

RESEARCH ARTICLE

Nighttime Visibility Classification Based on Stable Light Sources

ZHUORAN LIANG¹, YU CAO^{1,2}, ZHILEI WANG², YONGQIANG LI^{1,2}, ZAN CHEN^{1,2}, AND TING SUN³

¹Hangzhou Meteorological Bureau, Hangzhou, Zhejiang 310051, China

²College of Information Engineering, Zhejiang University of Technology, Hangzhou, Zhejiang 310014, China

³School of Culture and Tourism, Zhejiang International Studies University, Hangzhou, Zhejiang 310000, China

Corresponding author: Ting Sun (tingsun811@zisu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 62002327, in part by the National Science Foundation of Zhejiang Province under Grant LQ21F020017 and Grant LZ21F030003, in part by the Agricultural and Social Development Foundation of Hangzhou under Grant 202004A07, and in part by the Intergovernmental International Science and Technology Innovation Cooperation under Grant 2019YFE0124800.

ABSTRACT To enhance the accuracy of existing nighttime visibility estimation methods, this study proposes a classification algorithm for nighttime visibility levels based on stable light sources. Initially, a target detection network identifies all stable streetlights in the image and extracts the light source blocks. Subsequently, these blocks undergo fog classification through a classification network. The blocks are then sorted by brightness values and assigned corresponding weights. Finally, the classification results and weights are combined to categorize the nighttime image visibility levels. Experimental results show that the accuracy of our nighttime visibility classification algorithm reaches 77.6% on real-world datasets, outperforming existing methods and demonstrating good generalization across different scenes.


INDEX TERMS Night images, visibility classification, light source block, generalization.

I. INTRODUCTION

Visibility has significant impacts on various aspects of daily life, such as transportation, logistics, and production. Nighttime visibility is closely related to the occurrence of accidents. The primary factors influencing visibility are solid particles and small liquid droplets suspended in air [1]. Accurate classification of nighttime visibility is crucial for timely and effective implementation of related protective measures, reducing the occurrence of accidents, and playing an important role in social development [2]. Existing visibility estimation instruments mainly include forward-scattering visibility meters, which detect visibility based on atmospheric scattering principles [3]. However, issues such as high instrument cost, large size, and high maintenance expenses hinder the widespread deployment of these devices.

As computer vision develops, visibility estimation is no longer limited to physical instruments. Researchers have

previously proposed visibility estimation methods based on video images [4]. Approaches for constructing visibility and images can be divided into two main ways: physical model-based [5] and deep learning-based methods [6]. The visibility estimation method using a physical model primarily relies on the atmospheric scattering model and associated theories. This type of method, while interpretable to some extent, demands specific conditions and relies on additional parameters like distance and depth information. And, they have poor performance in terms of generalizability as they require matching different clear images for different scenarios, and the issue of atmospheric light values in multi-light source nighttime scenes has not been fully addressed. With the rapid development of CNN, researchers have been studying visibility estimation algorithm driven by images [7]. To fully exploit the learning capabilities of CNNs, these methods map foggy images directly to visibility levels using CNNs. However, most simply use CNNs for visibility estimation without extracting specialized features for this task. Additionally, the majority of current deep learning-based visibility estimation techniques rely on

The associate editor coordinating the review of this manuscript and approving it for publication was Yiming Tang .

CNNs, which are prevalent across various fields and excel at extracting low-level features. High-level visual semantic information typically focuses on how elements interconnect to form objects and how spatial relationships between objects create scenes, all of which merit consideration. In summary, existing visibility estimation methods face several issues, including excessive constraints and lower accuracy.

To overcome the drawbacks of current methods, we present a novel algorithm for classifying nighttime visibility based on street-lights. In the detection network, we use ResNet50-*vd* as the backbone for feature extraction. Inspired by the tricks proposed by [23], we replace the final 3×3 convolutional layer of ResNet50-*vd* with a deformable convolutional layer to optimize the backbone and enhance its performance. Additionally, to effectively leverage multi-scale information, we also employ the Path Aggregation Network (PAN) to process the extracted features. Unlike the methods mentioned above, the proposed method concentrates exclusively on analyzing the regions within the image that contain stable light sources. In our method, This approach enables the classification of the entire nighttime image while mitigating interference. In addition, a dataset comprising authentic nighttime surveillance images is made. Experimental results demonstrate that the our method outperforms the counterparts in terms of classification accuracy, particularly when dealing with limited sample data, underscoring the proposed method robust generalization capability.

The main contributions of this paper are summarized as follows:

- This paper proposes a classification algorithm for nighttime visibility levels based on streetlights, utilizing a target detection network to identify and extract all light source blocks in images.
- A classification network is employed to categorize the extracted light source blocks into four visibility level, and Color space conversion is applied, and the blocks are sorted by luminance values to determine light source weights. Combined with fog classification results and weights, the overall visibility level of the image is assessed.
- Performances of the proposed approaches are verified by extensive experiments on our dataset. The results show that the proposed visibility level estimation method outperforms its competitive methods.

The rest of this paper is organised as follows: Section II introduces the related works. Section III presents the proposed methods that contains the object detection network and the fog classification. Section IV gives the experimental validation. Finally, Section V concludes this paper.

II. PRELIMINARY WORKS

In this section, we present current status and progress of research in image-based visibility estimation, including physical model-based methods in section A and deep leaning-based methods in section B.

A. PHYSICAL MODEL-BASED METHODS

Visibility estimation methods based on physical models relies on Koschmieder's law, proposed by Lee and Shang [8], which describes the process of radiative attenuation of the entire observed atmosphere from an object to a sensor. This type of methods rely on atmospheric values, which is different in daytime and nighttime. The imaging of daytime images follows the principle of atmospheric scattering [22], as depicted by Eq.(1),

$$I(x) = J(x)e^{-\beta d(x)} + A(1 - e^{-\beta d(x)}) \quad (1)$$

where x represents the pixel value at a specific position in the image. $I(x)$ and $J(x)$ correspond to the pixel values of the hazy and clear images, respectively. $e^{-\beta d(x)}$ represents the transmission rate, and β is the extinction coefficient. $d(x)$ represents the distance between the object and the camera. A denotes the global atmospheric light value in the image. Fig.1(a) provides a visual representation of daytime imaging processing, with $J(x)e^{-\beta d(x)}$ representing the pixel value of an object that traverses through haze before reaching the camera, and $A(1 - e^{-\beta d(x)})$ representing the pixel value of atmospheric light scattering detected by the camera. Based on this law, He et al. proposed the dark channel prior algorithm [9], which involves extracting the top of 0.5% brightest points in each channel as the atmospheric light value. The transmission value and extinction coefficient are obtained by comparing the clear image with the hazy one, by which the visibility value is inversely calculated. This approach performs well when there is only one light source during the daytime. However, in nighttime, the imaging process varies.

The imaging of nighttime scenes is defined as:

$$I(x) = J(x)e^{-\beta d(x)} + A(1 - e^{-\beta d(x)}) + A_\alpha \cdot \text{APSF} \quad (2)$$

where, α represents a local region within the image, A_α denotes the atmospheric light value in region of the image, APSF is the atmospheric point spread function, and $A_\alpha \cdot \text{APSF}$ refers to the halo effect generated within that particular region. As illustrated in Fig.1(b), during nighttime, when the illumination comes from artificial light sources and is non-unique, the atmospheric light values across the entire image are inconsistent. Consequently, it would lead to significant deviations using this method to detect nighttime visibility. To address the issue, Pei et al. proposed a preprocessing step, mapping the color space from RGB to LAB [10], and then optimizing the image's extinction coefficient using the dark channel prior and bilateral filtering. Li et al. introduced a new nighttime scattering model considering the interference of halo in nighttime images on visibility [11], [12], which separates the halo layer from the image and performs visibility estimation using the dark channel prior algorithm. He et al. assumed the maximum brightness value of each channel in the clear image to be 1 [9], promoting the maximum reflection prior algorithm. Yu et al. proposed a pixel-based alpha blending method for estimating the transmission

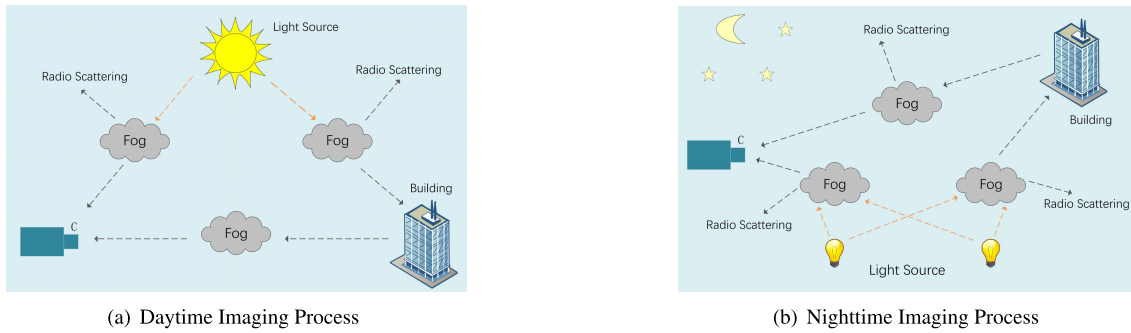


FIGURE 1. Daytime and nighttime imaging process diagram.

map [14], which is guided by luminance-aware weights derived from dark channel estimation. The method preserves high-frequency edge information through the application of Retinex theory. Yang et al. estimated the atmospheric light value using a superpixel algorithm and obtained the extinction coefficient by combining image halo layering and dark channel algorithms [15]. Gallen et al. performed visibility classification by calculating the attenuation of all light source illumination directions in the image [16]. However, due to its strict requirements on the scene and numerous necessary auxiliary parameters, that method is difficult to apply in the complex scene.

B. DEEP LEARNING-BASED METHODS

Li et al. utilized a CNN-RNN network to extract image features for estimating image visibility [17]. They employ a pre-trained convolutional neural network (CNN) to automatically extract the visibility features instead of manual method and design a generalized regression neural network (GRNN) for intelligent visibility evaluation based on these deep learning features. Zhang et al. trained the HazDesNet using paired images of the same scene with and without haze [18], employing the Structural Similarity Index (SSIM) score between the two images as the regression target to estimate the visibility in outdoor images. But the variability of haze levels in the environment poses significant challenges for the model's adaptability and responsiveness to different degrees of haze, and a single SSIM objective may not suffice to achieve such detailed differentiation. Choi et al. converted images to the HSL (Hue, Saturation, and Lightness) and used the VGG (Visual Geometry Group) to predict image visibility [19]. Yang et al. presented a self-enhanced image dehazing framework called D4 (Dehazing via Decomposing transmission map into Density and Depth) [20]. This framework overcomes overfitting issues in synthetic image dehazing and calculates the scattering coefficient of the original image. However, it usually over-estimates the transmission of extreme bright area, which will mislead the depth estimation network to predict low depth value for over-bright areas. Sharma et al. proposed a method for decomposing images into high-frequency and low-frequency

feature maps [21]. The low-frequency feature maps are processed using deep learning networks, and then combined with high-frequency features to obtain clear nighttime images. The visibility is calculated by comparing the transmission rate ratio of the clear image to the foggy image. However, deep learning-based algorithms classify visibility levels of input images directly through neural networks, which requires substantial amounts of data and necessitates a high level of network complexity to achieve satisfactory classification results.

In conclusion, physical model-based methods struggle with the complex and varied nature of night lighting. Deep learning-based approaches also face challenges such as low accuracy and high network complexity. Therefore, we propose to estimate visibility by conducting research on stable light sources during the night.

III. METHODOLOGY

In this part, Section A is the structure of our method. Section B and C introduce our light sources Detection network, visibility level estimation network and its related theories, Section D introduces our training process.

A. THE OVERVIEW OF OUR METHOD

As depicted in Fig.2, a detection network is designed to identify all light source blocks. Subsequently, these identified blocks undergo a fog classification process leveraging a specialized classification model. Meanwhile, color space conversion from RGB to LMN is executed on the light source blocks, facilitating the assignment of weights. These weights are derived from the sorting of the light source blocks by brightness values across all detected blocks. Finally, the visibility level is obtained by combing the classification results and the light values analysis.

B. DETECTION NETWORK

The framework for light source blocks detection [23], [24] is illustrated in Fig.3. We adopt the ResNet50-vd-dcn (the final convolutional layer replaced by a deformable convolutional network) to process the input images, generating feature maps. The top three feature maps l_1 , l_2 and l_3 are aggregated

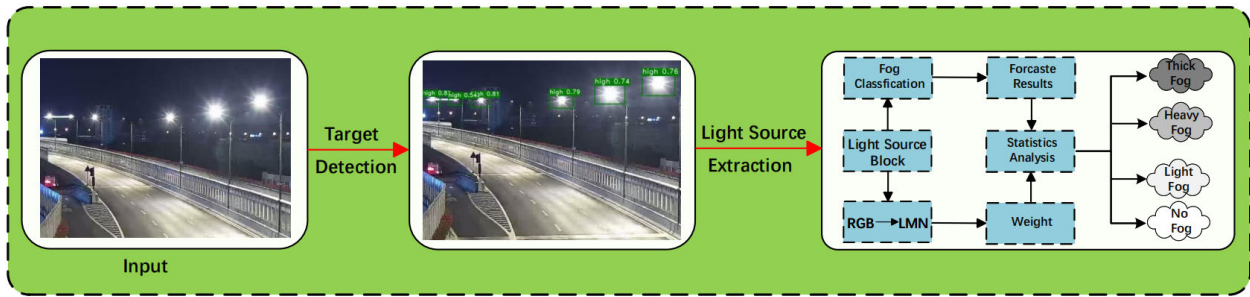


FIGURE 2. Framework diagram of the streetlight-based night visibility classification algorithm.

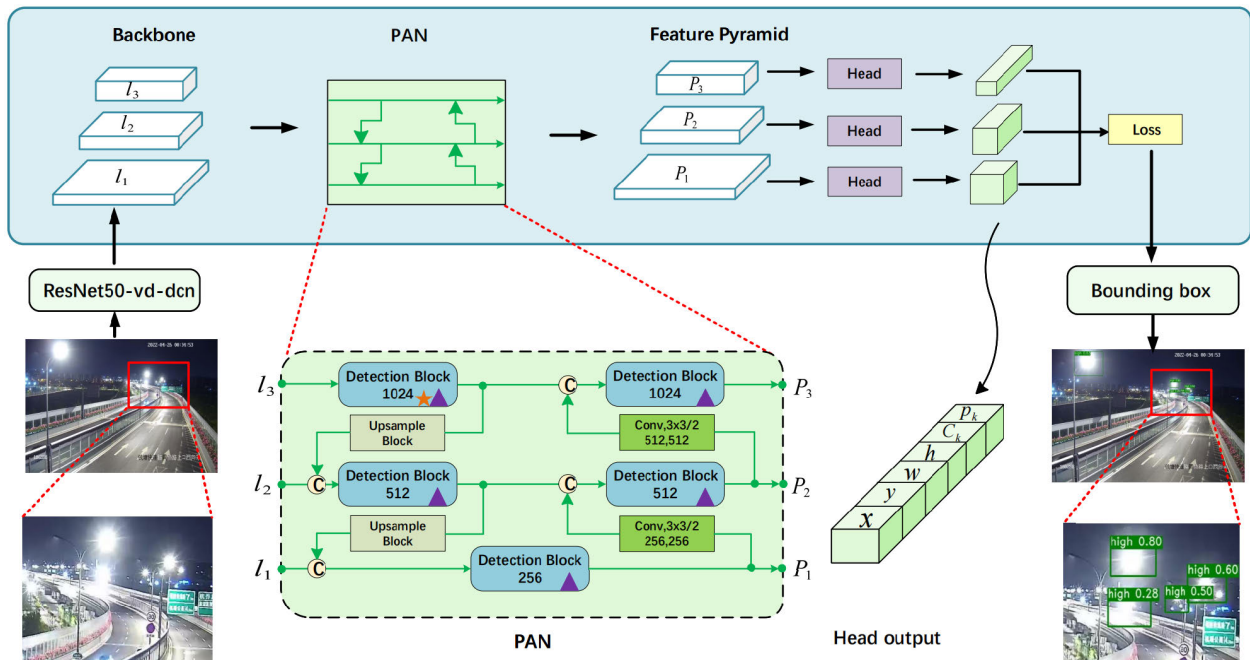


FIGURE 3. Schematic framework diagram of detection network.

using Path Aggregation Network(PAN) to form a Feature Pyramid, enhancing detection accuracy through the utilization of information from each feature map. Each feature map is fed into the Head for detection and loss calculation, with the total loss computed after aggregating information from each extracted depth image.

1) HEAD

The detection head comprises two convolutional layers, employing a 3×3 convolution followed by a 1×1 convolution layer for obtaining the final predictions. The network divides each image into $S \times S$ grids, with each grid corresponding to B anchors (we set to three), using k-means clustering to determine B scales bounding boxes priors for light sources identification. The network regresses the prior bounding box that has the highest IOU value with the ground truth, outputs $\{x_k, y_k, w_k, h_k\}_{k=1}^{S^2}, \{p_k\}_{k=1}^{S^2}, \{c_k\}_{k=1}^{S^2}$, representing the values of the center coordinate, height, width of the predicted bounding box, the class probability and the object score. The output channel of each final prediction

is $3(C + 5)$, where C is the number of classes. For each anchor, the first C channels predict the probabilities for C classes, followed by 4 channels predicting the location of the bounding box. The last channel predicts the object score. Cross-entropy loss is used for classification, and L1 loss for localization. The object loss supervises the object score, identifying the presence of an object.

2) INPUT AND OUTPUT

We manually annotated 1750 real-world night-time images from surveillance footage, marking all light sources on each image. The training set images were resized to 640×640 , with a batch size set to 1. We detected all light sources within the entire image. The first row (a, b, c) displays three scenarios, the second row (d, e, f) shows the detection results, as shown in Fig. 4.

C. VISIBILITY ESTIMATION NETWORK

All light sources $\{b_1, b_2 \dots b_n\}$, after being normalized based on their sizes and brightness, are fed into the



FIGURE 4. Target detection results for three scenarios.

network, conducting fog classification and weights calculating, as depicted Fig 5. A classification network, consisted of three convolutional layers, pooling layers, and two fully connected layers, is employed to classify these blocks into four visibility level. Each image is labeled with v , and the light source blocks are fed into the classification network. The final output of the network is $q(v)$, a 1×4 tensor $[q_1, q_2, q_3, q_4]$, where q represents the categorical probability of the image.

In each image, light source blocks vary in scale and foggy images tend to have lower and more uniform brightness around light sources. Therefore, we weight all light source blocks on the entire image that the brighter and bigger the light source block, the greater its representative weight is considered. Initially, each light source undergoes color space conversion. The brightness feature extracted from the RGB color space brightness channel does not adequately describe color distortion. The LMN color space is well-defined on a physical basis, and the conversion can eliminate the correlation between the luminance component (L) and the chrominance components (M, N) [22]. The L channel information is primarily used when weighting the light sources, as Eq.(3):

$$L = 0.06R + 0.63G + 0.27B \quad (3)$$

Secondly, drawing from the principle of image super-resolution, light sources are sorted based on their brightness values. And, each light source is weighted according to its scale and intensity within the image. The weight is defined as follows:

$$g_{L,i} = \frac{1}{|Q_i|} \sum_{(m,n) \in Q_i} I_L(m, n), \quad i = 1, 2 \dots \eta \quad (4)$$

$$f_{L,i} = \log_2(1 + \frac{Rank_i}{\eta}), \quad i = 1, 2 \dots \eta \quad (5)$$

$$W_{L,i} = \frac{\sum_1^\eta (f_{L,i} \cdot |Q_i|)}{\sum_1^\eta f_{L,i}}, \quad i = 1, 2 \dots \eta \quad (6)$$

where, $I_L(m, n)$ represents the pixel values of L channel (m, n) , Q_i denotes the coordinate set of the i_{th} light source patch in an image. $|Q_i|$ signifies the number of pixels in Q_i , η indicating the count of light source patches present in the image. $Rank_i$ denotes the index that reflects the brightness values ordering of light source patches within the entire image. $g_{L,i}$ represents the average pixel value of the i_{th} light source patch in L channel. $f_{L,i}$ denotes the brightness weight value assigned to the i_{th} light source patch, while $W_{L,i}$ represents the maximum weight value among all light source patches.

We multiply the weight of each light source block by its classification results and sum these products by category. Finally, we regard the weighted sum of the category with the highest total as the visibility level, which is calculated as:

$$V_{est} = \sum_{i=1}^{\eta} W_{L,i} \cdot q(v_i) \quad (7)$$

D. TRAINING PROCESS

1) DATASETS AND IMPLEMENTATIONS DETAILS

Since the limited quantity of real-world images is insufficient to adequately assess the algorithm's accuracy, we create a novel nighttime image dataset for this study. The selected nighttime images encompass complete light sources. The dataset consists of 46 sites, totaling 4861 images. Among these, 37 sites are designed for training, encompassing around 3551 images, while the remaining 9 sites are reserved for testing and validation, comprising 1310 images. The image collection period for the selected 46 sites spanned from April 2022 to May 2023, ensuring a comprehensive coverage of all seasons throughout a full year. In line with practical societal requirements, the nighttime visibility levels are categorized into four distinct classes: 0-0.2 km (thick fog), 0.2-1.0 km (heavy fog), 1-10.0 km (light fog), and >10.0 km (clear). Fig.6 presents illustrative examples of images representing different visibility levels. It is evident from Fig.6 that the selected real-world surveillance site images encompass a

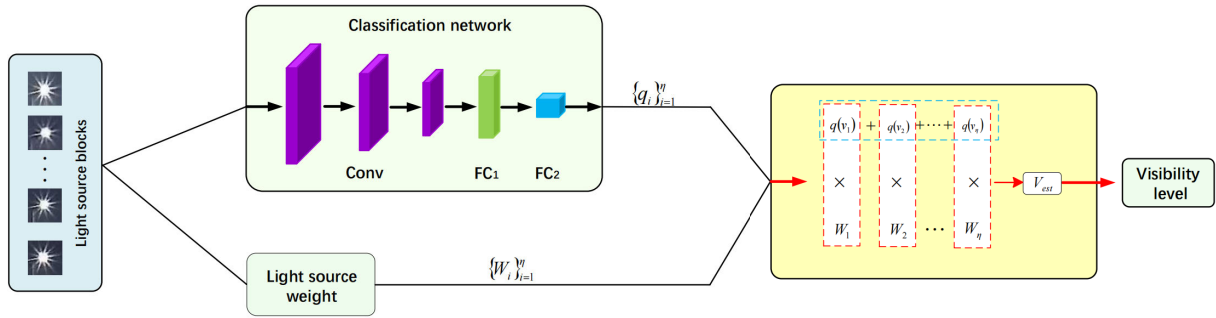


FIGURE 5. The structure of visibility estimation network.

variety of perspectives, including aerial views, horizontal views, and upward views, thereby catering to a diverse range of societal needs.

Given the absence of real visibility values at surveillance sites, the nearest meteorological monitoring site is determined by correlating the latitude and longitude information provided by the meteorological bureau with that of the surveillance sites. We select the meteorological station located within a radius of approximately 3km from each surveillance site to obtain the actual visibility values. we also conduct visual inspection and classification of the images to ensure compliance with the desired criteria.

We uses two models, a detection model and a classification model, which we train separately. Training the detection model we use the SGD optimizer, momentum is set to 0.9. The initial learning rate is 0.005 / 12 and the batch size is set to 1, for a total of 30 epochs. Training the classification model we use the adadelta optimizer, the initial learning rate is set to 0.001, epsilon defaults to 10^{-6} , and the batch size is 8, for a total of 200 epochs.

2) LOSS

The losses output from the three detection heads are added to constitute the total loss for network updates. The loss function is as follows:

$$Loss_{total} = Loss_{coord} + Loss_{obj} \quad (8)$$

The coordinate prediction loss function is as follows:

$$Loss_{coord} = \lambda_{coord} \sum_{k=1}^{S^2} \sum_{j=1}^B I_{k,j}^{obj} [(x_k - \hat{x}_k)^2 + (y_k - \hat{y}_k)^2] + \lambda_{coord} \sum_{k=1}^{S^2} \sum_{j=1}^B I_{k,j}^{obj} [(h_k - \hat{h}_k)^2 + (w_k - \hat{w}_k)^2] \quad (9)$$

where λ_{coord} is the weight of the coordinate error.

The object Loss is defined as follows:

$$Loss_{obj} = \sum_{k=1}^{S^2} \sum_{j=1}^B I_{k,j}^{obj} (C_k - \hat{C}_k)^2$$

$$+ \lambda_{noobj} \sum_{k=1}^{S^2} \sum_{j=1}^B I_{k,j}^{obj} (C_k - \hat{C}_k)^2 \quad (10)$$

where the parameter λ_{noobj} is the weight, default 0.5, \hat{C}_k is the predicted confidence, and C_k is the true confidence, defined as $sigmoid(\hat{C}_k)$.

The loss for this classification is defined as:

$$Loss_{cate} = -\frac{1}{N} \sum_{i=1}^N [v_i \cdot \log_2(q(v_i))] \quad (11)$$

In Eq.(11), N denotes the number of images.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In this paper, experimental evaluations are performed on the NVIDIA GeForce RTX 3060 GPU to validate the efficacy of the proposed algorithm. Real-world image datasets obtained from surveillance sites are employed for the purpose of conducting rigorous testing and validation.

A. EVALUATION METHOD

The confusion matrix is often used as a evaluation index in deep learning, including precision, recall, accuracy. A confusion matrix is used to evaluate the proposed method in this paper. Equations 12, 13 and 14 show their calculation methods.

$$precision_i = \frac{TP_i}{TP_i + FP_i} \quad i = 1, 2, 3, 4 \quad (12)$$

$$recall_i = \frac{TP_i}{TP_i + FN_i} \quad i = 1, 2, 3, 4 \quad (13)$$

$$accuracy = \frac{\sum_{i=1}^4 TP_i}{\sum_{i=1}^4 TP_i + FP_i} \quad (14)$$

where i dnates the i_{th} visibility range, TP, FP and FN are separately presented as true positive, false positive and false negative.

B. COMPARISON EXPERIMENT

To validate the effectiveness of the proposed method in this paper, we compare it with the following three methods.

In [2], the atmospheric light value is estimated using atmospheric modeling theory. The ratio of the extinction



FIGURE 6. Example of images with different visibility levels.

coefficients is derived through the application of the dark channel prior and edge-collapse algorithm. And, the region of interest (ROI) is determined using the dark channel thresholding method. Finally, the visibility value is computed by integrating the ratio of transmission rates and extinction coefficients for both clear and blurred regions within the ROI area.

In [11] involves decomposing the nighttime image into a halo layer and a background layer. It assumes that the atmospheric light value remains locally constant for the background layer. Then, the image is divided into smaller blocks, and the atmospheric light value is estimated for each block. Finally, the atmospheric light value for the entire image is obtained using guided filtering, enabling the computation of the image visibility.

In [26], a deep residual learning framework named ResNet50 is developed. This framework incorporates a 152-layer residual network, where each layer adheres to residual mapping principles. Within the scope of this study, this approach is employed for classifying nighttime image visibility. A training dataset comprising 2448 nighttime social video surveillance images from four distinct categories is selected for model training.

We compare the time it takes our method to infer a picture with several competitive methods, as shown in TABLE 1, and compare the performance of algorithm in different scenarios, as shown in the following tables.

TABLE 2 provides the experimental outcomes for the entire test dataset, analyzing the data, it is evident that the proposed method exhibits superior accuracy compared to the counterparts across the first three visibility ranges. Particularly in the low visibility range of 0-0.2 km, ours

TABLE 1. Comparison of the computational efficiency of our method with competitive methods.

Methods	[2]	[11]	[26]	Our
Times(s)	0.05	133.82	0.19	1.25

precision and recall achieve 83.7% and 58.5%, respectively. Moreover, the overall accuracy of our method surpasses that of the other three methods, reaching 77.6%. Our method demonstrates significant advancements in the prediction and classification of low visibility images compared to alternative methods.

TABLE 3 presents the comparison results of various algorithms in Scene 1 (view at eye level), which is a front-view image on an elevated bridge. At night when there is heavy vehicle traffic, it is common for headlights to directly face the camera. Algorithms based on physical models in [2] and [11], tend to produce high atmospheric light values in such scenarios, leading to overestimated visibility predictions. Thus, while these methods are fairly accurate for high visibility image classification, they exhibit significant errors in low visibility conditions. The direct illumination from vehicle lights causing image overexposure can lead to misjudgments by the algorithm proposed in [26]. In contrast, our algorithm focuses solely on fixed light sources like street lamps, reducing interference from vehicles and other variables.

TABLE 4 shows the comparison results of various algorithms in Scene 2 (view from below), where the images are taken from an upward-looking angle, causing the street lamps to appear higher in the frame, leading to overall image blurriness. In this scenario, the algorithm proposed in [2]

TABLE 2. Test result of each method at real social points (%).

Visibility range	Number/Piece	[2]			[11]			[26]			our		
		recall	precision	accuracy	recall	precision	accuracy	recall	precision	accuracy	recall	precision	accuracy
0-0.2 km	92	17.4	16.7	69.9	17.4	16.2	71.1	35.9	27.7	67.2	83.7	58.5	77.6
0.2-1.0 km	205	64.4	39.4		65.5	68.4		48.8	62.1		70.2	73.0	
1.0-10.0 km	221	67.0	60.7		62.4	46.9		49.8	38.9		70.6	53.0	
>10.0 km	792	78.3	90.1		81.3	91.1		80.4	85.3		80.8	94.0	

TABLE 3. Test results of each method in Scene 1 (view at eye level)(%).

Visibility range	Number/Piece	[2]			[11]			[26]			our		
		recall	precision	accuracy	recall	precision	accuracy	recall	precision	accuracy	recall	precision	accuracy
0-0.2 km	35	22.9	17.0	76.5	25.7	19.6	71.1	28.5	18.5	67.9	88.6	36.9	74.2
0.2-1.0 km	106	64.2	66.7		56.6	60.0		46.2	53.9		50.9	63.5	
1.0-10.0 km	106	76.4	52.3		66.0	41.2		52.8	36.4		85.8	49.5	
>10.0 km	497	82.9	93.2		78.5	91.2		78.5	89.7		75.7	96.1	

TABLE 4. Test results of each method in Scene 2 (view form below)(%).

Visibility range	Number/Piece	[2]			[11]			[26]			our		
		recall	precision	accuracy	recall	precision	accuracy	recall	precision	accuracy	recall	precision	accuracy
0-0.2 km	37	10.8	14.8	74.0	8.1	12.5	71.3	35.1	54.2	66.9	83.8	72.1	83.6
0.2-1.0 km	83	62.7	70.3		73.5	68.5		55.4	48.8		90.4	79.8	
1.0-10.0 km	48	81.3	56.5		60.4	43.3		75.0	58.1		45.8	55.0	
>10.0 km	185	89.7	91.7		85.9	85.9		90.3	92.8		90.3	90.3	

TABLE 5. Test results of each method in Scene 3 (view form above)(%).

Visibility range	Number/Piece	[2]			[11]			[26]			our		
		recall	precision	accuracy	recall	precision	accuracy	recall	precision	accuracy	recall	precision	accuracy
0-0.2 km	20	25.0	41.7	66.2	25.0	50.0	71.4	50.0	50.0	53.1	75.0	93.8	79.8
0.2-1.0 km	16	75.0	38.7		81.3	44.8		31.3	23.8		93.8	78.9	
1.0-10.0 km	67	41.8	68.3		58.2	66.1		26.9	34.6		64.2	76.8	
>10.0 km	110	87.3	74.4		86.4	76.0		72.7	66.7		88.2	77.8	

results in excessively high overall pixel values, while the algorithm from [11] causes overexposure at the top of the image when removing halo effects, making the image appear overly bright. The algorithm mentioned in [26] tends to misjudge the upper light sources, perceiving the image as overexposed. Our algorithm improves accuracy by training on light source blocks under various conditions, enhancing the precision of judgments.

TABLE 5 presents the comparison results of various algorithms in Scene 3(view form above), where images are captured from an aerial surveillance camera in a downward-looking view. Nighttime aerial views often include images of buildings, and the sporadic brightening and dimming of buildings at night lead to errors in the transmission maps calculated by the algorithms proposed in [2] and [11]. The algorithm in [26] experiences a decrease in accuracy due to interference from vehicle and building information.

Our algorithm, focusing solely on street light sources, minimizes errors caused by buildings and vehicle lights, thereby enhancing classification accuracy.

In conclusion, the test results for the 3 scenes depicted in TABLE 3, 4 and 5 demonstrate that our method outperforms the others, indicating that our algorithm maintains high accuracy across different viewing angles.

C. ABLATION EXPERIMENT

To further validate the impact of each module described in this paper on the experimental results, we remove the model that classifies light source blocks and instead perform visibility level classification directly within the detection network, named as DIS_cla. In detection network, we manually label the dataset by assigning category labels to the light sources based on image visibility for training purposes. Additionally,

TABLE 6. The result of ablation experiment(%).

Visibility range	Number/Piece	DIS_cla		DIS_rgb			our			
		recall	precision	accuracy	recall	precision	accuracy	recall	precision	accuracy
0-0.2 km	92	59.8	41.4	70.2	25.0	17.4	61.9	83.7	58.5	77.6
0.2-1.0 km	205	56.1	68.5		39.5	43.1		70.2	73.0	
1.0-10.0 km	221	57.5	42.9		46.2	38.2		70.6	53.0	
>10.0 km	792	78.7	87.9		76.4	83.7		80.8	94.0	

we eliminate the color space conversion module and define the method as DIS_rgb.

The classification loss in detection network is depicted as follows:

$$Loss_{cls} = \sum_{k=1}^{S^2} \sum_{j=1}^B I_{k,j}^{obj} \sum_{c \in classes} (p_k(c) - \hat{p}_k(c))^2 \quad (15)$$

where S^2 is the number of grids in the input image, and B is the number of bounding boxes generated by each grid. $I_{k,j}^{obj} = 1$ denotes that the object falls into the j_{th} bounding box in grid i , $p_k(c)$ refers to the true probability that the object belonging to class c is in grid k . $\hat{p}_k(c)$ is the predicted value. c refers to the class to which the detected target belongs. In training, we set c to 4, representing four visibility level.

TABLE 6 displays the results of the ablation study. In TABLE 6, using the DIS_cla method leads to a significant decrease in the accuracy of night-time visibility level classification, with notably lower accuracy in classifying images of low visibility. This suggests that while the object detection model is highly accurate in identifying street light, the diversity of light types results in lower accuracy during fog classification, thereby reducing the overall classification accuracy of the images. Furthermore, eliminating the color space conversion module (DIS_rgb) results in larger errors in classification accuracy across various intervals, indicating that the inter-channel effects in the RGB color space are significant and not conducive to improving classification accuracy.

V. CONCLUSION

This paper proposes a night-time visibility estimation algorithm based on street lights, utilizing a detection network to identify and extract all light source blocks in an image, which are then classified into four categories using a classification network. Within the classification network, the light source blocks are sorted by brightness and weighted according to their scales, finally, combining the classification results and weights, an estimation of the visibility level is made. Experiments show that in a small dataset, compared to other methods, our algorithm demonstrates stronger interference resistance and higher night-time visibility classification accuracy in complex real-world scenarios. In real traffic monitoring images, where scenes are highly variable and contain many sources of interference, our algorithm focuses on information from light blocks, reducing interference

and enhancing classification accuracy. Additionally, our method does not require extensive training data, nor does it need to calculate the transmittance ratio between clean and foggy images, making it less time-consuming and more generalizable than other methods. Future work will focus on optimizing the algorithm for night-time adverse weather conditions like rain or snow, aiming to enhance the accuracy of visibility estimation.

ACKNOWLEDGMENT

(Zhuoran Liang and Yu Cao are co-first authors.)

REFERENCES

- [1] J. You, S. Jia, X. Pei, and D. Yao, "DMRVisNet: Deep multi-head regression network for pixel-wise visibility estimation under foggy weather," 2022, *arXiv:2112.04278*.
- [2] Q. Li, Y. Li, and B. Xie, "Single image-based scene visibility estimation," *IEEE Access*, vol. 7, pp. 24430–24439, 2019, doi: 10.1109/ACCESS.2019.2894658.
- [3] K. Zhang, X. Zheng, X. Yan, S. Chen, X. Wan, J. Zhang, Q. Zhao, B. Wang, and H. Zhao, "Research on visibility detection based on forward scattering technology," *Proc. SPIE*, vol. 11191, 2019, Art. no. 111911A, doi: 10.1117/12.2540728.
- [4] C. Wang, Y. Dai, W. Zhou, and Y. Geng, "A vision-based video crash detection framework for mixed traffic flow environment considering low-visibility condition," *J. Adv. Transp.*, vol. 2020, pp. 1–11, Jan. 2020.
- [5] Y. Guo, Q. Xu, Y. Su, and S. Jiang, "Visibility detection based on the recognition of the preceding vehicle's taillight signals," *IEEE Access*, vol. 8, pp. 206105–206117, 2020.
- [6] Z. Liu, Y. Chen, X. Gu, J. K. W. Yeoh, and Q. Zhang, "Visibility classification and influencing-factors analysis of airport: A deep learning approach," *Atmos. Environ.*, vol. 278, Jun. 2022, Art. no. 119085.
- [7] H. Ullah, K. Muhammad, M. Irfan, S. Anwar, M. Sajjad, A. S. Imran, and V. H. C. de Albuquerque, "Light-DehazeNet: A novel lightweight CNN architecture for single image dehazing," *IEEE Trans. Image Process.*, vol. 30, pp. 8968–8982, 2021.
- [8] Z. Lee and S. Shang, "Visibility: How applicable is the century-old Koschmieder model?" *J. Atmos. Sci.*, vol. 73, no. 11, pp. 4573–4581, Nov. 2016.
- [9] 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, Dec. 2011.
- [10] S.-C. Pei and T.-Y. Lee, "Nighttime haze removal using color transfer pre-processing and dark channel prior," in *Proc. 19th IEEE Int. Conf. Image Process.*, Sep. 2012, pp. 957–960.
- [11] Y. Li, R. T. Tan, and M. S. Brown, "Nighttime haze removal with glow and multiple light colors," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 226–234.
- [12] Y. Li and M. S. Brown, "Single image layer separation using relative smoothness," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2752–2759.
- [13] J. Zhang, Y. Cao, S. Fang, Y. Kang, and C. W. Chen, "Fast haze removal for nighttime image using maximum reflectance prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 7016–7024.

- [14] T. Yu, K. Song, P. Miao, G. Yang, H. Yang, and C. Chen, "Nighttime single image dehazing via pixel-wise alpha blending," *IEEE Access*, vol. 7, pp. 114619–114630, 2019.
- [15] M. Yang, J. Liu, and Z. Li, "Superpixel-based single nighttime image haze removal," *IEEE Trans. Multimedia*, vol. 20, no. 11, pp. 3008–3018, Nov. 2018.
- [16] R. Gallen, A. Cord, N. Hautière, É. Dumont, and D. Aubert, "Nighttime visibility analysis and estimation method in the presence of dense fog," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 1, pp. 310–320, Feb. 2015.
- [17] S. Li, H. Fu, and W.-L. Lo, "Meteorological visibility evaluation on webcam weather image using deep learning features," *Int. J. Comput. Theory Eng.*, vol. 9, no. 6, pp. 455–461, 2017.
- [18] J. Zhang, X. Min, Y. Zhu, G. Zhai, J. Zhou, X. Yang, and W. Zhang, "HazDesNet: An end-to-end network for haze density prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 4, pp. 3087–3102, Apr. 2022.
- [19] Y. Choi, H.-G. Choe, J. Y. Choi, K. T. Kim, J.-B. Kim, and N.-I. Kim, "Automatic sea fog detection and estimation of visibility distance on CCTV," *J. Coastal Res.*, vol. 85, pp. 881–885, May 2018.
- [20] Y. Yang, C. Wang, R. Liu, L. Zhang, X. Guo, and D. Tao, "Self-augmented unpaired image dehazing via density and depth decomposition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 2037–2046.
- [21] A. Sharma and R. T. Tan, "Nighttime visibility enhancement by increasing the dynamic range and suppression of light effects," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 11972–11981.
- [22] Y. Zheng, J. Zhan, S. He, J. Dong, and Y. Du, "Curricular contrastive regularization for physics-aware single image dehazing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 5785–5794.
- [23] X. Long, K. Deng, G. Wang, Y. Zhang, Q. Dang, Y. Gao, H. Shen, J. Ren, S. Han, E. Ding, and S. Wen, "PP-YOLO: An effective and efficient implementation of object detector," 2020, *arXiv:2007.12099*.
- [24] X. Huang, X. Wang, W. Lv, X. Bai, X. Long, K. Deng, Q. Dang, S. Han, Q. Liu, X. Hu, D. Yu, Y. Ma, and O. Yoshie, "PP-YOLOv2: A practical object detector," 2021, *arXiv:2104.10419*.
- [25] T. Xiang, Y. Yang, and S. Guo, "Blind night-time image quality assessment: Subjective and objective approaches," *IEEE Trans. Multimedia*, vol. 22, no. 5, pp. 1259–1272, May 2020.
- [26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.



ZHILEI WANG received the master's degree from the Zhejiang University of Technology in 2024. His research interests include computer vision and image processing and deep learning.



YONGQIANG LI received the Ph.D. degree from Beijing Jiaotong University. He is an Associate Professor with the College of Information Engineering, Zhejiang University of Technology. His research interests include reinforcement learning and control theory.



ZAN CHEN received the Ph.D from Xi'an Jiaotong University in 2019. He was a Visiting Scholar at the University of East Anglia (UEA), Norwich, U.K., in 2018. Currently, he is an Associate Professor with the College of Information Engineering, Zhejiang University of Technology. His research interests include compressive sensing, computer vision, and medical image processing.



ZHUORAN LIANG was born in 1983. He received the master's degree. He is currently a Senior Engineer. He is interested in image processing and atmospheric science.



YU CAO is a master's student with the Zhejiang University of Technology. His research interests include artificial intelligence, computer vision and deep learning.



TING SUN received the Ph.D. degree from Northwest University. She is currently a Lecturer with Zhejiang International Studies University. Her research interests include application of geographic information science and cartography.

...