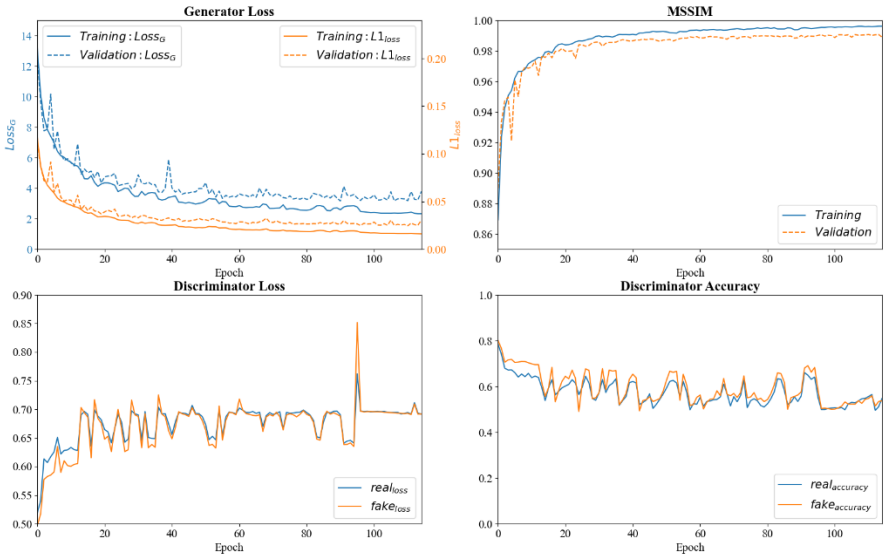

Figure 3 The learning process of the pix2pix model trained with the wall dataset.

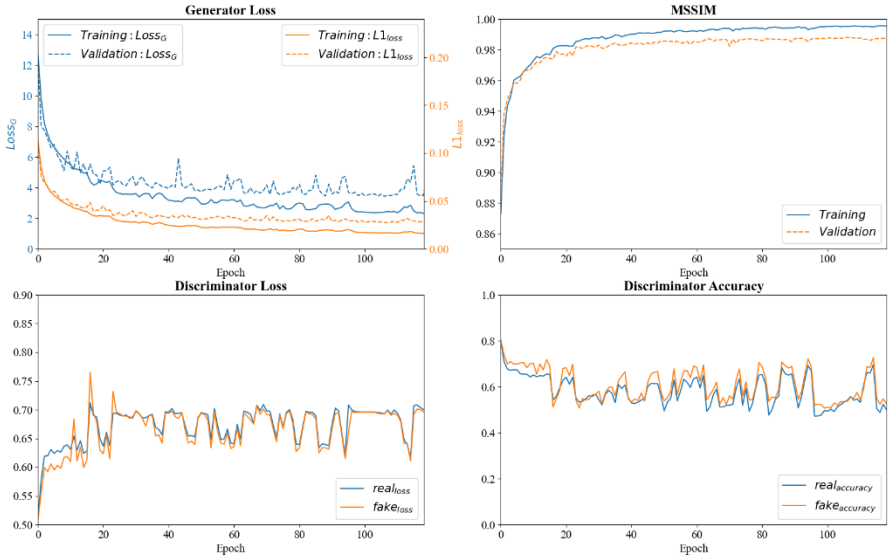

Figure 4 The learning process of the pix2pix model trained with the monitor dataset.

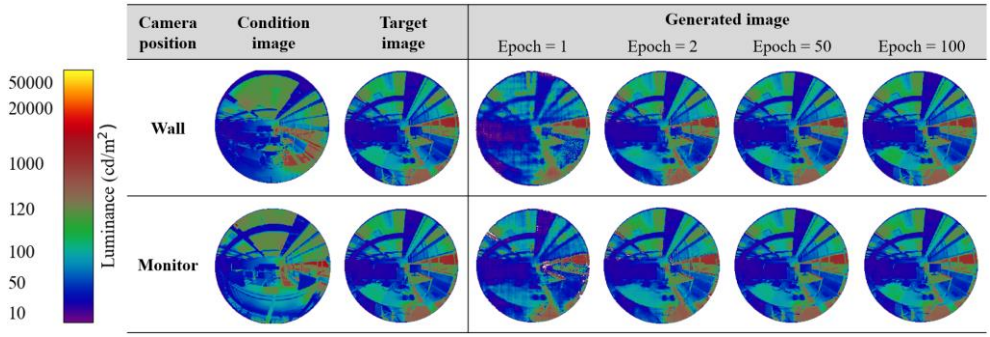

Figure 5 Visual inspections of generated images predicted using a condition image measured from the wall (top) and the monitor (bottom) aiming for the same target FOV image.

Output images from the generator at various stages of the training are recorded for visual inspections and quality check. Figure 5 displays generated images at the end of different epochs, predicted using a condition image (from the

validation set) measured from the wall (top) or the monitor (bottom), aiming for the same target FOV image. After the very first epoch, the model is already capable of roughly predicting the size and intensity of major objects. This capability arises because the model has been updated 238 times, which corresponds to the number of images in the training set divided by the batch size of 4. As training progresses, the model starts to sharpen edges of objects and focus on finer details. By the 50th epoch, generated images are less blurry, and it becomes challenging to visually differentiate the target and generated image. Ultimately, the model learns to accurately predict the intensity of smaller objects, and the generated image looks almost identical to the target one.

3.2 Evaluation of generated images

3.2.1. Luminance differences in target vs generated FOV images

Figure 6 shows a comparison of target and generated FOV images for three representative cases (visual scenes) and their corresponding condition images measured from the wall or the monitor. Absolute difference luminance maps (last column) illustrate errors in the generated image compared to the target counterpart in the range from 30 cd/m^2 up to the maximum luminance difference. This metric is useful because, although the target and generated images appearing nearly identical, there are cases with maximum luminance difference from 777-7314 cd/m^2 focused on the rightmost unshaded portion of the window. For the rest of the scene, the absolute luminance errors are minimal.

To assess the frequency of significant errors and to determine if there is an issue with overfitting, the distribution of maximum absolute luminance differences between all pairs of target and generated images in both the validation and training sets are displayed (Figure 7). Similar distributions of maximum absolute errors are observed across both the validation and training sets from both datasets, suggesting that the notable luminance differences in some cases are not attributable to overfitting. Moreover, over 85% of the cases in both the validation and training sets exhibit an absolute error of less than 2,500 cd/m^2. This indicates that the current pix2pix model and dataset generally perform well across most scenarios.

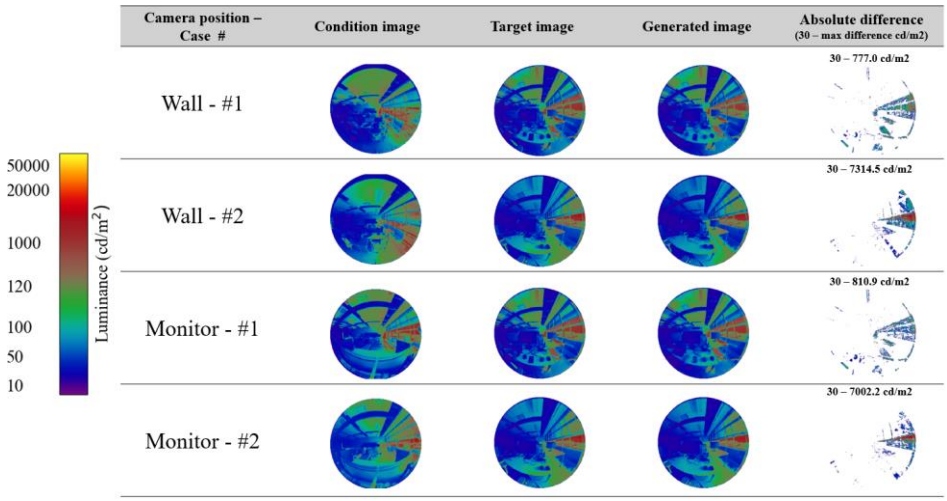

Figure 6 Luminance comparison and absolute difference luminance maps: target vs generated FOV images from the validation set, and their corresponding condition images measured from the wall or the monitor.

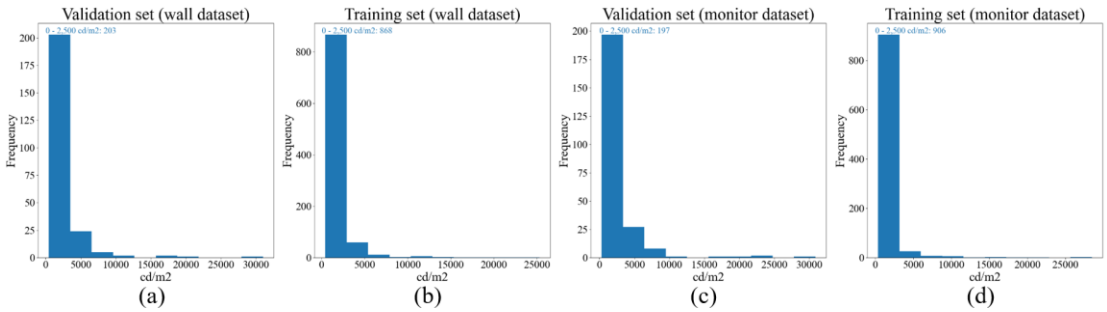

Figure 7 Distributions of maximum absolute luminance difference between target and generated images in the validation and training set from the wall (a, b) and monitor (c, d) dataset.

3.2.2. Errors distribution by feature region in the visual scene

Using each pair of generated (fake) and target (real) luminance maps, error metrics, mean luminance, and MSSIM are calculated for the Overall FOV and further analyzed by feature regions: Window, Background and Workplane (Figure 8). Pixelwise luminance differences between generated and target luminance maps are evaluated by Mean Bias Error (MBE) and Mean Absolute Percentage Error (MAPE) following equations (8-9).

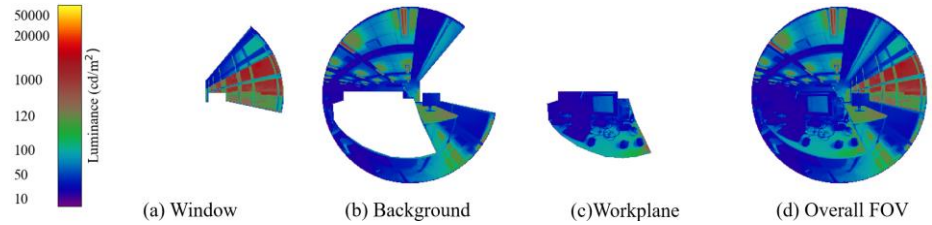

Figure 8 Feature regions: (a) Window, (b) Background, (c) Workplane, and (d) Overall FOV.

$$MBE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i) \tag{8}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}(\frac{\hat{y}_i - y_i}{y_i}) \tag{9}$$

where $\hat{y}_i$ is the pixelwise luminance value of the generated map, y_i is the pixelwise luminance value of the target map, and n is the total number of pixels values in each feature region.

Table 1 summaries mean values of evaluation metrics using the validation and training set from both the wall and monitor model. Figures 9-10 show the prediction performance of validation and training set using the wall, respectively. The distribution of evaluation metrics in the validation and training sets are similar. The number of outliers and mean values of these evaluation metrics show that the training set performs only slightly better than the validation set, indicating that there are no significant overfitting issues.

Table 1 Comparison of generated and target luminance maps: Mean values of MBE, MAPE, and MSSIM.

Dataset	Feature	MBE (cd/m²)		MAPE (%)		MSSIM (%)	
		Validation	Training	Validation	Training	Validation	Training
Wall	Window	-44.2	-38.4	15.7	12.8	95.1	96.7
	Background	-1.5	-1.1	8.3	6.7	99.2	99.6
	Workplane	0.0	0.2	10.0	7.9	99.5	99.7
	Overall FOV	-8.4	-7.1	10.5	8.6	98.6	99.1
Monitor	Window	-18.9	-14.3	15.7	10.6	95.8	98.4
	Background	-0.8	-0.6	8.5	6.8	99.1	99.5
	Workplane	0.2	0.3	11.1	8.9	99.4	99.6
	Overall FOV	-3.7	-2.7	11.3	8.7	98.6	99.4

For the wall dataset, the Window region has a mean MBE of -44.2 cd/m^2 and -38.4 cd/m2 for the validation and training set respectively, indicating that the luminance distributions in the Window region are underestimated, while the MBE of other regions is centered around 0, showing that the prediction for those regions is accurate with no systematic errors. Using the monitor dataset, predicted luminance values of the Window region are closer to their true values compared to the wall dataset, with a lower MBE of -18.9 cd/m^2 and -14.3 cd/m^2 for the validation and training set respectively. Similar conclusions can be drawn from distributions of MAPE. A higher MAPE of the Window region of the validation and training set (from both monitor and wall datasets) shows that it is harder to accurately predict luminance distributions of the Window, while it is relatively easier to predict luminance values of the other regions. Despite some outliers, the average MAPE for the Window region is 15.7% for the validation set of both datasets, while the average MAPE for other regions and Overall FOV is around or below 10%.

The mean luminance values within each feature region of the generated maps and target maps are also shown side by side for comparison. The Background, Workplane, and Overall FOV have similar distributions of mean luminance values, and model predictions are fairly accurate (negligible differences between generated and real values for this metric). The mean luminance values of the Window are slightly underestimated in both the validation and training set.

MSSIM of the Entire FOV in the validation set from both datasets exceeds 98%, indicating that the generated luminance maps closely resemble the target maps in terms of contrast, luminance, and structure similarity. The Window region has a relatively lower MSSIM (but still between 96-98%), while the Background and Workplane region have a MSSIM above 99%.

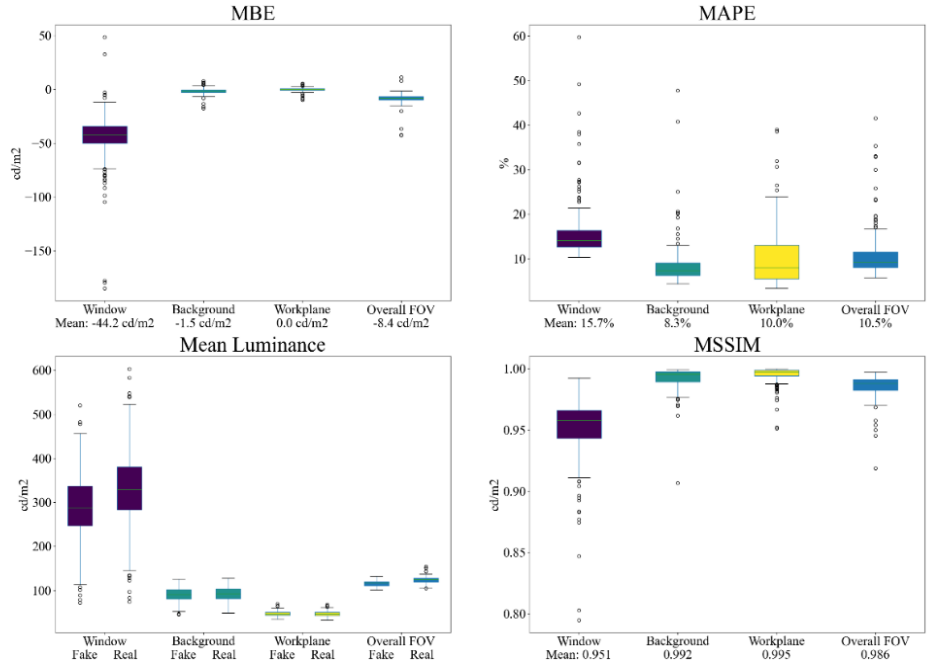

Figure 9 Comparison of generated and target luminance maps breakdown by feature regions: Window, Background, Workplane, and Overall FOV (Case: validation set from the wall dataset)

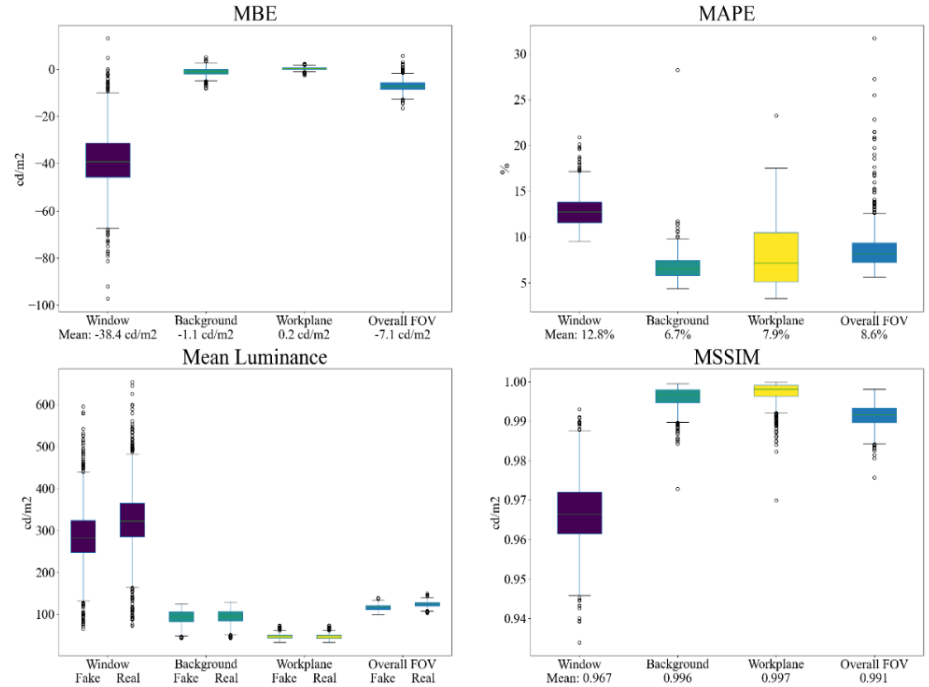

Figure 10 Comparison of generated and target luminance maps breakdown by feature regions: Window, Background, Workplane, and Overall FOV (Case: training set from the wall dataset)

4. CONCLUSION

This study aims to demonstrate that luminance information between FOV and non-intrusive viewpoints is consistent and transferrable through deep learning techniques. Pix2pix, a CGAN, is used to predict pixelwise luminance distributions within FOV based on those measured from a non-intrusive camera position (either from a wall or a monitor). To evaluate prediction performance of the pix2pix model, the generated and measured FOV luminance maps were compared in terms of several error metrics and MSSIM which are further breakdown into different feature regions: Window, Background, Workplane, and Entire FOV. The Entire FOV region of the predicted luminance maps exhibits a remarkable MSSIM of above 98%, an MBE centered around 0, an average MAPE of around 10%, and similar mean luminance values compared to the measured luminance maps; similar and even better prediction performance is observed for the Background and Workplane region. Most luminance errors are observed in the Window region, although more than 85% of the cases result in a small MBE relative to high luminance values present in the Window region. The highest errors occur when exterior brighter light sources are present in FOV but are occluded in non-intrusive positions; the condition images upon which the pix2pix model is based lack such information for accurately predicting luminance distributions within FOV in some cases. Overall, this paper shows that it is feasible to monitor luminance distributions within occupants' FOV using a non-intrusive camera through deep learning neural networks and low-cost camera sensors. This is the first proof of concept demonstrating that it is possible to evaluate visual preferences and enable human-centered daylighting operation without intrusive luminance monitoring.

REFERENCES

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. Advances in neural information processing systems, 27.

He, Q., Li, Z., Gao, W., Chen, H., Wu, X., Cheng, X., & Lin, B. (2021). Predictive models for daylight performance of general floorplans based on CNN and GAN: A proof-of-concept study. Building and Environment, 206, 108346.

Inanici, M. N. (2006). Evaluation of high dynamic range photography as a luminance data acquisition system. Lighting Research & Technology, 38(2), 123-134.

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).

Kim, M., & Tzempelikos, A. (2021). Semi-automated luminance map re-projection via high dynamic range imaging and indoor space 3-D reconstruction. Automation in Construction, 129, 103812.

Kim, M., & Tzempelikos, A. (2022). Performance evaluation of non-intrusive luminance mapping towards human-centered daylighting control. Building and Environment, 213, 108857.

Kruisselbrink, T. W., Dangol, R., & van Loenen, E. J. (2020). Feasibility of ceiling-based luminance distribution measurements. Building and Environment, 172, 106699.

Li, X., Yuan, Y., Liu, G., Han, Z., & Stouffs, R. (2024). A predictive model for daylight performance based on multimodal generative adversarial networks at the early design stage. Energy and Buildings, 305, 113876.

Mah, D., & Tzempelikos, A. (2024). Inferring personal daylighting preferences using HDRI and deep learning techniques, under review.

Mentens, A., Martin, S., Descamps, F., Lataire, J., & Jacobs, V. A. (2021a). Daylight glare probability prediction for an office room. Proc. CIE Midterm Meet.

Mentens, A., Scheir, G. H., Ghysel, Y., Descamps, F., Lataire, J., & Jacobs, V. A. (2021b). Optimizing camera placement for a luminance-based shading control system. In Proceedings of CIE Midterm Meeting.

OpenCV. (2024). Opencv-python 4.9.0.80. https://pypi.org/project/opencv-python/

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18 (pp. 234-241). Springer International Publishing.

Songwa, P. U., Saeed, A., Bhardwaj, S., Kruisselbrink, T. W., & Ozcelebi, T. (2021). LumNet: Learning to Estimate Vertical Visual Field Luminance for Adaptive Lighting Control. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 5(2), 1-20.

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing, 13(4), 600-612.