Halation Based Nighttime PM$_{2.5}$ Estimation

Yun Xiang, *Member, IEEE*, Kaihua Zhang, Tony Zhang, *Graduate Student Member, IEEE*, Zuohui Chen,
Qi Xuan, *Member, IEEE*, Robert P. Dick, *Senior Member, IEEE*

Abstract— Camera-based inference techniques can be used to estimate PM$_{2.5}$ concentrations in air based on the aggregate effects of particles on light scattering and absorption. These techniques can be spatially fine-grained, operate in real time, and substantially improve accuracy compared to particle counting sensors. However, existing camera-based techniques fail at night, when pollution exposure and production remain important. We describe the first vision-based technique for nighttime PM$_{2.5}$ concentration estimation. The design approach differs substantially from that of daytime systems because the primary source of daytime information, progression of color toward "airlight" color with increasing depth, is much less useful at night and the primary source of night-time information, the glowing halation regions around artificial light sources, is insignificant during the day. We describe a night-time pollution estimation technique that builds upon novel "Illumination Map" feature. We describe an Illumination Map based Dual-Channel Squeeze-and-Excitation Convolutional Neural Network (DSECNet) is to estimate PM$_{2.5}$ concentrations. This method is evaluated on real-world data and images and outperforms the most advanced related existing (daytime) haze estimation methods, achieving a mean absolute error of 8.65μg/m^3, which is 16.99% lower than the state-of-art baseline method. To the best of our knowledge, this is the first vision-based nighttime nighttime PM$_{2.5}$ estimation method.

Index Terms— Air Quality Estimation, Dataset, Glow Effect, Nighttime Image, PM$_{2.5}$ Concentration.
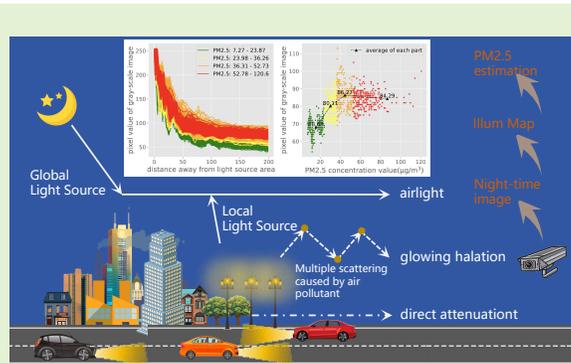
I. INTRODUCTION

PM$_{2.5}$, or ambient fine particulate matter, is an important pollutant consisting of particles with diameters less than 2.5 μm. It is a class I carcinogen certified by the World Health Organization (WHO) – a great threat to human health [1]. Long-term exposure to high concentrations of PM$_{2.5}$ damage the cardiovascular and respiratory systems, leading to respiratory disease, heart disease, stroke, and many other health problems. Long-term exposure to high PM$_{2.5}$ concentrations is correlated with increased mortality rate and reduced life expectancy [2]–[4], hence the importance of systems for large-scale PM$_{2.5}$ monitoring [5], [6].

Researchers have long relied on the laboratory results based on weighing, the micro-oscillation balance method, or the β-ray method [7] to measure PM$_{2.5}$ concentrations. Although these method are accurate, they have high maintenance costs and limited coverage, because each sensor can only measure concentration at one location [8]. Self-contained light scattering based air quality sensor networks have similar cost limitations [9]. Moreover, remote sensing technology is used for urban monitoring PM$_{2.5}$ estimation but is expensive and requires clear weather conditions [10], [11], e.g., cloud cover interferes with its use. In contrast, ground-deployed cameras are inexpensive and widely available and enable for near-ground PM2.5 monitoring. The scattering and absorption of light due to atmospheric pollutants influences captured images, which can be used to estimate pollutant concentrations. There are several vision-based techniques for daytime air quality estimation that use spatial contrast, dark channel features, and variations in sky-earth colors [12], [13] for pollutant concentration estimation. However, these techniques fail at night.

At night, global illumination is low, yet the camera's noise floor remains constant, pushing pollution-relevant signals below the noise floor. Key image attributes commonly used in daytime algorithms—such as saturation and brightness, which are largely determined by sunlight—become unreliable at night. For example, objects close to the artificial light source will show significantly increased saturation and brightness at night. Additionally, flare and halation caused by artificial light sources require transmission models that are highly spatially heterogeneous compared to those sufficient in daytime systems (see Section III for details). These discrepancies render con-

ventional daytime pollution estimation algorithms ineffective in nighttime environments.

To address the nighttime vision-based $PM_{2.5}$ monitoring problem, we design a technique that uses halation [14], the glow around (possibly artificial) lights, to estimate $PM_{2.5}$ concentration at night. Halation is common in night-time images of populated regions, where $PM_{2.5}$ influences the most people. Halation results from photons scattering off particles near the light source one or more times, appearing (from the camera's perspective) to come from glowing region surrounding the light source. The particle concentrations are correlated to the intensity of halation, making it possible to estimate $PM_{2.5}$ concentrations by scattered-light intensity (pixel value). Other features, such as contrast and dark channel, the patch-wise minimum of all color channels, depend on scene visualization and are undermined by noise at night, changes in atmospheric light (moonlight, cloud), and varying artificial light sources.

Specifically, we propose using features related to medium transmission and luminance attenuation from the source image, combining them to generate Illumination Map. We designed a Dual-Channel Squeeze-and-Excitation Convolutional Neural Network (DSECNet) for nighttime Illumination Map based pollution estimation. It consists of dual-channel parallel convolutional networks that extract features separately. A channel attention mechanism then uses raw image features as channel weights to improve accuracy.

The contributions of this work follow.

1) This is the first work solving the problem of vision-based night-time $PM_{2.5}$ estimation. The key idea behind the proposed method is the Illumination Map, which combines scene radiance transmission and light source estimation from a single image and uses this feature to predict the concentration.

2) We describe a Dual-Channel Squeeze-and-Excitation Convolutional Neural Network (DSECNet) to extract and fuse information from the raw image and the Illumination Map. It does not require light source brigness estimates, making the automated application of this technique practical in real-world scenarios.

3) We collected a dataset containing 11,753 multi-location images in night scenes with corresponding environment parameters, including $PM_{2.5}$, PM_{10}, temperature, and humidity. This dataset is of high temporal resolution and has been made publicly available to other researchers.

Experimental results show that, in comparison with the most advanced related work (which was designed for daytime use), our method achieves highest $PM_{2.5}$ accuracy, with an error of $8.65\,\mu g/m^3$. This improves on the second-best method, Retrained PE-MobileNetV2 [15], by 16.99%.

The rest of this paper is organized as follows. Section II presents related work. Section III describes details the proposed method. Section IV describes the data collection and analysis process. Section V presents the experimental results. Section VI discusses the potential limitations and concludes this paper.

II. RELATED WORK

Related work can be placed into four categories: $PM_{2.5}$ monitoring, vision-based air quality estimation, remote sensing based $PM_{2.5}$ estimation, and image dehazing.

A. $PM_{2.5}$ Monitoring

Filter-based air quality sensors can accurately measure $PM_{2.5}$ concentration. However, their deployment can be time-consuming, which limits coverage [16], [17]. Moreover, $PM_{2.5}$ concentration varies dramatically over time and space [18]. $PM_{2.5}$ concentration can vary up to $10\,\mu g/m^3$ within a 10-minute interval [19]. Therefore, long-term, large-scale, and real-time monitoring is impractical for this type of sensor.

For large-scale fine-grained monitoring, existing sensor network based approaches use interpolation algorithms to fuse multi-point environmental information, estimating values at non-deployed locations or during non-monitoring periods. For example, Krishan et al. consider the spatial diffusion and long-term dependence of pollutant concentration and develop an air quality prediction model based on long short-term memory (LSTM) [20]. Ma et al. propose a bidirectional LSTM (BLSTM) model for air quality prediction [21]. Guo et al. propose an unsupervised $PM_{2.5}$ estimation method using a time distributed convolutional gated recurrent unit (TCGRU) and k-nearest neighbor inverse distance weighted (KIDW) interpolation to monitor areas without air monitoring stations [22]. Zhang et al. propose a CNN-LSTM hybrid network to model the spatio-temporal correlations between haze images and $PM_{2.5}$ concentrations [23]. It uses multi-level attention to forecast $PM_{2.5}$ concentration. Although these methods improve the spatial or temporal resolution of estimation, they are not applicable to areas without monitoring station deployment.

B. Vision-Based Air Quality Estimation

Vision-based estimation methods have much higher spatial resolutions than particle counters. There are a variety of open datasets available for air quality measurements, but images are all captured during the daytime, not at night [15], [24]–[26].

Zhang et al. describe a method that uses scattering and absorption features for the concurrent estimation of multiple pollutants [27]. Su et al. describe an end-to-end CNN to estimate multiple atmospheric environmental parameters [28]. Wang et al. developed a dual-channel air quality measurement method that combines RGB frame and corresponding semantic segmentation video frames [13]. Yang et al. designed ImgSensingNet, a vision-guided aerial-ground sensing system that consists of unmanned aerial vehicles (UAVs) and a ground sensor network. It combines vision-based air quality monitoring and networked terrestial point sensors to improve accuracy [12].

Fang et al. [15] designed a prior-enhanced (PE) framework that learns from both the input image and its associated prior maps, including the dark channel (DC) and inverted saturation (IS). Utomo et al. [29] predict the particle concentration based on multi-modal information, including the depth domain and texture domain. However, the feature extractor is trained on clear images with sufficient airlight, making it unsuitable for

night scenes with insufficient and uneven lighting. Wang et al. [30] and Yang et al. [12] fused multi-image feature in their work to predict PM$_{2.5}$ concentration. These vision-based haze estimation methods are built upon designed or learning features, with a focus on image structure loss, chroma, and scene depth. However, existing methods are designed for daytime images, and are unsuitable for nighttime pollution estimation because they assume and require bright global illumination, which does not exist at night. In contrast, our proposed method makes no such assumption and automatically uses natural and artificial light sources of varying sizes, intensities, and colors for pollution estimation. It focuses on the impact of pollutant concentration on halation effects near (artificial) light sources.

C. Remote Sensing Based PM$_{2.5}$ Estimation

Remote sensing technology is widely used in urban monitoring. Advanced technologies such as model-data coupling and semantic segmentation are applied to improve performance and improve generalization capabilities [31], [32]. In contrast, our method is expected to achieve easy-to-deploy monitoring of near-ground PM2.5 concentrations using an inexpensive, ground-based camera monitoring.

Recent works on nighttime PM$_{2.5}$ concentration estimation are based on remote sensing techniques. They rely on homogeneous lighting at night, however, they can not be applied in areas that are not consistently and well lit. Weng et al. propose to use thermal channel data and aerosol light absorption to estimate PM$_{2.5}$ [10]. Wang et al. analyze the relationship between nighttime light radiance, meteorological elements, and topographic elements. Then they use multiple linear regression and random forest methods to develop seasonal and annual PM$_{2.5}$ concentration estimation models [11]. However, satellite remote sensing techniques are limited by cost and require the absence of cloud cover. Our method is expected to achieve fine-grained monitoring of near-ground PM$_{2.5}$ concentrations using an inexpensive, ground-based camera monitoring.

D. Image Dehazing

Image dehazing and vision-based air quality estimation techniques are closely related since they both process haze effect in images. Image dehazing aims to remove the haze effect and enhance image quality, while air quality estimation quantifies the magnitude of the haze effect.

He et al. propose to use the dark channel prior and global atmospheric light to estimate transmission in the haze images [33]. However, although the dark channel is widely used for daytime air quality estimation [34], [35], it is ineffective for nighttime images because its premises do not hold. He et al. attribute the low intensity in the dark channel to visible factors, such as colorful objects, dark surfaces, or shadows. In contrast, at night low intensituy is generally due to low illumination. Li et al. propose to separate the glow from nighttime images based on a smoothness prior [36]. However, this method is prone to noise and color shift problems in the resulting feature maps, reducing accuracy. In contrast, our nighttime PM$_{2.5}$ estimation technique is based on the novel concept of Illumination Map (defined in Section III), which

enables improvements of 10.10% in MAE and 13.36% in MRE compared with other feature maps, e.g., dark channel, He's transmission, and Li's glow feature.

III. METHODOLOGY

In this section, we first present our nighttime haze imaging model. Then we describe how to extract halation-related features and the corresponding model to estimate PM$_{2.5}$ concentration. The flow of our system is shown in Fig. 1.

A. Nighttime Haze Imaging Model

The nighttime haze imaging model can be viewed as a generalization and extension of the optical daytime model [37], where the observed intensity at pixel x is modeled as a linear combination of the direct attenuation [38] $\mathcal{D}(x)$ and the airlight [38] $\mathcal{A}(x)$ as follows:

$$\begin{aligned} \mathcal{I}(x) &= \mathcal{D}(x) + \mathcal{A}(x) \\ &= \mathcal{J}(x)t(x) + \boldsymbol{A}[1 - t(x)], \end{aligned} \quad (1)$$

where $\mathcal{I}(x)$ is the observed intensity at pixel x, $\mathcal{J}(x)$ is the scene radiance assuming pollution-free air, and $\boldsymbol{A}$ is the atmospheric light constant. $t(x)$ is the medium transmission that indicates the portion of scenes reaching the camera. It is defined as follows:

$$t(x) = e^{-\beta d(x)}, \quad (2)$$

where β is the atmosphere scattering coefficient and d is the scene depth. $\mathcal{J}(x)t(x)$ represents the perceived scene radiance after attenuation, i.e., direct attenuation. $\mathcal{A}(x)$ is the airlight indicating the particle veil induced by the atmospheric light scattering, which varies with location. In daytime, when the regional atmospheric light derived from artificial light sources is insignificant compared with global atmospheric light derived from sunlight, $\mathcal{A}(x)$ is equal to induced by global atmospheric light.

At night, global atmospheric light, e.g., from the moon, is intensified and dominated by other regional light sources, e.g., artificial lights. These regional light sources produce glowing regions near them. The observed glow effect G can be modeled as the convolution of a light source with an atmospheric point spread function (APSF) expressed by Legendre polynomial [14].

$$\mathcal{I}(x) = \mathcal{D}(x) + \mathcal{A}(x) + G \text{ and} \quad (3)$$

$$G = \boldsymbol{A}_L(x) * APSF, \quad (4)$$

where $\boldsymbol{A}_L$ is the active light source and its intensity is convolved with $APSF$ to derive the G. Glow extraction is the process of separating glow intensity from a hazy image, i.e., the elimination of airlight $\mathcal{A}(x)$ and direct attenuation $\mathcal{D}(x)$. Note that we distinguish glow and airlight according to their transmission paths. Specifically, Glow is the scattered light that reaches the lens from the visible light source. It is centered at the light source and its intensity changes exponentially with distance from the source. Airlight is the perceived intensity excluding direct attenuation and glow. It is produced by portion of the aggregate of all light sources in which global source homogeneously influencing the entire scene and local sources influencing their respective regions.
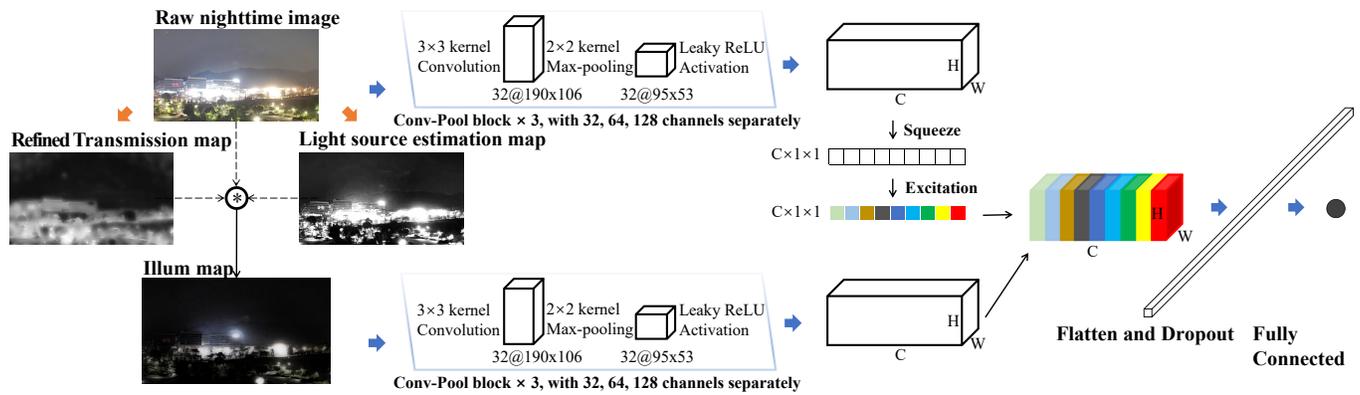
Fig. 1. The architecture and forward process of the proposed method. It involves halation-related features extraction (left, Illumination Map) and network training (right, DSECNet which fuses information from the raw image and Illumination Map). Our method first extracts a refined transmission map and a light source estimation map from the original image and combines them into an Illumination map. This illumination map is then fed into the network together with the original image to estimate PM$_{2.5}$ concentration using the light source glow effect.

B. Feature Maps

As shown in (3), a nighttime hazy image is the combination of three types of intensity sources: glow, direct attenuation, and airlight. To eliminate direct attenuation and airlight, we use refined transmission map and light source estimation map to restrain the effect of scene radiance and distinguish glow from airlight. They are then combined to generate the Illumination Map.

Refined Transmission Map (ReTM): Transmission $t(x)$ is defined as the proportion of $\mathcal{D}(x)$, the direct attenuation including reflected light and light source direct light, reaching the lens. $1 - t(x)$ represents the proportion of scattered light, i.e., $\mathcal{A}(x)$ and G. He et al. propose using the dark channel for transmission map (TM) estimation as follows:

$$1 - \tilde{t}(x) = \omega \min_{c \in \{r,g,b\}} \left(\min_{x' \in \Omega(x)} \frac{\mathcal{I}^c(x')}{\boldsymbol{A}^c} \right), \qquad (5)$$

where x is the pixel index, Ω is a local patch centered at x, x' is the location index inside the patch, ω is haze-retention constant (fixed at 0.95), and $\mathcal{I}^c(x)$ is the pixel value at x position on channel c.

In the case of daytime, $\boldsymbol{A} = \{A^r, A^g, A^b\}$ is a constant vector modeling the aggregate, homogeneous component of lighting, called atmospheric light, which is estimated as the global brightest pixel value. $\min_c(\min_{x' \in \Omega(x)} \frac{\mathcal{I}^c(x')}{A^c})$ is the ratio of the local minimum to the global maximum, representing the visibility of local patch color (non-white), and proportion of scattered light, e.g., when scattered light dominates the patch, the visibility of scene diminishes and the local minimum is close to the global maximum. This globally homogeneous component is large during the day, but small at night making it important to model the light due to each artificial light source individually. Thus, global atmospheric light and regional atmospheric light work together to affect airlight. Treating atmospheric light as a global constant is inappropriate at night. Modeling (varying) atmospheric light in small regions is more appropriate.

We define $\boldsymbol{A}$ as the combination of global and regional atmospheric light, i.e., the maximum in Ω, and refined transmission is

$$\tilde{t}(x) = 1 - \omega \min_c \left(\min_{x' \in \Omega(x)} \frac{\mathcal{I}^c(x')}{\boldsymbol{A}^c(x)} \right) \quad \text{and} \qquad (6)$$

$$\boldsymbol{A}^c(x) = \max_{x' \in \Omega(x)} \mathcal{I}^c(x'). \qquad (7)$$

We use guided image filtering [39] to reduce the halos and block artifacts introduced by the patch Ω; thus the map can capture the edges of objects.

At night, the glow effect is negligible in regions distant from the light source, as expressed in (1). However, in patches close to the light source, non-uniform illumination significantly enhances atmospheric light $\boldsymbol{A}$, glow G, and attenuates transmission $t(x)$, making direct attenuation irrelevant, i.e., $\frac{t(x)}{\boldsymbol{A}^c} \to 0$. By combining (1) and (3), we have

$$\frac{\mathcal{I}(x)}{\boldsymbol{A}} = \frac{t(x)}{\boldsymbol{A}} \mathcal{J}(x) + 1 - t(x) + \frac{G}{\boldsymbol{A}}. \qquad (8)$$

It can be further approximated as

$$\begin{aligned} \mathcal{I}(x) &\approx G + \boldsymbol{A}(1 - t(x)) \\ &\approx G + \boldsymbol{a}, \end{aligned} \qquad (9)$$

where a is an airlight-related constant vector. Note that each image patch corresponds to a local region in which there is typically little scene depth and atmospheric light variation; this enables $\boldsymbol{A}$ and $t(x)$ to be approximated with constants.

Light Source Estimation Map (LSEM): Light source estimation determines the distance between pixels and their most closely associated light sources.

In night images, pixels with high values across all three color channels are generally close to light sources. The channel values may vary due to light source colors. The channel difference, which is defined as the gap between the maximum and minimum values of each color channel at every pixel position, is used to generate the light source estimation map. By merging the maximum value and channel difference, we can derive the light source estimation map as shown in the following equation.

$$L(x) = \min \left((2 * \max_c \frac{\mathcal{I}^c(x)}{255} - \min_c \frac{\mathcal{I}^c(x)}{255})^\gamma, 1 \right), \qquad (10)$$

where $c \in \{r, g, b\}$, γ is a constant to modify the distribution of the normalized light source estimation map.

Illumination Map (IM): The halation (or glow) effect originates primarily from nighttime artificial light sources. Near light sources, glow is the dominant contributor to intensity, either saturating light sensors or requiring reduced sensitivity. This reduces the amount of information gathered about backgrounds near light sources. The refined transmission, $t(x)$, measures the amount of information as the proportion of direct attenuation intensity. Therefore, we use $\boldsymbol{A}^c(x)(1 - t(x))$ to represent the scattered proportion of light, i.e., airlight and glow, and eliminate the impact of direct attenuation.

After eliminating scene radiance, we further eliminate airlight from scattered light based on the exponential attenuation characteristics [14] of the glow. Both the Illumination Map and light source estimation map are used for artificial light sources, their attenuation patterns remain consistent. Therefore, we use the light source estimation map to retain the exponentially varying intensity as glow. The final Illumination Map follows:

$$G^c(x) = \boldsymbol{A}^c(x) \cdot (1 - t(x)) \cdot L(x), \tag{11}$$

where x is the location of pixel, $c \in \{r, g, b\}$ is the color channel.
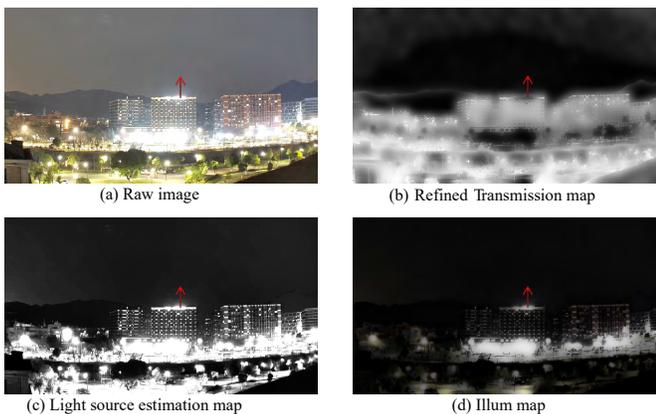
Fig. 2. The raw image and its corresponding feature maps described in III-B. The refined transmission map captures direct attenuation of light source. The light source estimation map indicates the area in which the light source has a dominant effect. The Illumination Map eliminates the effects of direct attenuation and airlight, representing halation.

As shown in Fig. 2, the raw image contains direct attenuation, airlight, and glow simultaneously. The refined transmission map evaluates direct attenuation of light source, and is used to eliminate the effects of direct reflected light in the image. The light source estimation map eliminates airlight but contains glow intensity. Only the regions with sufficient light intensity and smooth variation are identified as having glow feature. Fig. 3 shows the pixel values at the red arrow locations in Fig. 2. The red line in Fig. 3(a) represents the glow component, which is in accordance with the visual change of glow. Fig. 3(b) shows that the variation of glow is consistent with the raw image. The gap here between the raw image and the Illumination Map is the airlight related constant a introduced in (9).
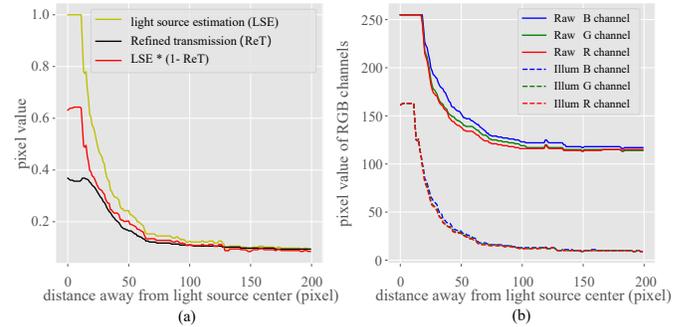
Fig. 3. Visualization of pixel values along red arrow line in Fig. 2. The x-axis is the number of pixels away from the arrow tail along the vertical direction and the y-axis is the corresponding pixels value.

C. Mapping Algorithm
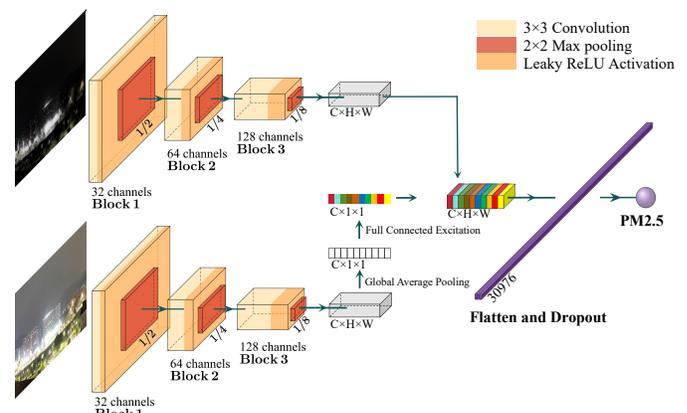
Fig. 4. Architecture of Dual-Channel Squeeze-and-Excitation Convolutional Neural Network (DSECNet). It employs dual channels to receive both the raw image and the Illumination Map as inputs, uses an attention mechanism to enhance feature discrimination, and outputs the PM$_{2.5}$ concentration.

DSECNet extracts and fuses information from the raw image and the Illumination Map. As shown in Fig. 4, both the raw image and extracted Illumination Map are resized to $108 \times 192 \times 3$ to reduce the computation cost. For features extraction, we set the convolutional kernel size to 3×3 and 2×2 max pooling. The leaky ReLU function is employed as the activation function to avoid the dying ReLU problem [40]. This Conv-Pool block is repeated three times consecutively. Raw image and the Illumination Map are processed by two individual Conv-Pool blocks and are bridged by an attention block, which uses the raw image channel to boost feature discriminability [35], [41] in the glow channel. Finally, the flattened vector is input to a fully connected layer to generate the concentration of PM$_{2.5}$. During training, dropout (p=0.5) is employed for the fully connected layer, to enable learning more robust and comprehensive feature representations.

IV. DATA COLLECTION AND PROCESSING

Given that this is the first work on halation-based pollution estimation, evaluating the algorithm required a new dataset. This section describes our dataset, the deployment of our sensor network, and data analysis methods.

A. Overview

The dataset consists of 11,753 nighttime images and the corresponding particle counter based ground truth pollutant concentrations collected from Nov. 2022 to Mar. 2023 in west Hangzhou. The longitude and latitude are $120°02'49.631''$E, $30°13'56.022''$N, respectively. Particulate matter is the main local atmospheric pollutant at night [42]. 75% of the missing samples were a result of inclement weather. 75% of the missing samples were a result of inclement weather. The dataset is used to validate our algorithms. It has been publicly released for use by other researches.

Our dataset contains frames captured every 5 minutes, with $1{,}080 \times 1{,}920$ resolution. Due to the high pollutant concentration ($PM_{2.5} > 80\,\mu g/m^3$) only accounting for 2.3% of the total data, we increase the sampling rate in the high-pollution subset during training to balance the distribution. The testing set is of course unchanged. Resampling is only used for our experiments and has no impact on the released dataset. Specifically, we set the sampling rate to one per minute from 06:20 P.M. to 08:00 P.M. on Feb. 19, during which the $PM_{2.5}$ concentration is between $58\,\mu g/m^3$ and $119\,\mu g/m^3$.

The overall distribution of image capture time is shown in Fig. 5. The images are taken between approximately 06:20 P.M. and 05:00 A.M. when the sun has set and the illumination is low. The experiment lasts for 5 months, from November to March in the following year. The figure also shows the $PM_{2.5}$ concentration distribution during this period. It exhibits a short-tailed pattern: few readings exceed $80\,\mu g/m^3$. There are two methods of evaluating the accuracy of a technique over a range of actual pollutant concentrations. First, one might set the pollutant concentration in a controlled environment. Second, one might compare a new technique against an existing, accurate measurements in the presence of substantial pollutant variation. Our goal is to determine how well proposed technique works in the real world, in the presence of real-world sensing errors and uncertainty. Only the second evaluation method is applicable for real-world scenarios and we therefore use it. Our reported accuracy is in the presence of all real-world sensing errors and uncertainties, e.g., not being provided with light source intensities and camera noise. The technique can be expected to perform no worse, and potentially better, in a laboratory environment in which some uncertainties and noise sources are eliminated. Again, our primary goal is to determine its accuracy in real-world use.

B. Data Collection

To collect the ground truth $PM_{2.5}$ readings and the corresponding images, we deploy sensor and camera networks in downtown Hangzhou, which contains many residential communities and schools. The cameras are placed along a main city road, as demonstrated in Fig. 6. Meanwhile, an air quality sensing platform is placed in the center of the region to record ground truth pollutant readings.

Previous work [43] has reported that regional pollutants exhibit low spatial variability, e.g., using the chemical transport model to predict $PM_{2.5}$ concentration in Pittsburgh in winter, the maximum concentration value increases from $12.9\,\mu g/m^3$
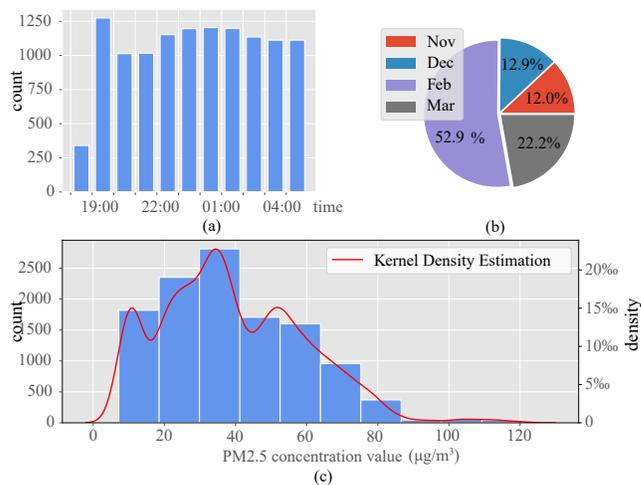
Fig. 5. The time (a, b) and (c) $PM_{2.5}$ distributions of samples in our gathered data. Data are collected after sunset.
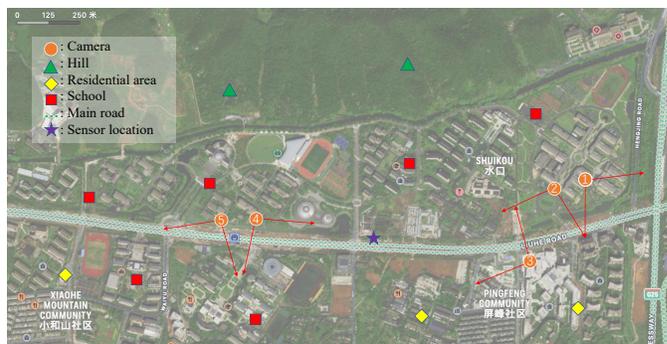
Fig. 6. Sensor and cameras locations. The cameras directions are marked by red arrows. The maximum distance between the sensing platform and cameras is about 1.0 km.

to $16.4\,\mu g/m^3$ when the simulation grid increase from 4×4km to 1×1km. Therefore, we believe that one air quality sensor is sufficient for our experimental setup.

We developed a portable sensing platform to collect and store data, including a Nova particulate matter sensor module, temperature and humidity sensors, and an Arduino UNO controller (see Fig. 7). This platform is powered by lithium batteries and records data once per second.

The deployments and parameters for the Nova PM sensor are shown in Tables I and II. These sensors measure $PM_{2.5}$ (0.0-$999.9\,\mu g/m^3$) and PM_{10} (0.0-$1999.9\,\mu g/m^3$). They operate between $-10\,°C$ to $50\,°C$ and below 70% humidity. They have 1-second measurement response times and detect particles as small as $0.3\,\mu m$, with a $\pm 15\%$ error margin.

We employ five XiaoMi intelligent cameras to capture videos of the monitored area (see Fig. 7). Further details regarding the camera parameters can be found in Table II. The key camera parameters are lens and aperture sizes. Increasing those sizes allows monitoring of large areas in low-light conditions. Each camera covers an approximate area of $2\,km^2$, adopting to spatial variation of $PM_{2.5}$ concentration [19], [44], [45]. It is important to note that we use inexpensive, commodity off-the-shelf camera modules: the technique does not require specialized sensors. These cameras are strategically

1. Temperature and humidity sensor
2. Particulate matter sensor
3. Arduino UNO
4. Lithium battery
5. XiaoMi cameras
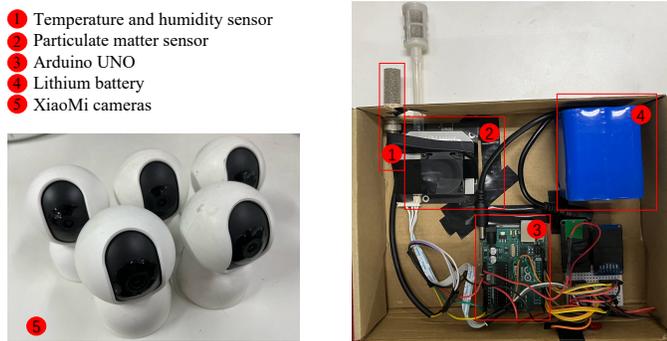
Fig. 7. Sensing platform and cameras used in our deployment. Components are marked in numbers, see Table II for more detailed device parameters.

positioned within buildings situated in university and residential areas, spanning a distance of approximately 4.8 km along the main road. Detailed information about the specific camera locations can be found in Table I and Fig. 6. We adjust the camera pitch angles to capture comprehensive scenes including both the ground and the sky.

TABLE I
DEPLOYMENT SPECIFICATIONS FOR ENVIRONMENTAL SENSORS AND CAMERAS.

Device	GPS Location	Height (m)	Pitch Angle	Acquisition Interval
Camera1	120°03′16.488″E 30°14′15.946″N	18	20°	5 minutes
Camera2	120°03′12.236″E 30°14′12.787″N	18	10°	5 minutes
Camera3	120°03′21.372″E 30°14′08.083″N	33	5°	5 minutes
Camera4	120°02′38.086″E 30°13′47.709″N	18	20°	5 minutes
Camera5	120°02′34.189″E 30°13′45.393″N	15	5°	5 minutes
Sensors	120°02′53.592″E 30°13′54.323″N	18	-	1 second

C. Nighttime Observations

In theory, the formation, propagation, and dissipation of PM$_{2.5}$ are affected by climate and weather conditions. For instance, low night-time temperatures can cause pollutants to stay closer to the ground. However, in our experiments, we found the influence of these environmental factors on PM concentration to be small. We calculate the R^2 correlation coefficients for the two environmental parameters and discover that the correlations are weak, as demonstrated in Table III.

Air pollution shows significant temporal variation throughout the day [46]. In particular, decreased human activities during the night can lead to reduced pollutant concentration. The drop in temperature during the night can also cause atmospheric pollutants to be trapped and accumulate.

To assess the differences in the rate of change of PM$_{2.5}$ between day and night, we use both absolute and relative rates of variation. Since a portion of our PM$_{2.5}$ data does not

TABLE II
PARAMETERS OF PM DEVICE AND CAMERAS.

Nova PM sensor :	
Sensor range	[PM$_{2.5}$] 0.0 to 999.9 µg/m^3
	[PM$_{10}$] 0.0 to 1999.9 µg/m^3
Operating temperature	-10 to 50 °C
Maximum operating humidity	70%
The response time	1 second
Serial port data output frequency	1 Hz
Minimum resolution particle size	0.3 µm
The relative error	Max. ±15% and ±10µg/m^3
	(Note: 25 °C,50%RH)
Standard certification	CE/FCC/RoHS
Camera :	
Lens	FOV: 110°
	Aperture: f/1.4
	Shooting Range: 0.6 m to ∞
Video Resolution	1080 × 1920, MP4 (H.265/HEVC)
Operating temperature	-10 to 45 °C

TABLE III
CORRELATION WITH ENVIRONMENTAL FACTORS

	PM$_{2.5}$	PM$_{10}$	Temperature	Humidity
PM$_{2.5}$	1.0	0.758	0.325	0.349
PM$_{10}$	0.758	1.0	0.253	0.302
Temperature	0.325	0.253	1.0	0.787
Humidity	0.349	0.302	0.787	1.0

cover a full 24-hour period, we select 13 days that contain 24-hours of data and calculate their average. The absolute rate of variation v_a and the relative rate of variation v_r are calculated as follows:

$$v_a = \frac{1}{N} \sum_{t=1}^{N} |p_{t+1} - p_t| \text{ and} \tag{12}$$

$$v_r = \frac{1}{N} \sum_{t=1}^{N} |p_{t+1} - p_t|/p_t, \tag{13}$$

where $N = \frac{A}{s}$ represents the length of the sequence, A is the number of data points covering 24 hours, s is the sampling interval, and p_t denotes the PM$_{2.5}$ concentration at the t-th sampling point.

The results are presented in Fig. 8, which clearly demonstrates that the level of PM$_{2.5}$ varies less at night than during the day.

D. Glow Effect

We first design an algorithm, as shown in Algorithm IV-D, to determine the attenuation rate of light intensity near the light source, which is closely related to the PM$_{2.5}$ concentration. Specifically, we put a 101×101 gray-scale image patch P

This article has been accepted for publication in IEEE Sensors Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2025.3525712

8   IEEE SENSORS JOURNAL, VOL. XX, NO. XX, XXXX 2024
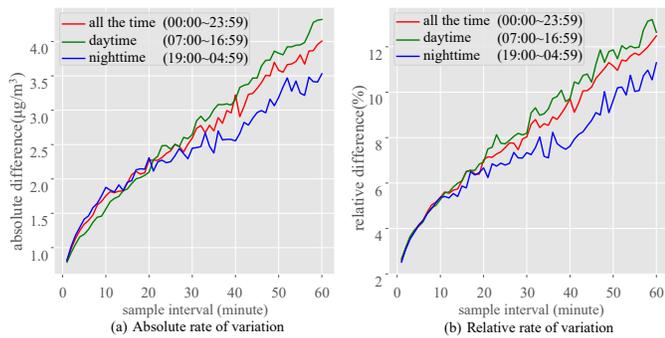
Fig. 8. Diurnal difference in PM$_{2.5}$ variation rate. Daytime, nighttime, and all the time are marked in different colors. Absolute (a) and relative (b) rates of variation. The level of PM$_{2.5}$ varies less at night than during the day.

Fig. 9. Light source of location 3 (see Fig. 6). The red arrow represents the transmission direction for attenuation analysis.

Algorithm 1 Calculation of Light Source

Input: an image patch P; coordinates set $\mathbb{X}$

Output: light source binary map M; light source central coordinate $\boldsymbol{x_c}$

1: Init 1: $\mathbb{X}_{255} \leftarrow \{\boldsymbol{x} \mid \boldsymbol{x} \in \mathbb{X}, P(\boldsymbol{x})=255\}$,
2: $\quad M \leftarrow O_{101 \times 101}$
3: Init 2: $\boldsymbol{x_c} \leftarrow \text{Average}(\mathbb{X}_{255})$, $M(\boldsymbol{x_c}) \leftarrow 1$
4: Step 1: Create 1-D array $\mathbb{A}$ **s.t.** $\forall\ 1 \leq i < j \leq |\mathbb{A}|$ and $E(\mathbb{A}(i), \boldsymbol{x_c}) \leq E(\mathbb{A}(j), \boldsymbol{x_c})$
5: **for** $(i \leftarrow 2; \boldsymbol{x} \leftarrow \mathbb{A}(i), i \leq |\mathbb{A}|; i++)$ **do**
6: $\quad$ Step 2.1:
7: $\quad$ **if** $P(\boldsymbol{x})=255$ and $\hat{\boldsymbol{x}} \in \mathbb{A}$ and $M(\hat{\boldsymbol{x}}) * H(\hat{\boldsymbol{x}}, \boldsymbol{x})=1$ **then**
8: $\quad\quad$ $M(\boldsymbol{x}) \leftarrow 1$ // identified as light source
9: $\quad$ **end if**
10: $\quad$ Step 2.2:
11: $\quad$ **if** $i > 200$ and $\sum_{j=i-199}^{i} M(\boldsymbol{x_j}) < 20$ **then**
12: $\quad\quad$ break
13: $\quad$ **end if**
14: **end for**
15: Step 3: $\boldsymbol{x_c} \leftarrow \text{Average}(\mathbb{X}_M)$,
16: $\quad$ $\mathbb{X}_M \leftarrow \{\boldsymbol{x} \mid \boldsymbol{x} \in \mathbb{X}, M(\boldsymbol{x})=1\}$
17: **return** $M, \boldsymbol{x_c}$

---

in a Cartesian coordinate system and use $\boldsymbol{x} = (x_1, x_2)$ as the pixel coordinate. $\mathbb{X}$ is the coordinates set of all pixels. We first initialize the light source central coordinate $\boldsymbol{x_c}$ by the average coordinates of saturated pixels. The light source binary map M is initialized with zeros, with the $M(\boldsymbol{x_c})$ set to 1. $\mathbb{A}$ is the ordered array of elements in $\mathbb{X}$ sorted by their Euclidean distance to $\boldsymbol{x_c}$. Subsequently, we traverse the coordinates in $\mathbb{A}$ and mark it as the light source in M if it is saturated and satisfies the Euclidean distance ($E(\cdot)$) and the Hamiltonian distance ($H(\cdot)$) requirement. We terminate this process when more than 200 pixels are traversed and there are less than 20 pixels marked as the light source in the last traversal step. In the end, we update $\boldsymbol{x_c}$ by M.

We use the light source located in Fig. 9 to evaluate the correlation between the glow effect and PM$_{2.5}$ concentration. 1,000 images are selected randomly and divided into four sets based on their concentration ranges. The region with a gray-scale pixel values of 255 is defined as the light source. We

verified that all images contain such regions and did not use normalization. We indicate the region 200 pixels away from the edge of the light source for illustration; at this distance, there is generally little glow. We plot the pixel values as a function of distance away from the edge of the light source along the red arrow.
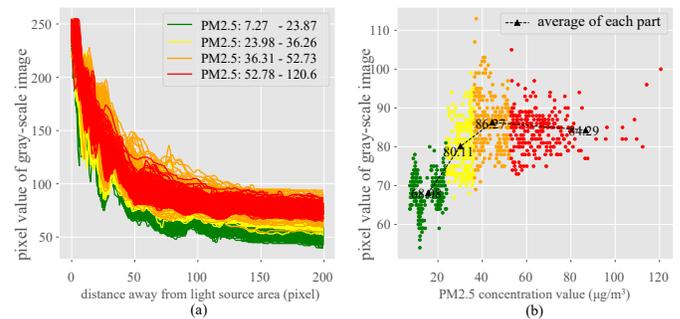
Fig. 10. Attenuation of the glowing region: (a) The x-axis is the number of pixels away from the light source along the horizontal direction and the y-axis is the corresponding pixel value; (b) The x-axis is PM$_{2.5}$ concentration value of each line and the y-axis is pixel value 100-pixels from the light source edge. The stratification (a) and linear relationship (b) illustrate the existence of a correlation between PM$_{2.5}$ concentration and light source glow.

As shown in Fig. 10(a), each line represents an image, and its color indicates the corresponding PM$_{2.5}$ range. There is an obvious layering by color, indicating that the pixel values decreased more rapidly with a decrease in PM$_{2.5}$ concentration, thus confirming the correlation between the glow effect and PM$_{2.5}$ concentration. Specifically, using pixels 100 pixels away from the light source (along the horizontal direction) as an example, Fig. 10(b) shows that the average dependence of intensity on PM$_{2.5}$ is 0.63 pixel intensity units every μg/m^3 when the concentration is lower than 50 μg/m^3. Note that the red portion does not exhibit this trend, possibly due to increased PM$_{2.5}$ increasing the overall brightness and the intelligent camera decreases the exposure automatically. Moreover, the most significant attenuation in pixel values occurred within a distance of 100 pixels. This guides our selection of hyperparameters in later experiments.

We make no assumption that users will provide actual brightnesses of artificial light sources.

V. EXPERIMENTAL RESULTS

In this section, we introduce the experimental setup and the corresponding experimental design and results.

A. Experimental Setup
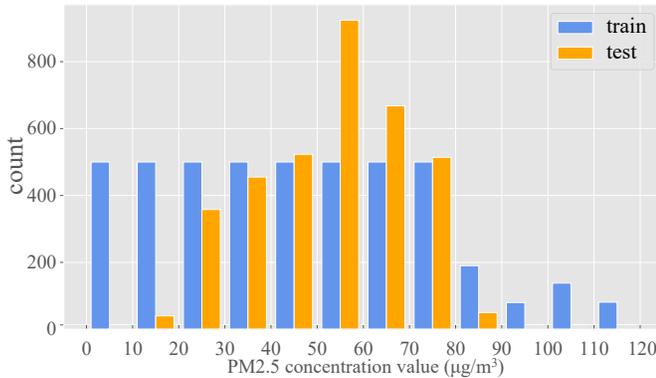
Fig. 11. Distribution of experimental data. There are 534 samples from highly polluted environments (PM$_{2.5}$ > 80 µg/m^3) after replication. After correction, the distribution of the training set is more balanced.

We select the data obtained in Feb. and Mar., which contains 8,829 images. After sorting by capture time, they are divided 3:2 into training and testing data sets. For the training set, since there were few high-concentration samples (PM$_{2.5}$ >80 µg/m^3), we resample the training set to balance the distribution (see Section IV). Fig. 11 shows the sample distribution of the training set and testing set. The algorithm is developed using PyTorch (version 1.11.0, CUDA 10.2), Adam optimizer with a batch size of 128 [47], and mean squared error loss function. We employ multi-step learning rate decay strategy, where the learning rate is initialized to $5\,e^{-3}$ and stepped down to $1\,e^{-5}$ in 400 epochs.

The model is trained on a server equipped with a Tesla V100 GPU and 2.30 GHz Intel(R) Xeon(R) Gold 5218 CPU. As suggested in [48], we use the mean absolute error (MAE), mean relative error (MRE), and rooted mean squared error (RMSE) as the evaluation metrics.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|, \tag{14}$$

$$MRE = \frac{1}{N}\sum_{i=1}^{N}\frac{|y_i - \hat{y}_i|}{y_i}, \tag{15}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}, \tag{16}$$

where N is the total number of samples, $\hat{y}_i$ is the predicted value of i-th sample, and y_i is the ground truth value.

B. Evaluation of Illumination Map Based Model

Our work is the first on nighttime vision-based pollution estimation. We compare our technique with the most closely related state-of-the-art PM$_{2.5}$ estimation techniques, which were developed for daytime use. As shown in Table IV, our method outperforms all daytime methods consistently. The poor performance on the pre-trained PE-MobileNetV2 [15] indicates that the vision-based daytime model cannot be directly applied to night scenarios. Compared with the SOTA daytime methods, our technique reduces MAE by 16.99% and MRE by 8.02%.

TABLE IV
COMPARISON WITH EXISTING (DAYTIME) METHODS (µg/m^3)

Method	MAE	MRE	RMSE
Our Method	**8.65**	**16.74%**	**11.02**
Pre-trained PE-MobileNetV2 [15]	18.15	40.04%	22.77
Re-trained PE-MobileNetV2 [15]	10.42	18.20%	13.73
Utomo [29]	12.44	21.81%	16.16
MIFF [30]	11.13	21.97%	14.27
ImgSensingNet [12]	10.96	22.83%	13.98

C. Ablation Analysis

We argue that the glow feature is key enabler of accurate image-based nighttime pollution estimation. Using the same architecture, we compare the proposed Illumination Map and popular feature maps to determine its impact. The Illumination Map is replaced by baseline features in the network, including Li's glow feature [36], dark channel, and transmission [33]. As shown in Table V, our method, which uses refined transmission and light source estimation for glow extraction, ourperforms other methods, indicating the benefit of using Illumination Map features. Compared with the second-best feature (Dark Channel), our technique reduces MAE by 13.33% and MRE by 10.10%.

The effectiveness of the transmission map-based method is diminished at night [49], as it tends to mistakenly identify artificial light sources as light sources at infinity. This issue arises due to the uneven distribution of brightness caused by low illumination and artificial light sources. Consequently, many regions in the transmission map exhibit pixel values that are either close to 0 (black regions) or 255 (white regions). This disrupts the overall relationship between transmission and scene depth, resulting in poor transmission map performance.

On the other hand, the dark channel-based method seeks the minimal value of RGB channels within a patch to generate a coarse output. However, this approach neglects a significant amount of valuable image information.

Li's approach addresses the color shift problem by imposing a global RGB channel constraint and extracting a smooth layer using a spectrum-based algorithm. However, when dealing with large images containing multi-color lights, unexpected color shifts can occur, and object edges may exist. Compared with Li's method, which relies on statistical prior knowledge that the gradient histogram of the smooth layer has a short tail distribution, our Illumination Map estimates the proportion of radially scattered light in a raw image based on physical imaging model. Experimental results demonstrate that our Illumination Map improves MAE by 25.56% relative to the "Glow Map" described by Li et al.

TABLE V
COMPARISON OF DIFFERENT FEATURES USING OUR MODEL ($\mu g/m^3$).

Method	MAE	MRE	RMSE
DSECNet + IM	**8.65**	**16.74%**	**11.02**
DSECNet + Raw image	10.30	18.88%	13.06
DSECNet + He's dark channel [33]	9.98	18.62%	12.92
DSECNet + Li's glow feature [36]	11.62	21.18%	14.93
DSECNet + TM [33]	10.09	19.45%	12.51
DSECNet + LSEM	10.79	19.34%	13.65
DSECNet + ReTM	9.17	17.09%	11.57

TABLE VII
IMPACT OF VARYING INPUT SIZE AND RECEPTIVE FIELD

Input size	Receptive field	MAE ($\mu g/m^3$)	MRE	Flops	Parameters
(252,448)	(94,94)	9.54	18.05%	2.19G	395,137
(216,384)	(110,110)	9.34	16.97%	1.58G	341,889
(180,320)	(132,132)	9.43	17.17%	1.08G	291,969
(144,256)	(165,165)	9.83	18.85%	675.17M	256,129
(108,192)	**(220,220)**	**8.65**	**16.74%**	**365.37M**	**225,665**
(81,144)	(293,293)	9.46	16.53%	192.54M	211,073
(54,96)	(440,440)	11.35	20.21%	79.26M	201,089
(27,48)	(880,880)	12.06	22.28%	13.59M	195,201

One limitation of our method is that it is most accurate when trained on data from the deployment location. We evaluate its generalizability in unseen places by dividing the test set and training set according to location, that is, images collected in a specific location are used for testing and the other four locations are for training. The results are shown in Table VI, where we select models from 25 and 100 training epochs to test under the case of preliminary fitting and sufficient fitting. It indicates the error of our model increases significantly when applied to unseen places. In the worst case, the MAE is $40.12\,\mu g/m^3$ in Location3.

TABLE VI
GENERALIZATION IN DIFFERENT LOCATIONS ($\mu g/m^3$)

	The Epoch is 25		The Epoch is 100	
	MAE	RMSE	MAE	RMSE
Location1	10.68	13.97	13.93	19.97
Location2	27.29	30.33	25.06	29.42
Location3	40.12	42.49	17.53	21.17
Location4	9.57	12.99	17.08	22.47
Location5	9.37	13.42	19.25	26.87

D. Receptive Field

Fig. 10(a) shows that the attenuation rate of the light intensity ranges from 47.75% to 79.41% within a 100 pixels radius. Therefore, we expect the receptive field of the neuron in the final Conv-Pool block to cover an area of approximately 200×200 pixels in the original image. To test this hypothesis, we train models with varying input sizes and evaluate each case, as shown in Table VII. We use a scaling factor (γ) that ranges from 4 to 40, to scale down the original $1,080\times1,920$ image. The results indicate that the best input size is (108,192) with the corresponding receptive field of (220,220), i.e., neurons of the last layer have a field of view that approximates (220,220) in the original image. The receptive field size (RFS) is calculated as follows:

$$RFS = (((1 * 2 + 2) * 2 + 2) * 2 + 2)/\gamma, \qquad (17)$$

where sf is the scaling factor and other constants are determined by the network structure. Increasing the input resolution beyond that point leads to a decrease in the receptive field and an increase in MAE and MRE. Moreover, decreasing the input resolution causes a loss of image details and also a deterioration in network accuracy.

E. Sky Region Impact

The sky is crucial in estimating daytime $PM_{2.5}$ concentration in outdoor environments [11], [13], [15], [29]. Typically, the daytime airlight is assumed to be the intensity of the sky region at an infinite distance. However, sky-based estimation is much less effective at night due to reduced illumination.

(a) Illu map   (b) Grad-CAM of the Illum map

(c) Raw image   (d) Grad-CAM of the raw image
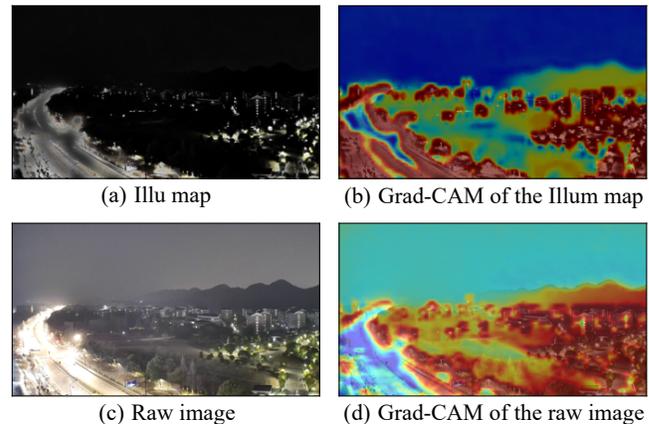
Fig. 12. Grad-CAM of the first convolution layer of Illu map and raw image. Compared to the raw image, the Illu Map places less emphasis on the non-light source regions, airlight, and background, while focusing more on the glow itself.

To investigate the impact of the sky, we partition the image into two regions: the sky and non-sky. As shown in Table VIII, training the model using only the sky region results in an additional error of $5.965\,\mu g/m^3$ compared to using both regions. Removing the nighttime sky region has little impact (MAE is $1.98\,\mu g/m^3$) on the overall model accuracy.

TABLE VIII
EVALUATION OF SKY AND NON-SKY REGIONS

	sky region	non-sky region	sky + non-sky
MAE ($\mu g/m^3$)	14.62	10.63	**8.65**
MRE (%)	27.65%	19.65%	**16.74%**
RMSE ($\mu g/m^3$)	18.54	13.17	**11.02**

We then use gradient-weighted class activation mapping (Grad-CAM) [50] to visualize the network attention on input images. Grad-CAM generates a heatmap of each pixel's contribution to the final prediction by using back-propagation

of gradients. Since the resolution of the Grad-CAM heatmap is consistent with the target hidden layer, we select the first convolution layer to illustrate the fine-grained details of the image. As shown in Fig. 12, the network focuses its attention on the glow of light sources. In comparison to the raw image, the Illumination Map pays less attention to the non-light source region, airlight, and background while focusing more on the glow itself.

F. Performance on PM$_{10}$

As shown in Fig. 13, the Pearson correlation coefficient between PM$_{2.5}$ and PM$_{10}$ in the testing set is 0.9594, indicating a strong positive correlation. When the experimental data are replaced with PM10, the proposed method achieves a RMSE of 24.73 μg/m^3, a MRE of 20.07%, and a MAE of 19.88 μg/m^3. Notably, the MRE is comparable to the model's performance on PM$_{2.5}$, suggesting that the proposed method is also effective for PM$_{10}$ estimation.

The increase in MAE and RMSE for PM$_{10}$ compared to PM$_{2.5}$ is attributed to the broader range of PM$_{10}$ values, which reaches over 140 μg/m^3. Despite this, the model remains robust across both particle size measurements.
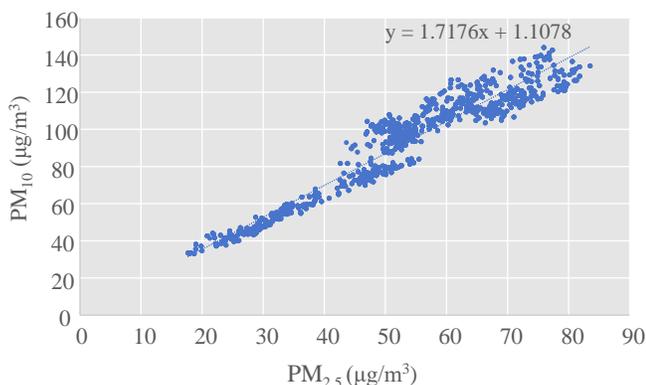
Fig. 13. The correlation between PM$_{\mathbf{2.5}}$ and PM$_{\mathbf{10}}$ in the testing set, where the Pearson coefficient is 0.96. The strong positive correlation suggest that the proposed method is also effective for PM10 estimation.

Fig. 14. Color light extraction results. The first row is the mask for non-color and color light, while the second row is the original images after applying masks.

TABLE IX
IMAPCT OF COLOR

	RMSE	MAE	MRE
Non-color (μg/m^3)	13.76	11.59	20.14
Color (μg/m^3)	14.68	11.97	20.44

G. Impact of Color

The color of artificial light sources can vary, which differs from the daytime pollution estimation problem in which sunlight is the primary source of illumination. We therefore evaluated the effect of color on estimation accuracy. In our dataset, we distinguish the intensity of glow color by the Saturation (S) in HSV space. The poor uniformity of the RGB color space makes it inappropriate to define color difference as the distance between RGB color vectors. HSV Saturation indicates the purity and intensity of a color. Specifically, we first select the light source area according to the brightness map (B¿0.7), then convert this area into HSV to calculate the saturation value S, where S equals 0 indicates that the area is non-colored. Finally, we distinguish the colored area according to whether the S value is greater than 0.

As shown in Fig. 14, we use a black mask to get colored and non-colored light data. The impact of color on performance is shown in Table IX, the predictions on the colored data are slightly less accurate than the non-colored ones. This may be because white (non-colored) light has a wider spectrum than colored light and therefore contains more information. Whether particular (coarse, given the RGB representation) wavelengths are better suited to vision-based pollution estimation remains an open question. However, our reported accuracies were for data containing color variation in a real-world environment, so the approach appears to be robust to this variation.

VI. CONCLUSION AND DISCUSSIONS

We have described a nighttime PM$_{2.5}$ concentration estimation technique based on light source Illumination Map extraction. We have also presented a publicly released dataset including PM$_{2.5}$ and PM$_{10}$ concentrations, humidity, and temperature. The experimental results demonstrate that our Illumination Map based approach enables a 16.99% improvement in MAE compared to the existing daytime methods.

During deployment, due to the relatively low pollution level in Hangzhou, most PM$_{2.5}$ values are under 80 μg/m^3. Moreover, rainy weather sometimes prevented data collection. Our method is designed for use at night and does best when artificial light sources are present within the scene. Therefore, it may not work well during the day, at dawn, or at dusk when the sun dominates artificial light sources; under these conditions, existing (daytime) approaches are adequate.

Future work includes considering other environmental parameters, such as wind and rain, to improve the accuracy of PM$_{2.5}$ concentration estimation. Moreover, collecting a more diverse and larger-scale nighttime image dataset might enable further improvements in model accuracy.

DATA AVAILABILITY

Our dataset is available at https://github.com/Kaihua-Zhang/Light-Illum based-Nighttime-PM2.5-Estimation.

The following files are available free of charge:

- image dataset and corresponding environment labels and
- code for Illumination Map implementation in Python.

REFERENCES

[1] J. Park, E. H. Park, J. J. Schauer, S. M. Yi, and J. Heo, "Reactive oxygen species (ROS) activity of ambient fine particles (PM2.5) measured in Seoul, Korea," *Environ. Int.*, vol. 117, pp. 276–283, 2018.

[2] M. Rohrer, A. Flahault, and M. Stoffel, "Peaks of fine particulate matter may modulate the spreading and virulence of COVID-19," *Earth Syst. Environ.*, vol. 4, no. 4, pp. 789–796, 2020.

[3] S. M. Ali, F. Malik, M. S. Anjum, G. F. Siddiqui, M. N. Anwar, S. S. Lam, A. S. Nizami, and M. F. Khokhar, "Exploring the linkage between PM2.5 levels and COVID-19 spread and its implications for socio-economic circles," *Environ. Res.*, vol. 193, p. 110421, 2021.

[4] R. K. Chakrabarty, P. Beeler, P. Liu, S. Goswami, R. D. Harvey, S. Pervez, A. van Donkelaar, and R. V. Martin, "Ambient PM2.5 exposure and rapid spread of COVID-19 in the United States," *Sci. Total Environ.*, vol. 760, p. 143391, 2021.

[5] M. Alvarado-Molina, A. Curto, A. J. Wheeler, R. Tham, E. Cerin, M. Nieuwenhuijsen, R. Vermeulen, and D. Donaire-Gonzalez, "Improving traffic-related air pollution estimates by modelling minor road traffic volumes," *Environ. Pollut.*, vol. 338, p. 122657, 2023.

[6] V. V. Le, D. H. Nguyen, H.-D. Wang, B.-H. Liu, and S.-I. Chu, "Efficient UAV scheduling for air pollution source detection from chimneys in an industrial area," *IEEE Sensors J.*, vol. 22, no. 20, pp. 19983–19994, 2022.

[7] Y. Peng, Z. i. Sui, Y. Zhang, T. Wang, P. Norris, and W. P. Pan, "The effect of moisture on particulate matter measurements in an ultra-low emission power plant," *Fuel*, vol. 238, pp. 430–439, 2019.

[8] B. Feenstra, V. Papapostolou, S. Hasheminassab, H. Zhang, B. Der Boghossian, D. Cocker, and A. Polidori, "Performance evaluation of twelve low-cost PM2.5 sensors at an ambient air monitoring site," *Atmos. Environ.*, vol. 216, p. 116946, 2019.

[9] Y. Lu, G. Giuliano, and R. Habre, "Estimating hourly PM2.5 concentrations at the neighborhood scale using a low-cost air sensor network: A Los Angeles case study," *Environ. Res.*, vol. 195, p. 110653, 2021.

[10] F. Weng, H. Huang, and X. Han, "Monitoring PM2.5 distributions over China from geostationary satellite observations," in *IGARSS IEEE Int.l Geosci. Remote Sens. Symp.* IEEE, 2020, pp. 5581–5583.

[11] M. Wang, Y. Wang, F. Teng, S. Li, Y. Lin, and H. Cai, "Estimation and analysis of PM2.5 concentrations with NPP-VIIRS nighttime light images: A case study in the Chang-Zhu-Tan Urban Agglomeration of China," *Int. J. Environ. Res. Public Health*, vol. 19, no. 7, p. 4306, 2022.

[12] Y. Yang, Z. Hu, K. Bian, and L. Song, "ImgSensingNet: UAV vision guided aerial-ground air quality sensing system," in *IEEE INFOCOM Conf. Comput. Commun.* IEEE, 2019, pp. 1207–1215.

[13] Z. Wang, S. Yue, and C. Song, "Video-based air quality measurement with dual-channel 3-D convolutional network," *IEEE Internet Things J.*, vol. 8, no. 18, pp. 14372–14384, 2021.

[14] S. G. Narasimhan and S. K. Nayar, "Shedding light on the weather," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1. IEEE, 2003, pp. I–I.

[15] X. Fang, Z. Li, B. Yuan, and Y. Chen, "Image-based PM2.5 estimation from imbalanced data distribution using prior-enhanced neural networks," *IEEE Sensors J.*, vol. 24, no. 4, pp. 4677–4693, 2024.

[16] M. L. Zamora, J. Rice, and K. Koehler, "One year evaluation of three low-cost PM2.5 monitors," *Atmos. Environ.*, vol. 235, p. 117615, 2020.

[17] P. Wong, H. Lee, Y. Chen, Y. Zeng, Y. Chern, N. Chen, S. C. Lung, H. Su, and C. Wu, "Using a land use regression model with machine learning to estimate ground level PM2.5," *Environ. Pollut.*, vol. 277, p. 116846, 2021.

[18] S. Zhai, Y. Zhang, J. Huang, X. Li, W. Wang, T. Zhang, F. Yin, and Y. Ma, "Exploring the detailed spatiotemporal characteristics of PM2.5: Generating a full-coverage and hourly PM2.5 dataset in the Sichuan Basin, China," *Chemosphere*, vol. 310, p. 136786, 2023.

[19] Z. Chen, T. Zhang, Z. Chen, Y. Xiang, Q. Xuan, and R. P. Dick, "HVAQ: A high-resolution vision-based air quality dataset," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–10, 2021.

[20] M. Krishan, S. Jha, J. Das, A. Singh, M. K. Goyal, and C. Sekar, "Air quality modelling using long short-term memory (LSTM) over NCT-Delhi, India," *Air Qual. Atmos. Health*, vol. 12, pp. 899–908, 2019.

[21] J. Ma, Y. Ding, V. J. Gan, C. Lin, and Z. Wan, "Spatiotemporal prediction of PM2.5 concentrations at different time granularities using IDW-BLSTM," *IEEE Access*, vol. 7, pp. 107897–107907, 2019.

[22] C. Guo, G. Liu, L. Lyu, and C.-H. Chen, "An unsupervised PM2.5 estimation method with different spatio-temporal resolutions based on KIDW-TCGRU," *IEEE Access*, vol. 8, pp. 190263–190276, 2020.

[23] T. Zhang and R. P. Dick, "Image-based air quality forecasting through multi-level attention," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*. IEEE, 2022, pp. 686–690.

[24] M. Dao, K. Zettsu, and U. K. Rage, "IMAGE-2-AQI: Aware of the surrounding air qualification by a few images," in *Adv. Trends Artif. Intell. Theor. Pract.* Springer, 2021, pp. 335–346.

[25] K. Gu, J. Qiao, and X. Li, "Highly efficient picture-based prediction of PM2.5 concentration," *IEEE Trans. Ind. Electron.*, vol. 66, no. 4, pp. 3176–3184, 2019.

[26] O. Vernik, A. Degani, and B. Fishbain, "Mathematical estimation of particulate air pollution levels by aerosols tomography," *IEEE Sensors J.*, vol. 22, no. 9, pp. 8274–8281, 2022.

[27] T. Zhang and R. P. Dick, "Estimation of multiple atmospheric pollutants through image analysis," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*. IEEE, 2019, pp. 2060–2064.

[28] P. Su, Y. Liu, S. Tarkoma, A. Rebeiro-Hargrave, T. Petäjä, M. Kulmala, and P. Pellikka, "Retrieval of multiple atmospheric environmental parameters from images with deep learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[29] S. Utomo, A. John, A. Pratap, Z.-S. Jiang, P. Karthikeyan, and P.-A. Hsiung, "Aix implementation in image-based PM2.5 estimation: Toward an AI model for better understanding," in *Proc. Int. Conf. Knowl. Smart Technol. (KST)*, 2023, pp. 1–6.

[30] G. Wang, Q. Shi, H. Wang, K. Sun, Y. Lu, and K. Di, "Multi-modal image feature fusion-based PM2.5 concentration estimation," *Atmos. Pollut. Res.*, vol. 13, no. 3, p. 101345, 2022.

[31] C. Li, B. Zhang, D. Hong, X. Jia, A. Plaza, and J. Chanussot, "Learning disentangled priors for hyperspectral anomaly detection: A coupling model-driven and data-driven paradigm," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–14, 2024.

[32] D. Hong, B. Zhang, H. Li, Y. Li, J. Yao, C. Li, M. Werner, J. Chanussot, A. Zipf, and X. X. Zhu, "Cross-city matters: A multimodal remote sensing benchmark dataset for cross-city semantic segmentation using high-resolution domain adaptation networks," *Remote Sensing of Environment*, vol. 299, p. 113856, 2023.

[33] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, 2010.

[34] C. Hsieh, K. Chen, M. Jiang, J. Liaw, and J. Shin, "Estimation of PM2.5 concentration based on support vector regression with improved dark channel prior and high frequency information in images," *IEEE Access*, vol. 10, pp. 48486–48498, 2022.

[35] H. Chen, W. Yang, and Q. Liao, "Two-stream non-uniform concentration reasoning network for single image air pollution estimation," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*. IEEE, 2022, pp. 501–505.

[36] Y. Li, R. T. Tan, and M. S. Brown, "Nighttime haze removal with glow and multiple light colors," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2015, pp. 226–234.

[37] H. Koschmieder, *Theorie der horizontalen Sichtweite*. Keim & Nemnich, 1925.

[38] R. T. Tan, "Visibility in bad weather from a single image," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2008, pp. 1–8.

[39] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, 2012.

[40] A. K. Dubey and V. Jain, "Comparative study of convolution neural network's relu and leaky-relu activation functions," in *Appl. Comput., Automat. Wireless Syst. Elect. Eng.* Springer, 2019, pp. 873–880.

[41] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018, pp. 7132–7141.

[42] Hangzhou People's Government, "2022 hangzhou ecological environment status bulletin," https://www.hangzhou.gov.cn/art/2023/6/5/art_1228974625_59081990.html, 2023, accessed April 10, 2023.

[43] P. Garcia Rivera, B. T. Dinkelacker, I. Kioutsioukis, P. J. Adams, and S. N. Pandis, "Source-resolved variability of fine particulate matter and human exposure in an urban area," *Atmospheric Chemistry and Physics*, vol. 22, no. 3, pp. 2011–2027, 2022.

This article has been accepted for publication in IEEE Sensors Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2025.3525712

XIANG *et al.*: LIGHT-ILLUM BASED NIGHTTIME PM$_{2.5}$ ESTIMATION
13

[44] M. Eeftens, M.-Y. Tsai, C. Ampe, B. Anwander, R. Beelen, T. Bellander, G. Cesaroni, M. Cirach, J. Cyrys, K. de Hoogh *et al.*, "Spatial variation of PM2.5, PM10, PM2.5 absorbance and PM coarse concentrations between and within 20 European study areas and the relationship with NO2–results of the ESCAPE project," *Atmos. Environ.*, vol. 62, pp. 303–317, 2012.

[45] Y. Xie, Y. Wang, K. Zhang, W. Dong, B. Lv, and Y. Bai, "Daily estimation of ground-level PM2.5 concentrations over Beijing using 3km resolution MODIS AOD," *Environ. Sci. Technol.*, vol. 49, no. 20, pp. 12 280–12 288, 2015.

[46] P. Y. Wong, H. J. Su, S. C. C. Lung, and C. D. Wu, "An ensemble mixed spatial model in estimating long-term and diurnal variations of PM2.5 in Taiwan," *Sci. Total Environ.*, p. 161336, 2023.

[47] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[48] P. Motto Ros, M. Laterza, D. Demarchi, M. Martina, and C. Bartolozzi, "Event-driven encoding algorithms for synchronous front-end sensors in robotic platforms," *IEEE Sensors J.*, vol. 19, no. 16, pp. 7149–7161, 2019.

[49] X. Jiang, H. Yao, S. Zhang, X. Lu, and W. Zeng, "Night video enhancement using improved dark channel prior," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*. IEEE, 2013, pp. 553–557.

[50] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 618–626.

Zuohui Chen received the B.S. degree in automation from the Zhejiang University of Technology, Hangzhou, China, in 2019. He is working toward the PhD. degree in control theory and engineering from the Zhejiang University of Technology. His current research interests include computer vision, AI application, and AI security.

Yun Xiang (M'11) is an Associate Professor of the Institute of Cyberspace Security, College of Information Engineering at the Zhejiang University of Technology. He received his Ph.D. degree from University of Michigan in 2014 and his B.S. degree from Zhejiang University in 2008. His current research interests include cyberspace security, machine learning, embedded systems, and computer vision.

Qi Xuan (M'18) received the BS and PhD degrees in control theory and engineering from Zhejiang University, Hangzhou, China, in 2003 and 2008, respectively. He was a Post-Doctoral Researcher with the Department of Information Science and Electronic Engineering, Zhejiang University, from 2008 to 2010, respectively, and a Research Assistant with the Department of Electronic Engineering, City University of Hong Kong, Hong Kong, in 2010 and 2017. From 2012 to 2014, he was a Post-Doctoral Fellow with the Department of Computer Science, University of California at Davis, CA, USA. He is a member of the IEEE and is currently a Professor with the Institute of Cyberspace Security, College of Information Engineering, Zhejiang University of Technology, Hangzhou, China. His current research interests include network science, graph data mining, cyberspace security, machine learning, and computer vision.

Kaihua Zhang is an Electrical Engineer of the Planning Company, PowerChina Hubei Electric Engineering Company Corporation Limited. He received the B.S. degree in energy power from the Hohai University, Nanjing, China, in 2016 and his M.D. degree in control science and engineering from the Zhejiang University of Technology, Hangzhou, China, in 2024. His current research interests include theory and method of measurement, forecasting and scheduling.

Robert P. Dick (S'95-M'02-Sr'20) is a Professor of Electrical Engineering and Computer Science at the University of Michigan. He received his Ph.D. degree from Princeton University in 2002 and his B.S. degree from Clarkson University in 1996. He worked as a Visiting Professor at Tsinghua University's Department of Electronic Engineering in 2002, as a Visiting Researcher at NEC Labs America in 1999, and was on the faculty of Northwestern University from 2003-2008. Robert received an NSF CAREER award and won his department's Best Teacher of the Year award in 2004. In 2007, his technology won a Computerworld Horizon Award and his paper was selected as one of the 30 in a special collection of DATE papers appearing during the past 10 years. His 2010 work won a Best Paper Award at DATE. He served as the Technical Program Committee Co-Chair of the 2011 International Conference on Hardware/Software Codesign and System Synthesis, as an Associate Editor of IEEE Trans. on VLSI Systems, and as a Guest Editor for ACM Trans. on Embedded Computing Systems and IEEE Design & Test of Computers. He was also CEO of Stryd, Inc., which produces wearable electronics for athletes.
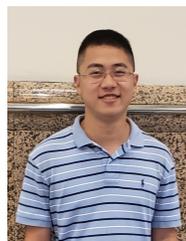
Tony Zhang received the Ph.D. degree with the Department of Electrical and Computer Engineering at the University of Michigan, Ann Arbor, MI, USA in 2023. He also received the M.S. degree in electrical and computer engineering from the University of Michigan, Ann Arbor, MI, USA. in 2019. His research interests include machine learning, computer vision, air quality, and data science.