

# Causally Aware Generative Adversarial Networks for Light Pollution Control

Yuyao Zhang, Ke Guo, Xiao Zhou\*

Gaoling School of Artificial Intelligence, Renmin University of China  
{2020201710,2021201515,xiaozhou}@ruc.edu.cn

## Abstract

Artificial light plays an integral role in modern cities, significantly enhancing human productivity and the efficiency of civilization. However, excessive illumination can lead to light pollution, posing non-negligible threats to economic burdens, ecosystems, and human health. Despite its critical importance, the exploration of its causes remains relatively limited within the field of artificial intelligence, leaving an incomplete understanding of the factors contributing to light pollution and sustainable illumination planning distant. To address this gap, we introduce a novel framework named Causally Aware Generative Adversarial Networks (CAGAN). This innovative approach aims to uncover the fundamental drivers of light pollution within cities and offer intelligent solutions for optimal illumination resource allocation in the context of sustainable urban development. We commence by examining light pollution across 33,593 residential areas in seven global metropolises. Our findings reveal substantial influences on light pollution levels from various building types, notably grasslands, commercial centers and residential buildings as significant contributors. These discovered causal relationships are seamlessly integrated into the generative modeling framework, guiding the process of generating light pollution maps for diverse residential areas. Extensive experiments showcase CAGAN’s potential to inform and guide the implementation of effective strategies to mitigate light pollution. Our code and data are publicly available at [https://github.com/zhangyuuao/Light\\_Pollution\\_CAGAN](https://github.com/zhangyuuao/Light_Pollution_CAGAN).

## Introduction

When Thomas Edison’s electric light first illuminated a street in New York City, it marked the dawn of a modern era intertwined with artificial illumination (Chepesiuk 2009). Since that moment, our world has become saturated with artificial light, transforming environments and enhancing economic productivity. However, the seemingly benign nature of artificial light conceals substantial adverse effects on human health. Nocturnal exposure to artificial light suppresses the synthesis and secretion of pineal melatonin, a hormone crucial for regulating the circadian rhythm (Falchi et al. 2011; Cao, Xu, and Yin 2023; Tåhkämö, Partonen, and Pesonen 2019), leading to an increased risk of

metabolic disorders, cardiovascular issues, and heart disease (Brugger, Marktl, and Herold 1995; Falchi et al. 2011; Cao, Xu, and Yin 2023). Considering the widely acknowledged reality that urban areas serve as primary generators of artificial light, affecting both the adjacent sky and ground (Pothukuchi 2021), coupled with the rapid acceleration of urbanization leading more individuals to inhabit these landscapes, there is a compelling need to scrutinize the intricate implications of artificial light. This underscores the importance of managing and curbing light pollution, presenting a crucial challenge for the sustainable development of contemporary societies worldwide.

To uncover and probe the phenomenon of light pollution within urban areas, several prior studies have delved into this domain. Notably, some investigations utilized satellite-based night-time light (NTL) imagery to quantify and evaluate the extent of light pollution (Zheng et al. 2021; Mu et al. 2021; Cox et al. 2020; Hu and Zhang 2020). Others employed a combination of NTL imagery and Geographic Information System (GIS) techniques to analyze the intricate interplay of light pollution (Butt 2012; Chalkias et al. 2006). Moreover, the distribution of Points of Interest (POIs) in urban settings significantly shapes the occurrence of light pollution. As a valuable data source in urban computing, POIs play a crucial role in tasks such as discovering new venues (Zhou, Mascolo, and Zhao 2019), predicting chronic disease rates (Wang et al. 2018), allocating resources (Zhou et al. 2018), and evaluating investments’ impact on socio-economic indicators (Zhou et al. 2017). Of particular note, a recent study used NTL imagery and POI data to comprehensively assess citywide light pollution (Zhao et al. 2021).

However, existing studies on light pollution exhibit specific limitations. The majority of research tends to superficially quantify light pollution levels in urban areas, often neglecting the underlying causes. Consequently, practical insights for effective real-world policy development are scarce. Additionally, these investigations commonly concentrate on entire cities, overlooking nuanced analyses at smaller spatial scales or specific land use functions (Butt 2012; Chalkias et al. 2006; Zhao et al. 2021). Furthermore, the evaluation metrics employed in these studies are frequently uniform and rudimentary, disregarding the complexities of localized circumstances. For instance, equating illumination levels across different areas with the degree of light

\*Corresponding Author.

Copyright © 2024, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

pollution is illogical, as certain regions may legitimately require higher luminance without causing significant impacts. In summary, the current literature lacks a comprehensive framework that holistically addresses urban light pollution, spanning detection, assessment, and suggestion formulation, while considering solid causal relationships at finely-grained spatial and functional scales.

To bridge this gap, we introduce a novel framework, Causally Aware Generative Adversarial Networks (CAGAN), meticulously crafted to unveil causal relationships between urban areas and light pollution. These relationships subsequently steer the process of generating light pollution maps across diverse residential regions. To provide a comprehensive and objective assessment of light pollution levels, we incorporate three indices: *over illumination* (Mishra 2018), *light trespass* (Schreuder 1986), and *light clutter* (Rajkhowa 2014). By holistically considering these indices, we effectively gauge the intricate impact of light pollution on residential areas. Furthermore, our methodology includes a thorough exploration employing causal inference techniques, which unveil significant effects of different building types on light pollution levels. These identified causal relationships are seamlessly integrated into CAGAN, thereby enhancing the precision of generated light pollution maps in urban areas. We assess light pollution across seven globally significant metropolises, renowned for their robust economies and international importance. Rigorous experiments underscore CAGAN’s capacity to provide valuable insights and aid in the development and implementation of strategies to mitigate the adverse impacts of light pollution.

Our main contributions are summarized as follows:

- We propose an evaluation framework for light pollution that enables comprehensive and objective assessment of light pollution extent in various urban residential areas.
- We investigate the underlying factors contributing to light pollution across metropolises, unveiling causal relationships through advanced causal inference techniques.
- We utilize the identified causal relationships related to light pollution as conditional information to guide the process of generating light pollution maps, which empowers local administrators to make informed decisions and allocate lighting resources effectively.

## Related Work

In this section, we explore three categories of studies closely related to our topic and methodology. These include research on light pollution assessment, applications of causal inference, and the use of Generative Adversarial Networks (GANs) in the urban computing domain.

**Light Pollution Assessment.** The utilization of nighttime light images to evaluate light pollution levels has garnered considerable attention in research. Notably, Zhao et al. (2021) enhanced urban-scale modeling and analysis by integrating POI data into their methodology. Additionally, Tong et al. (2022) and Burt et al. (2023) introduced distinctive indices for assessing light pollution. The ALAN Lab at the University of Hong Kong has conducted extensive academic

investigations into local urban light pollution (HKU ALAN team 2023). In contrast, our research zeroes in on the impact of light pollution on residents in the residential areas of metropolises. We conduct a comprehensive assessment of light pollution levels in these areas, leveraging nighttime light images and POI data.

**Causal Inference in Urban Computing.** The exploration of causal relationships within the realm of social science is crucial, and various methodologies in causal inference have yielded substantial insights in urban computing and analysis (Baum-Snow and Ferreira 2015). Overcoming confounding factors lies at the heart of causal inference, particularly in scenarios with high-dimensional confounders. To address this challenge, Propensity and Propensity Score Matching were introduced by ROSENBAUM and RUBIN (1983). Additionally, Morgan and Winship (2015) proposed the Inverse of Propensity Weighting as an effective solution to confounder interference. The application of the causal inference framework by Zhang et al. (2022b) in individual healthy lifestyle and mobility-related health policymaking further underscores its efficacy. In this study, we employ the Debiased Machine Learning method, as proposed bychetverikov et al. (2016). This method focuses on diminishing the influence of confounding variables, integrating machine learning models, and providing confidence intervals to enhance stability and effectiveness.

**Application of GANs in Urban Computing.** Researchers have harnessed the potential of GANs to create diverse urban environments at various scales. CityGAN, for instance, focuses on synthesizing architectural features and building images (Bachl and Ferreira 2020). SG-GAN, introduced by Li et al. (2018), enhances semantic segmentation on Cityscapes using virtual data. Domain-adaptive networks based on CycleGAN, proposed by Guo et al. (2020), generate urban scenes from virtual video games to aid in segmentation. In a different way, Andrade and Fernandes (2022) employed conditional GANs to synthesize satellite-like urban images from historical maps. The work of Albert et al. (2018) delves into simulating hyper-realistic urban patterns through GANs. Zhang et al. (2022a) developed MetroGAN to utilize NTL images, water area data, and built-up area data to simulate urban morphology. However, our endeavor transcends the task of generating urban visuals or conducting style transfers for urban scenes. We leverage the capabilities of conditional variational auto-encoders to embed the contextual and semantic intricacies associated with light pollution into the generated maps. This infusion of information grants the maps interpretability, elucidating the implications of light pollution’s effects.

## Preliminaries

In this section, we outline methodologies for data collection, processing, index calculation, and constructing an evaluation system for residential light pollution.

## Data Collection and Processing

Before initiating our study, we first selected seven global metropolises that stand out for their scale of urbanization

and extent of economic development. These metropolises include New York, Los Angeles, Paris, London, Shanghai, Beijing, and Guangzhou. We established latitude and longitude boundaries for each metropolis and employed OpenStreetMap<sup>1</sup>'s API to retrieve POI data spanning various land use types. Subsequently, we sampled the corresponding nighttime light intensity value from the remote sensing image of night-time light (Chen et al. 2020) using the longitude and latitude coordinates of each POI. Taking into account the quantity of each land use type and the distribution of night-time light intensity values, we distilled a total of nine distinct POI categories in this study. These encompass brownfield, commercial, construction, farmland, forest, grass, industrial, residential, and retail.

### Light Pollution Evaluation

Following the sampling process, we characterize a residential area within the metropolis as a square region centered around the "residential" land use type POI. This designated residential area is intended to encompass a diverse range of buildings that cater to a variety of residents' needs, with each occupying an approximate area of four square kilometers. Subsequently, we identify the POI situated within this residential zone and partition them into distinct plots, as depicted in Figure 1.



Figure 1: The segmentation of POI in a residential area.

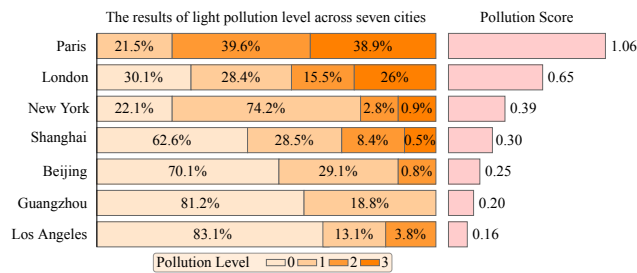


Figure 2: The results of light pollution assessment in all residential areas across the seven metropolises.

In assessing the impact of light pollution sources near residential areas, we drew inspiration from prior research that employed radial basis functions to gauge the extent of light pollution influence (Zhao et al. 2021). This influence is quantified using the following formula:

$$I_j^i = \text{NTL}_j \times \frac{1}{\sqrt{2\pi}D} \times \exp\left(-\frac{d_j^{i2}}{2D^2}\right) \quad (1)$$

where  $I_j^i$  represents the level of light pollution impact on residential area  $i$ 'th from  $j$ 'th pollution source,  $\text{NTL}_j$  represents the night-time light intensity of  $j$ 'th pollution source,  $d_j^i$  represents the distance between the light pollution source and the residential area, and  $D$  is the bandwidth in the radial basis function, which is a hyperparameter with a value of 1500m in this study.

We proceed to employ *Total Nighttime Light* (TNL), *Nighttime Light Disturbance* (NLD), and *Nighttime Light Standard Deviation* (NLSD) to quantify the three indices, namely *over illumination*, *light trespass*, and *light clutter*, as previously discussed (Burt et al. 2023; Tong et al. 2022). In this context, let  $r_i$  represent the  $i$ -th residential area, and  $p_j$  symbolize the  $j$ -th light pollution source, the corresponding light pollution scores can be obtained through:

$$\text{TNL}_i = \sum_{p_j \in \mathcal{C}(r_i)} I_j^i \quad (2)$$

$$\text{NLD}_i = \sum_{p_j \in \mathcal{C}(r_i)} I_j^i - I_i^i \quad (3)$$

$$\text{NLSD}_i = \sqrt{\frac{\sum_{p_j \in \mathcal{C}(r_i)} (I_j^i - \mu_j^i)^2}{|\mathcal{C}(r_i)|}} \quad (4)$$

$$\text{Score}_i = \text{TNL}_i + \text{NLD}_i + \text{NLSD}_i \quad (5)$$

where  $\mathcal{C}(r_i)$  represents the residential area centered on  $r_i$ .

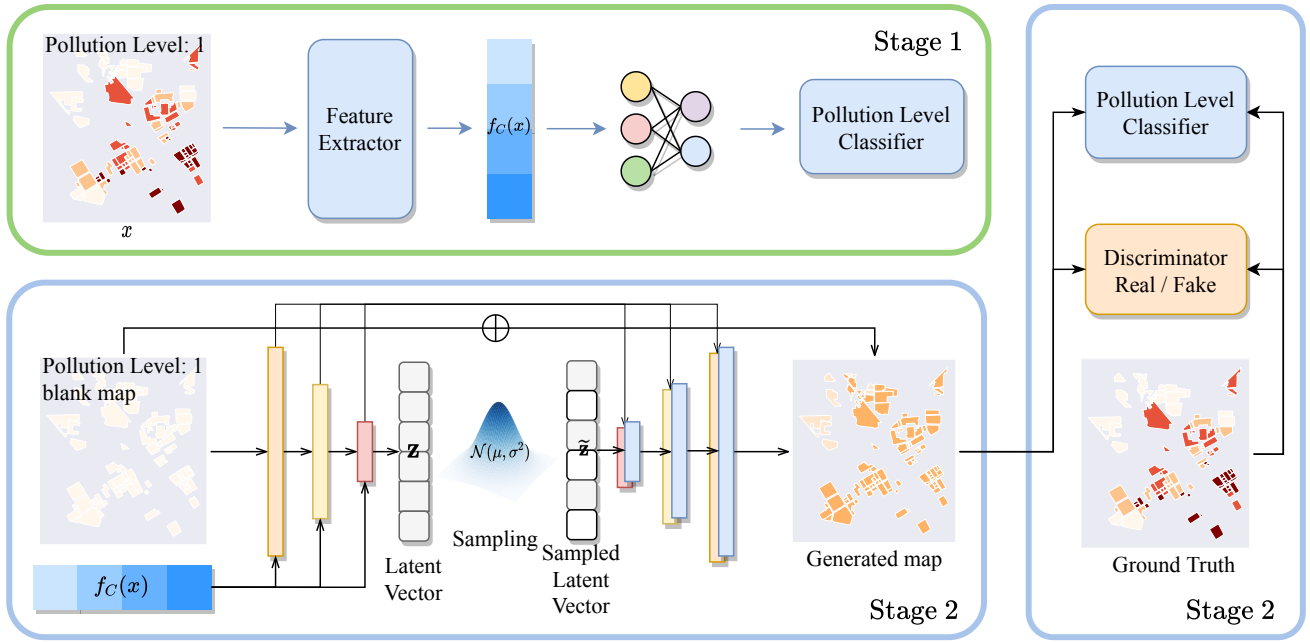
By computing these three light pollution indices, we can assess the influence of light pollution on residential areas from varied perspectives. Subsequently, we employ unsupervised learning algorithms to categorize these areas into five distinct levels of light pollution, discerning which residential regions are prone to enduring significant light pollution and which areas enjoy better conditions. Armed with this insight, we can offer recommendations for addressing severe light pollution in residential zones. This guidance aids local residents in enhancing the rational distribution of light and managing pollution levels around nearby POIs.

We meticulously calculate light pollution indicators for all residential areas within the seven metropolises. Subsequently, we choose a suitable number of centroids and apply the k-means clustering algorithm to divide light pollution levels into four distinct categories. The classification of residential areas in each metropolis, based on their pollution levels, is illustrated in Figure 2. Additionally, the light pollution scores for each city are displayed on the right side of the figure, facilitating a visual comparison of residential light pollution at the urban scale.

### Methodology

In this section, we introduce the Causally Aware Generative Adversarial Networks (CAGAN) framework for generating fine-grained light pollution maps in urban environments while considering causal relationships. The framework is divided into two main phases: 1) estimating the

<sup>1</sup><https://openstreetmap.org>



$f_C(x)$ : the conditional information extracted by a **feature extractor** from given pollution map

Figure 3: The comprehensive architecture of the proposed CAGAN, containing two training stages.

causal effects of light pollution through debiased machine learning, and 2) producing coherent light pollution maps by integrating the obtained conditional causal information into a novel GAN framework. Next, we delve into a comprehensive explanation of the techniques employed in each phase.

### Causal Inference

Causal inference is a scientific method crucial for identifying causal relationships among events or variables. It allows us to determine if one event truly causes another, moving beyond simple correlations or coincidences, by examining patterns and connections in observed phenomena. In our study, applying causal inference helps us investigate the complex links between residential light pollution and various POIs in a specific area, uncovering fundamental causal mechanisms. Through this approach, we make significant contributions to improving and managing light pollution.

We incorporate the causal inference within the latent outcome framework, employing a two-stage Debiased Machine Learning (DML) approach as the cornerstone of our research methodology. DML can be used to estimate treatment effects, particularly in scenarios where all potential confounding variables that exert a direct influence on both the treatment decision and observed outcome are accessible (Chetverikov et al. 2016). To elaborate, let  $Y$  denote the outcome variable encompassing three light pollution indices for each residential area. For each instance, we meticulously designate a specific category of POIs and calculate its quantity, average nighttime light intensity, and mean distance from the center of the residential area, all of which are classified as intervention variables labeled as  $T$ . Addi-

tionally, we account for other variables connected to POIs, denoted as  $X$ , which serve as potential confounders.

The DML simplifies the challenge by initially focusing on estimating two predictive tasks (see Figure 4):

1. Predicting the outcome  $Y$  from the confounders  $X$

$$\hat{Y} = f(X) + \epsilon, \mathbb{E}(\epsilon|X) = 0 \quad (6)$$

2. Predicting the treatment  $T$  from the confounders  $X$

$$\hat{T} = g(X) + \eta, \mathbb{E}(\eta|X) = 0 \quad (7)$$

Furthermore, it holds that  $\mathbb{E}(\epsilon \cdot \eta|X) = 0$ .

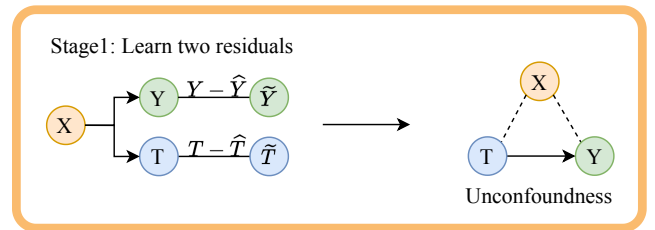


Figure 4: Regression from confounders  $X$  to treatment  $T$  and outcome  $Y$  to achieve unconfoundedness.

Subsequently, we combine these two residuals using predictive models during the final stage of estimation, thereby constructing a model for the Average Treatment Effect (ATE), as illustrated in Figure 5. Assuming that the ultimate

goal is to estimate  $\theta$ , the equation is as follows:

$$\tilde{Y} = Y - f(X) \quad (8)$$

$$\tilde{T} = T - g(X) = \eta \quad (9)$$

$$\tilde{Y} = \theta(X)\tilde{T} + \epsilon \quad (10)$$

Given that  $\mathbb{E}(\epsilon \cdot \eta|X) = 0$ , our objective is to estimate  $\theta(X)$  along with the optimal parameters, representing the desired ATE:

$$\theta^* = \arg \min_{\theta} \mathbb{E}[(\tilde{Y} - \theta(X)\tilde{T})^2] \quad (11)$$

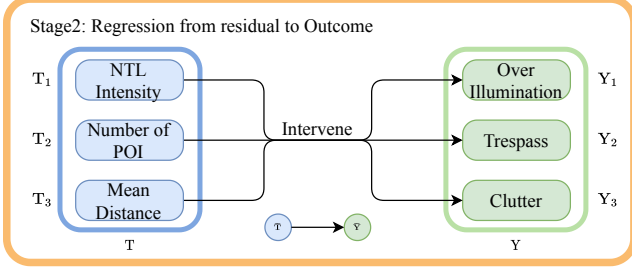


Figure 5: Estimation of Average Treatment Effect.

The novelty of DML resides in the orthogonal relationship between the hyperplane formed by confounders and residuals. Through the regression of the two residuals, we effectively mitigate the confounding bias originating from the confounders on the treatment variables. Thus, we attain an unbiased estimation of the causal effect of the treatment.

### Causally Aware Generative Adversarial Networks

The assessment of authenticity in generated images often concerns resolution and detail. However, for city light pollution maps, this criteria emphasizing high resolution and details may not be crucial. In this context, the emphasis shifts towards generating distinct blocks with clear boundaries.

The proposed Causally Aware Generative Adversarial Networks (CAGAN) comprises two distinct training stages, as illustrated in Figure 3. In the initial stage, a light pollution image classifier is trained with actual light pollution images, enabling it to establish accurate associations with given residential area light pollution maps. Subsequently, a conditional variational autoencoder is trained exclusively using the base map without colorization. The generator is then guided by the classifier and discriminator to progressively generate authentic light pollution maps. Upon completion of both stages, the model is equipped to generate accurate light pollution maps for residential areas based on the relevant conditional information.

Specifically, our variational inference module employs the ELBO (Evidence Lower Bound) loss function, which incorporates reconstruction error based on L1 norm and KL divergence error. The reconstruction error measures the discrepancy between generated images and the original input images. The KL divergence error is used to assess the difference between the generated latent variable distribution and

the prior latent variable distribution. By minimizing the divergence between the generated images and the prior distribution, the generative model learns meaningful latent representations. To be specific,

$$\mathcal{L}_{KL} = \frac{1}{2} (\boldsymbol{\mu}^\top \boldsymbol{\mu} + \text{sum}(\exp(\boldsymbol{\epsilon}) - \boldsymbol{\epsilon} - 1)) \quad (12)$$

$$\mathcal{L}_G = \|\mathbf{x} - \hat{\mathbf{x}}\|_1 \quad (13)$$

$$\mathcal{L}_{ELBO} = \lambda_1 \mathcal{L}_G + \lambda_2 \mathcal{L}_{KL} \quad (14)$$

where  $\boldsymbol{\mu}$  and  $\boldsymbol{\epsilon}$  are mean and covariance of the latent vector output by encoder, and  $\mathbf{x}$  is the original light pollution map.

Furthermore, our approach incorporates the following loss functions: cross-entropy loss, which measures the disparity between the predicted pollution level and the true labels during image classification; binary cross-entropy (BCE) loss, employed to discern the authenticity of the generated maps; and mean squared error (MSE), assessing the difference between the extracted features  $f_C(x)$  of the generated and real maps, as outlined in CVAE-GAN (Bao et al. 2017).

$$\mathcal{L}_C = - \sum_{i=1}^N \sum_{c=1}^C y_{i,c} \log(p_{i,c}) \quad (15)$$

$$\mathcal{L}_D = - \sum_{i=1}^N [y_i \log(p_i) + (1 - y_i) \log(1 - p_i)] \quad (16)$$

$$\mathcal{L}_{GD} = \|\mathbb{E}_{\mathbf{x} \sim P_r} f_D(\mathbf{x}) - \mathbb{E}_{\hat{\mathbf{x}} \sim P_{\hat{\mathbf{x}}}} f_D(\hat{\mathbf{x}})\|_2^2 \quad (17)$$

$$\mathcal{L}_G = \|f_D(\mathbf{x}) - f_D(\hat{\mathbf{x}})\|_2^2 + \|f_C(\mathbf{x}) - f_C(\hat{\mathbf{x}})\|_2^2 \quad (18)$$

Here,  $N$  represents the number of training samples,  $C$  denotes the number of pollution levels,  $y_{i,c}$  is the classification label,  $y_i$  denotes the label for true or fake maps, and  $f_D$  and  $f_C$  represent the features extracted from the input map  $\mathbf{x}$ .

Ultimately, the complete loss function for optimization is:

$$\mathcal{L} = \mathcal{L}_{ELBO} + \lambda_3 \mathcal{L}_C + \lambda_4 \mathcal{L}_D + \lambda_5 \mathcal{L}_{GD} + \lambda_6 \mathcal{L}_G \quad (19)$$

## Experiments

### Experimental Setup

**Data Processing.** After addressing the outliers, we divide the dataset into training, validation, and test sets in a 0.7:0.15:0.15 ratio, yielding 23,514 valid training samples. Additionally, we set various random number seeds to generate different data subsets, and subsequently test and confirm the effectiveness of the training.

**Causal Inference Experiment Setup.** We use a multi-task elastic net in linear models to efficiently fit two residuals. Additionally, we employ a low-dimensional linear regression for estimating the Average Treatment Effect (ATE). Our methodology involves the use of 3-fold cross-validation, setting the maximum number of iterations at 2000.

**CAGAN Experiment Setup.** In the initial stage, a classifier is trained using ground truth light pollution maps with pollution levels as labels. Employing a pre-trained ResNet50 as the backbone network, augmented by a fully connected layer for regressing conditional vectors, we utilize a classification head and the conditional vector for pollution level



classification. Subsequently, in the second stage, the classifier’s parameters are kept fixed, and training proceeds for the generator and discriminator networks. The models are trained using the methodology-specific loss functions. For our experiments, we set hyperparameters as follows:  $\lambda_1 = 1, \lambda_2 = 0.1, \lambda_3 = 1, \lambda_4 = 1, \lambda_5 = 0.001, \lambda_6 = 1$ .

## Experimental Results

**Identifying Contributors to Light Pollution.** We conducted comprehensive experiments covering all residential areas across seven metropolises, meticulously documenting the ATE of various building types on light pollution indices. The outcomes are detailed in Table 1. We categorized the nine building types into three distinct groups: urban living areas (including residential buildings, commercial centers, and retail spaces), urban development areas (encompassing brownfields, construction areas, and industrial zones), and urban landscaping (including grasslands, forests, and farmlands). Comparing their effects allows us to gain valuable insights into light pollution. Due to space constraints, we present results for three Chinese cities only: Beijing, Shanghai, and Guangzhou, as depicted in Figure 6.

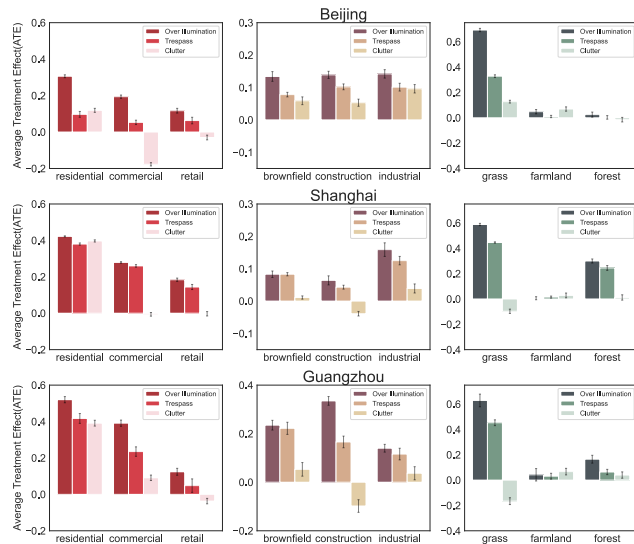


Figure 6: The results of DML in Chinese metropolises.

It is evident that the origins of light pollution vary considerably across different metropolises. Notably, *grasslands, commercial centers, and residential buildings have emerged as noteworthy contributors to light pollution*, consistently exhibiting high ATE values across various metropolises.

**Generating Light Pollution Maps.** We utilize a well-trained model to generate light pollution maps that facilitate the identification of buildings making notable contributions to light pollution for local residents. Furthermore, the model excels at accurately pinpointing sources of light pollution, thereby enabling a judicious allocation of nighttime lighting resources. Figure 7 displays the remarkable performance of CAGAN. The model delineates building boundaries within

residential areas and strategically leverages conditional information to guide the accurate coloring of land blocks.

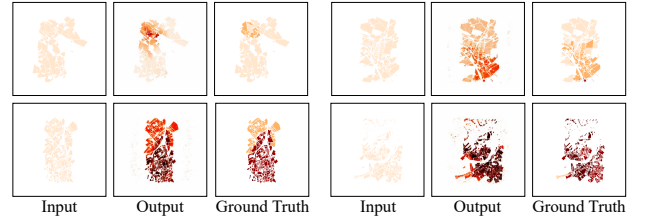


Figure 7: Light pollution maps generated by CAGAN.

The prowess of CAGAN lies in its capacity to produce light pollution maps grounded in causal relationships that may not correspond to reality. For instance, it can simulate the enhancement of an area previously plagued by intense light pollution. This ability can be attributed to two pivotal factors: firstly, the integration of a well-trained classifier that guides the generator’s training, infusing the generated images with both conditional information and interpretability; secondly, the inclusion of a variational inference module that empowers the model to grasp the underlying factors driving light pollution, endowing it with a sense of causality. We have manipulated light pollution levels in residential areas and generated corresponding maps, illustrated in Figure 8.

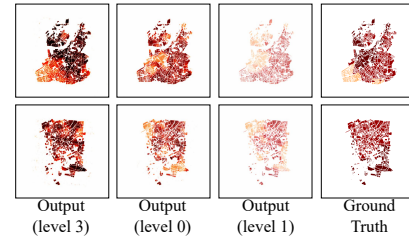


Figure 8: The improvement of severe light pollution.

## Evaluation of Generated Light Pollution Maps

We employ a range of metrics to comprehensively evaluate the quality of our generated light pollution maps, spanning from pixel-level analysis to perceptual assessment. These metrics encompass the Inception Score (IS), Peak Signal-to-Noise Ratio (PSNR), Mean Absolute Error (MAE), Multi-Scale Structural Similarity Index Measure (SSIM), Learned Perceptual Image Patch Similarity (LPIPS), and Relative Average Spectral Error (RASE).

Furthermore, we compute the Kullback-Leibler (KL) divergence to compare the probability distributions of pollution levels between the generated and actual light pollution maps, before classifying them using the softmax function. This comparison also extends to random and uniform distributions. A lower KL divergence value here suggests a higher resemblance in pollution levels between the two maps, emphasizing the semantic information of light pollution. The evaluation results are detailed in Tables 2 and 3.

Indices	City	Residential	Commercial	Retail	Brownfield	Construction	Industrial	Grass	Farmland	Forest
Illumination	Paris	0.318***	0.004	0.016	-0.089*	0.165**	-0.202***	1.414***	-0.026	0.303**
	London	0.243***	0.279***	0.350***	0.001	0.294***	0.038***	0.140***	0.175***	0.238***
	Newyork	0.312***	0.144***	-0.009	0.093**	0.150***	0.008	0.906***	-0.147***	0.082***
	Shanghai	0.422***	0.279***	0.184***	0.083***	0.065***	0.160***	0.573***	0.007	0.298***
	Beijing	0.301***	0.202***	0.120***	0.128***	0.133***	0.133***	0.681***	0.021	0.023
	Guangzhou	0.518***	0.421***	0.130***	0.225***	0.332	0.135***	0.638***	0.045	0.182***
Los Angeles	0.417***	0.145***	0.112***	0.108***	0.052***	0.051***	0.913***	-0.028	0.028***	
Trespass	Paris	0.195**	-0.019	-0.024	-0.139**	0.187**	0.057	0.959***	-0.012	0.325*
	London	0.311***	0.208***	0.294***	-0.009	0.266***	0.024*	0.057***	0.179***	0.244***
	Newyork	0.214***	0.084***	-0.051***	0.037	0.121***	-0.015	0.628***	-0.155***	0.021
	Shanghai	0.380***	0.267***	0.147***	0.085***	0.043***	0.125***	0.431***	0.017**	0.247***
	Beijing	0.107***	0.066***	0.075***	0.071***	0.095***	0.098***	0.360***	-0.003	0.008
	Guangzhou	0.414***	0.265***	0.055***	0.211***	0.170***	0.114***	0.458***	0.038*	0.077***
Los Angeles	0.020	0.330***	0.072**	0.182***	-0.321***	-0.148***	-0.064***	-0.000	0.032	
Clutter	Paris	0.310**	0.110	0.137	-0.018	-0.027	0.243**	0.276*	0.109	0.083
	London	0.290***	-0.019	0.154***	-0.016*	-0.041***	0.010	0.067***	-0.037**	0.044***
	Newyork	0.233***	0.193***	0.087	0.092*	0.113**	-0.015	0.040*	-0.024	-0.093***
	Shanghai	0.399***	-0.007	-0.006	0.013	-0.042***	0.038**	-0.097***	0.026	0.005
	Beijing	0.116***	-0.161***	-0.018	0.051	0.052***	0.091***	0.172***	0.062***	-0.053***
	Guangzhou	0.390***	0.102***	-0.039**	0.047***	-0.095***	0.040	-0.159***	0.068***	0.036
Los Angeles	0.009	0.089***	0.079***	0.051**	-0.146***	-0.068***	-0.070***	0.107**	0.068***	

Table 1: The ATE of different types of buildings on Over illumination, Trespass and Clutter. \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ .

	Pixel Level		Perceptual Level		
	PSNR $\uparrow$	MAE $\downarrow$	SSIM $\uparrow$	LPIPS $\downarrow$	RASE $\downarrow$
IS $\uparrow$	3.253	24.095	0.012	0.941	0.068
					126.953

Table 2: Quality evaluation of CAGAN-generated Maps.

	CAGAN	Random	Uniform
KL $\downarrow$	<b>7.2380</b>	36.1757	28.2567

Table 3: KL Divergence Comparison

As depicted in Tables 2 and 3, our Causally Aware Generative Adversarial Network (CAGAN) exhibits outstanding performance across both pixel-level and perceptual-level metrics, affirming the practicality of the generated light pollution images. Moreover, the generated light pollution maps successfully capture the probability distribution of light pollution levels. Notably, when compared to uniform and random distributions, our model demonstrates a lower Kullback-Leibler divergence, emphasizing its fidelity to the probability distribution observed in real images.

## Discussions

**Causality Interpretability.** The generated images exhibit inherent causal interpretability, facilitating a deeper understanding of the outcomes from causal experiments. For instance, deliberately adjusting the average nighttime light intensity of various building types within residential areas leads to observable changes in brightness across different regions. By adjusting the NTL values for Residential, Commercial, Construction, and Grassland categories and subsequently regenerating the light pollution maps, as illustrated in Figure 9, our model showcases its capacity to assist local administrators in making more informed decisions regarding the allocation of nighttime lighting resources.

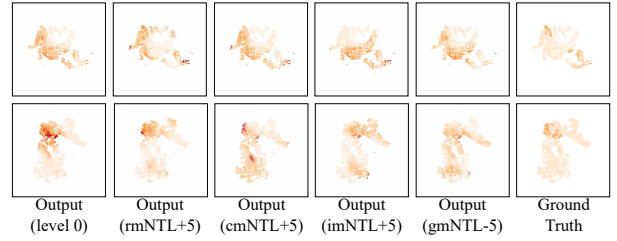


Figure 9: The residential areas after altering the Average NTL (Nighttime Light Intensity) of different buildings.

**Further Insights.** Considering both the generated images and causal outcomes in a holistic manner allows us to elucidate and dissect intricate light pollution phenomena. For example, commercial centers often exhibit continuous nighttime illumination, encompassing buildings, billboards, and lighting clutter. Additionally, in economically developed metropolises, densely populated residential zones often experience mutual light trespass. Furthermore, alterations in NTL values within grasslands yield broader effects. Urban landscaping areas like grasslands require nighttime illumination for visual enhancement. This necessity might lead to increased lighting installations around such spaces, thus contributing to elevated light pollution. Additionally, grassy areas commonly lack tall buildings and trees, rendering the scattering and reflection of light in the surrounding environment more pronounced, thereby influencing a larger area.

## Conclusion

In a nutshell, the proposed Causally Aware Generative Adversarial Network (CAGAN) introduces a novel approach to enhance interpretability and formulate management strategies for mitigating light pollution in urban residential areas. Through the exploration of causal relationships associated with light pollution in our framework, we have gained a

deeper understanding of this phenomenon, paving the way for the implementation of effective mitigation strategies. The investigation has demonstrated the efficacy of CAGAN in assessing residential light pollution across seven economically advanced metropolises. Looking forward, this method holds the potential for broader applications, covering different categories and levels of economic development.

### Acknowledgement

This work was supported by the National Natural Science Foundation of China (NSFC Grant No.62106274); the Fundamental Research Funds for the Central Universities, Renmin University of China (No.22XNKJ24). We also wish to acknowledge the support provided by the Intelligent Social Governance Platform, Major Innovation & Planning Interdisciplinary Platform for the "Double-First Class" Initiative.

### References

- Albert, A.; Strano, E.; Kaur, J.; and González, M. 2018. Modeling Urbanization Patterns with Generative Adversarial Networks. In *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, 2095–2098.
- Andrade, H. J. A.; and Fernandes, B. J. T. 2022. Synthesis of Satellite-Like Urban Images From Historical Maps Using Conditional GAN. *IEEE Geoscience and Remote Sensing Letters*, 19: 1–4.
- Bachl, M.; and Ferreira, D. C. 2020. City-GAN: Learning architectural styles using a custom Conditional GAN architecture. arXiv:1907.05280.
- Bao, J.; Chen, D.; Wen, F.; Li, H.; and Hua, G. 2017. CVAE-GAN: Fine-Grained Image Generation through Asymmetric Training. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*, 2764–2773. IEEE Computer Society.
- Baum-Snow, N.; and Ferreira, F. 2015. Chapter 1 - Causal Inference in Urban and Regional Economics. In Duranton, G.; Henderson, J. V.; and Strange, W. C., eds., *Handbook of Regional and Urban Economics*, volume 5 of *Handbook of Regional and Urban Economics*, 3–68. Elsevier.
- Brugger, P.; Marktl, W.; and Herold, M. 1995. Impaired nocturnal secretion of melatonin in coronary heart disease. *The Lancet*, 345(8962): 1408.
- Burt, C. S.; Kelly, J. F.; Trankina, G. E.; Silva, C. L.; Khalighifar, A.; Jenkins-Smith, H. C.; Fox, A. S.; Fristrup, K. M.; and Horton, K. G. 2023. The effects of light pollution on migratory animal behavior. *Trends in Ecology & Evolution*.
- Butt, M. J. 2012. Estimation of light pollution using satellite remote sensing and geographic information system techniques. *GIScience & Remote Sensing*, 49(4): 609–621.
- Cao, M.; Xu, T.; and Yin, D. 2023. Understanding light pollution: Recent advances on its health threats and regulations. *Journal of Environmental Sciences*, 127: 589–602.
- Chalkias, C.; Petrakis, M.; Psiloglou, B.; and Lianou, M. 2006. Modelling of light pollution in suburban areas using remotely sensed imagery and GIS. *Journal of environmental management*, 79(1): 57–63.
- Chen, Z.; Yu, B.; Yang, C.; Zhou, Y.; Yao, S.; Qian, X.; Wang, C.; Wu, B.; and Wu, J. 2020. An extended time-series (2000-2018) of global NPP-VIIRS-like nighttime light data.
- Chepesiuk, R. 2009. Missing the Dark: Health Effects of Light Pollution. *Environmental Health Perspectives*, 117(1): A20–A27.
- Chetverikov, D.; Demirer, M.; Duflo, E.; Hansen, C.; Newey, W.; and Chernozhukov, V. 2016. Double machine learning for treatment and causal parameters. 2016.
- Cox, D. T.; Sánchez de Miguel, A.; Dzurjak, S. A.; Bennie, J.; and Gaston, K. J. 2020. National scale spatial variation in artificial light at night. *Remote Sensing*, 12(10): 1591.
- Falchi, F.; Cinzano, P.; Elvidge, C. D.; Keith, D. M.; and Haim, A. 2011. Limiting the impact of light pollution on human health, environment and stellar visibility. *Journal of Environmental Management*, 92(10): 2714–2722.
- Guo, X.; Wang, Z.; Yang, Q.; Lv, W.; Liu, X.; Wu, Q.; and Huang, J. 2020. GAN-Based virtual-to-real image translation for urban scene semantic segmentation. *Neurocomputing*, 394: 127–135.
- HKU ALAN team. 2023. Artificial Light at Night (ALAN) Research Literature Database. Available at <https://nightsky.physics.hku.hk/en-hk/light-pollution-resources/reference-materials/articles-papers>. Accessed: 2nd August 2023.
- Hu, Y.; and Zhang, Y. 2020. Global nighttime light change from 1992 to 2017: Brighter and more uniform. *Sustainability*, 12(12): 4905.
- Li, P.; Liang, X.; Jia, D.; and Xing, E. P. 2018. Semantic-aware Grad-GAN for Virtual-to-Real Urban Scene Adaptation. arXiv:1801.01726.
- Mishra, S. P. 2018. Photoperiodic biodiversities under light pollution in India during Anthropocene epoch. *International Journal of Advanced Research (IJAR)*, 6(2): 1090–1106.
- Morgan, S. L.; and Winship, C. 2015. *Counterfactuals and causal inference*. Cambridge University Press.
- Mu, H.; Li, X.; Du, X.; Huang, J.; Su, W.; Hu, T.; Wen, Y.; Yin, P.; Han, Y.; and Xue, F. 2021. Evaluation of light pollution in global protected areas from 1992 to 2018. *Remote Sensing*, 13(9): 1849.
- Pothukuchi, K. 2021. City Light or Star Bright: A Review of Urban Light Pollution, Impacts, and Planning Implications. *Journal of Planning Literature*, 36(2): 155–169.
- Rajkhowa, R. 2014. Light pollution and impact of light pollution. *international Journal of Science and Research (IJSR)*, 3(10): 861–867.
- ROSENBAUM, P. R.; and RUBIN, D. B. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1): 41–55.
- Schreuder, D. 1986. Light trespass: Causes, remedies and actions. *Transport*, 22: 23.
- Tong, J. C.; Lau, E. S.; Hui, M. C.; Kwong, E.; White, M. E.; and Lau, A. P. 2022. Light pollution spatial impact assessment in Hong Kong: Measurement and numerical modelling on commercial lights at street level. *Science of The Total Environment*, 837: 155681.



Tähkämö, L.; Partonen, T.; and Pesonen, A.-K. 2019. Systematic review of light exposure impact on human circadian rhythm. *Chronobiology International*, 36(2): 151–170. PMID: 30311830.

Wang, Y.; Zhou, X.; Mascolo, C.; Noulas, A.; Xie, X.; and Liu, Q. 2018. Predicting the Spatio-Temporal Evolution of Chronic Diseases in Population with Human Mobility Data. *IJCAI*.

Zhang, W.; Ma, Y.; Zhu, D.; Dong, L.; and Liu, Y. 2022a. MetroGAN: Simulating Urban Morphology with Generative Adversarial Network. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '22, 2482–2492. New York, NY, USA: Association for Computing Machinery. ISBN 9781450393850.

Zhang, Y.; Xu, F.; Xia, T.; and Li, Y. 2022b. Quantifying the Causal Effect of Individual Mobility on Health Status in Urban Space. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 5(4).

Zhao, F.; Chu, C.; Liu, R.; Peng, Z.; Du, Q.; Xie, Z.; Sun, Z.; Zeng, H.; and Xia, J. 2021. Assessing Light Pollution Using POI and Luojia1-01 Night-Time Imagery From a Quantitative Perspective at City Scale. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14: 7544–7556.

Zheng, Z.; Wu, Z.; Chen, Y.; Guo, G.; Cao, Z.; Yang, Z.; and Marinello, F. 2021. Africa's protected areas are brightening at night: A long-term light pollution monitor based on nighttime light imagery. *Global Environmental Change*, 69: 102318.

Zhou, X.; Hristova, D.; Noulas, A.; Mascolo, C.; and Sklar, M. 2017. Cultural investment and urban socio-economic development: a geosocial network approach. *Royal Society open science*, 4(9): 170413.

Zhou, X.; Mascolo, C.; and Zhao, Z. 2019. Topic-enhanced memory networks for personalised point-of-interest recommendation. In *Proceedings of the 25th ACM SIGKDD International conference on knowledge discovery & data mining*, 3018–3028.

Zhou, X.; Noulas, A.; Mascolo, C.; and Zhao, Z. 2018. Discovering latent patterns of urban cultural interactions in wechat for modern city planning. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, 1069–1078.